NIST Special Publication 1275v1

NIST Conference Papers Fiscal Year 2018 Volume 1: Engineering Laboratory

Compiled and edited by: Information Services Office

This publication is available free of charge from: https://doi.org/10.6028/NIST.SP.1275v1



NIST Special Publication 1275v1

NIST Conference Papers Fiscal Year 2018 Volume 1: Engineering Laboratory

Compiled and edited by: Resources, Access, and Data Team Information Services Office

This publication is available free of charge from: https://doi.org/10.6028/NIST.SP.1275v1

January 2022



U.S. Department of Commerce Gina M. Raimondo, Secretary

National Institute of Standards and Technology James K. Olthoff, Performing the Non-Exclusive Functions and Duties of the Under Secretary of Commerce for Standards and Technology & Director, National Institute of Standards and Technology Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

> National Institute of Standards and Technology Special Publication 1275v1 Natl. Inst. Stand. Technol. Spec. Publ. 1275v1, 691 pages (January 2022) CODEN: NSPUE2

> > This publication is available free of charge from: https://doi.org/10.6028/NIST.SP.1275v1

Foreword

NIST is committed to the idea that results of federally funded research are a valuable national resource and a strategic asset. To the extent feasible and consistent with law, agency mission, resource constraints, and U.S. national, homeland, and economic security, NIST will promote the deposit of scientific data arising from unclassified research and programs, funded wholly or in part by NIST, except for Standard Reference Data, free of charge in publicly accessible databases. Subject to the same conditions and constraints listed above, NIST also intends to make freely available to the public, in publicly accessible repositories, all peer-reviewed scholarly publications arising from unclassified research and programs funded wholly or in part by NIST.

This Special Publication represents the work of Engineering Laboratory researchers at professional conferences, as reported in Fiscal Year 2018.

More information on public access to NIST research is available at https://www.nist.gov/ open.

Key words

NIST conference papers; NIST research; public access to NIST research.

Table of Contents

Cox, Daniel; Cutler, Harvey; Ellingwood, Bruce; McAllister, Therese; Paolo,
Gardoni; Peacock, Walter; van de Lindt, John. "Computational Environment for
Modeling and Enhancing Community Resilience: Introducing the Center for Risk-
Based Community Resilience Planning." Paper presented at Second International
Conference on Performance-based and Lifecycle Structural Engineering (PLSE
2015), Brisbane, Australia. December 9, 2015 - December 11, 2015
Libes, Donald; Shin, Seungjun; Woo, Jungyub. "Considerations and
Recommendations for Data Availability for Data Analytics for Manufacturing."
Paper presented at 2015 IEEE International Conference on Big Data (IEEE Big
Data 2015), Santa Clara, CA, United States. October 29, 2015 - November 1, 2015 SP-11
Cedillos, Veronica; Johnson, Laurie; McAllister, Therese; O'Rourke, Thomas;
Rojahn, Chris. "Community Resilience of Lifeline Systems: Societal Needs and
Performance Assessment." Paper presented at 16th World Conference on
Earthquake, 16WCEE 2017, Santiago, Chile. January 9, 2017 - January 13, 2017SP-19
Jain, Sanjay; Lechevalier, David; Libes, Donald. "Issues in Synthetic Data
Generation for Advanced Manufacturing." Paper presented at 2017 IEEE Big Data,
Boston, MA, United States. December 11, 2017 - December 14, 2017 SP-31
Ju, Feng; Lu, Yan. "Smart Manufacturing Systems based on Cyber-physical
Manufacturing Services (CPMS)." Paper presented at IFAC 2017 World Congress,
Toulouse, France. July 9, 2017 - July 14, 2017
Hunston, Donald; White, Christopher. "Kinetics Based Models to Describe
Environmental Degradation of Polymeric Sealants." Paper presented at Adhesion
Society Meeting, St. Petersburg, FL, United States. February 26, 2017 - March 1,
2017
Bentz, Dale; De Belie, Nele; Snoeck, Didier; Wang, Jianyun. "Bond Enhancement
of Repair Mortar via Biodeposition." Paper presented at INTERNATIONAL
CONFERENCE ON ADVANCES IN CONSTRUCTION MATERIALS AND
SYSTEMS, Chennai, India. September 3, 2017 - September 8, 2017SP-50

Gupta, Satyandra; Kootbally, Zeid; Kramer, Thomas; Schlenoff, Craig. "Overview
of an Ontology-based Approach for Kit-Building Applications." Paper presented at
First International Workshop on Semantic Robotics, San Diego, CA, United States.
January 29, 2017 - February 1, 2017
Bock, Conrad; Reichwein, Axel; Szarazi, Jerome. "Integrating Finite Element
Analysis with Systems Engineering Models." Paper presented at NAFEMS World
Congress 2017, Stockholm, Sweden. June 11, 2017 - June 14, 2017
Loftis, Jon 'Derek'; Nguyen, Cuong; Rhee, Sokwoo. "Emerging Flood Model
Validation Frameworks for Street-Level Inundation Modeling with StormSense."
Paper presented at Second International Workshop on Science of Smart City
Operations and Platforms Engineering (SCOPE), Pittsburg, PA, United States. April
21, 2017 - April 21, 2017
Bernstein, William; Ramani, Karthik; Ramanujan, Deverajan. "Design Patterns for
Visualization-Based Tools in Sustainable Product Design." Paper presented at
ASME 2017 International Design Engineering Technical Conference & Computers
and information in Engineering Conference. IDETC/CIE 2017, Cleveland, OH,
United States. August 6, 2017 - August 9, 2017
Ellingwood, Bruce; McAllister, Therese; van de Lindt, John. "The Science Behind
Community Resilience Assessment: An Overview of the Center for Risk-Based
Community Resilience Planning." Paper presented at 12th International
Conference on Structural Safety & Reliability, Vienna, Austria. August 6, 2017 -
August 10, 2017
Donmez, M; Lu, Yan; Witherell, Paul. "A Collaborative Data Management System
for Additive Manufacturing." Paper presented at 37th Computers and Information
in Engineering Conference (CIE), Cleveland, OH, United States. August 6, 2017 -
August 9, 2017
Brundage, Michael; Sexton, Thurston; Sharp, Michael. "Toward Semi-
Autonomous Information Extraction for Unstructured Maintenance Data in Root
Cause Analysis." Paper presented at APMS 2017 International Conference
Advances in Production Management Systems (APMS 2017), Hamburg,
Germany. September 3, 2017 - September 7, 2017

Gong, Mengyan; Poppendieck, Dustin. "An improved method for calibrating solid-

phase microextraction by direct loading." Paper presented at Healthy Buildings	
2017 Asia, Tianan, Taiwan Province of China. September 2, 2017 - September 5,	
2017	28
Franaszek, Marek; Michaloski, John; Proctor, Frederick. "Tolerances and	
Uncertainty in Robotic Systems." Paper presented at International Mechanical	
Engineering Congress & Exposition, Tampa, FL, United States. November 3,	
2017 - November 9, 2017	32
Brundage, Michael; Kulvatunyou, Boonserm; Li, Quanri. "Improvement Strategies	
for Manufacturers using the MESA MOM Capability Maturity Model." Paper	
presented at International Conference on Advances in Production Management	
Systems, the path to intelligent, collaborative, and sustainable manufacturing,	
Hamburg, Germany. September 3, 2017 - September 7, 2017	43
Archenti, Andreas; Donmez, M; Vogl, Gregory. "Identification of machine tools	
linear axes performance using on-machine embedded inertia measurement units."	
Paper presented at Lamdamap 12th International Conference & Exhibition,	
Gloucestershire, United Kingdom. March 15, 2017 - March 16, 2017	52
Ademujimi, Toyosi; Brundage, Michael; Prabhu, Vittaldas. "A review of current	
machine learning techniques used in manufacturing diagnosis." Paper presented at	
APMS 2017 International Conference Advances in Production Management	
Systems (APMS 2017), Hamburg, Germany. September 3, 2017 - September 7,	
2017	62
Bock, Conrad; Sprock, Timothy. "Incorporating Abstraction Methods into System-	
Analysis Integration Methodology for Discrete Event Logistics Systems." Paper	
presented at 2017 Winter Simulation Conference, Las Vegas, NV, United States.	
December 3, 2017 - December 6, 2017	71
Ju, Feng; Lu, Yan; Wang, Feifan. "A Study on Performance Evaluation and Status-	
based Decision for Cyber-Physical Production Systems." Paper presented at 13th	
IEEE Conference on Automation Science and Engineering, Xi'an, China. August	
19, 2017 - August 24, 2017	82
Emfinger, William; Koutsoukos, Xenofon; Neema, Himanshu; Potteiger, Bradley;	
Stouffer, Keith; Tang, CheeYee. "Evaluating the Effects of Cyber-Attacks on	
Cyber Physical Systems using a Hardware-in-the-Loop Simulation Testbed." Paper	

presented at International Symposium on Resilient Communications Systems,
Resilience Week, Wilmington, DE, United States. September 18, 2017 -
September 22, 2017
Brundage, Michael; Carolis, Anna; Kulvatunyou, Boonserm; Macchi, Marco;
Terzi, Sergio. "Maturity models and tools for enabling Smart Manufacturing
Systems: comparison and reflections for future developments." Paper presented at
IFIP 14th International Conference on Product Lifecycle Management, Seville,
Spain. July 9, 2017 - July 12, 2017
Choe, Lisa; Ramesh, Selvarajah; Seif, Mina. "Localized fire tests on steel beams
with different end restraints." Paper presented at The 9th International Symposium
on Steel Structures, Jeju, Republic of Korea. November 1, 2017 - November 4,
2017
Bao, Yihai; Lew, Hai; Main, Joseph; Thonstad, Travis; Weigand, Jonathan. "New
Connections for Enhancing Robustness of Precast Concrete Frame Structures."
Paper presented at 2018 PCI Convention and National Bridge Conference, Denver,
CO, United States. February 20, 2018 - February 24, 2018
Klinger, Alexander; Weiss, Brian. "Identification of Industrial Robot Arm Work
Cell Use Case Characteristics and a Test Bed to Promote Monitoring, Diagnostic,
and Prognostic Technologies." Paper presented at Annual Conference Of The
Prognostics And Health Management Society 2017, St. Petersburg, FL, United
States. October 2, 2017 - October 5, 2017
Evans, Christopher; Fox, Jason; Kim, Felix; Reese, Zachary. "Investigation of
complementary use of optical metrology and x-ray computed tomography for
surface finish in laser powder bed fusion additive manufacturing." Paper presented
at European Society for Precision Engineering and Technology Special Interest
Group, Leuven, Belgium. October 8, 2017 - October 11, 2017
Brundage, Michael; Hoffman, Michael; Morris, Katherine; Sexton, Thurston. "
Hybrid Datafication of Maintenance Logs from AI-Assisted Human Tags." Paper
presented at 2nd Symposium on Data Analytics for Advanced Manufacturing,
Boston, MA, United States. December 11, 2017 - December 14, 2017 SP-241
Fox, Jason; Grantham, Steven; Lane, Brandon; Yeung, Ho; Zarobila, Clarence. "
Performance Characterization of Process Monitoring Sensors on the NIST

Additive Manufacturing Metrology Testbed." Paper presented at Solid Freeform
Fabrication Symposium, Austin, TX, United States. August 7, 2017 - August 9,
2017
Dols, William; Katal, Ali; Polidoro, Brian; Wang, Liangzhu. "An Investigation of
Different Strategies for Solving Coupled Thermal Airflows by Multi-Zone
Network Method." Paper presented at COBEE 2018 Conference, Melbourne,
Australia. February 5, 2018 - February 9, 2018
Candell, Richard; Hany, Mohamed. "Characterization for an Industrial Wireless
Network in a Gas Sensing Scenario." Paper presented at 2017 ISA International
Instrumentation Symposium, Houston, TX, United States. November 6, 2017 -
November 9, 2017
Sharp, Michael; Vogl, Gregory. "Diagnostics of machine tool linear axes via
separation of geometric error sources." Paper presented at 2017 Annual Conference
of the Prognostics and Health Management Society, St. Petersburg, FL, United
States. October 2, 2017 - October 5, 2017
Cheok, Geraldine; Franaszek, Marek. "Propagation of orientation uncertainty of 3D
rigid object to its points." Paper presented at ICCV 2017 International Conference
on Computer Vision, Venice, Italy. October 22, 2017 - October 29, 2017 SP-283
Marvel, Jeremy. "Sensors for Safe, Collaborative Robots in Smart Manufacturing."
Paper presented at IEEE Sensors, Glasgow, United Kingdom. October 29, 2017 -
November 1, 2017
Manzello, Samuel; Naruse, Tomohiro; Suzuki, Sayaka. "Experiments to Quantify
Firebrand Production from Roofing Assembly Combustion." Paper presented at
55th Japanese Combustion Symposium, Toyama, Japan. November 13, 2017 -
November 15, 2017
Candell, Richard; Downs, Anthony; Huang, Hui-Min; Jacoff, Adam; Kimble,
Kenneth; Saidi, Kamel; Virts, Ann. "Events for the Application of Measurement
Science to Evaluate Ground, Aerial, and Aquatic Robots." Paper presented at The
15th IEEE International Symposium on Safety, Security, and Rescue Robotics
2017 (SSRR 2017), Shanghai, China. October 11, 2017 - October 13, 2017 SP-297

Manzello, Samuel; Suzuki, Sayaka. "Investigation on Firebrand Generation from

Reduced-scale Mock-ups." Paper presented at 55th Japanese Combustion
Symposium, Toyama, Japan. November 13, 2017 - November 15, 2017
Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb,
David. "Identifying and Quantifying the Resilience Dividend using Computable
General Equilibrium Models: A Methodological Overview." Paper presented at
2nd International workshop on Modelling of Physical, Economic and Social
Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December
15, 2017
Ak, Ronay; Ferguson, Max; Law, Kincho; Lee, Yung-Tsun. "Automatic
Localization of Casting Defects with Convolutional Neural Networks." Paper
presented at 2017 IEEE International Conference on Big Data (BigData 2017),
Boston, MA, United States. December 11, 2017 - December 14, 2017 SP-318
Andres Valiente, Blanca; Hoehler, Matthew. "Influence of Fire on the Seismic
Shear Capacity of Cold-Formed Steel Shear Walls." Paper presented at Eleventh U.
S. National Conference on Earthquake Engineering, Los Angeles, CA, United
States. June 25, 2018 - June 29, 2018
Butry, David; Fung, Juan; McCabe, Steven; Sattar, Siamak. "Cost Estimates for the
Seismic Retrofit of Federally Owned and Leased Buildings." Paper presented at
Eleventh US National Conference on Earthquake Engineering, Los Angeles, CA,
United States. June 25, 2018 - June 29, 2018
Kamarthi, Sagar; Lee, Yung-Tsun; Radhakrishnan, Srinivasan. "Estimation of
Online Tool Wear in Turning Processes Using Recurrence Quantification Analysis
(RQA)." Paper presented at 2017 IEEE Big Data, Boston, MA, United States.
December 11, 2017 - December 14, 2017 SP-342
Bentz, Dale; Cusatis, Gianluca; Di Luzio, Giovanni; Masoero, Enrico; Pathirage,
Madura. "A multiscale framework for the prediction of concrete self-desiccation."
Paper presented at EURO-C 2018 Conference - Computational Modelling of
Concrete and Concrete Structures, Bad Hofgastein, Austria. February 26, 2018 -
March 1, 2018
Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental
Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at
Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth,

TX, United States. April 19, 2018 - April 21, 2018	SP-352
Choe, Lisa; Gross, John; Grosshandler, William; Hoehler, Matthew; Ramesh,	
Selvarajah. "Design and Construction of Long-Span Composite Beam Specimens	
for Large Structural-Fire Tests." Paper presented at Structures Congress 2018, Fort	
Worth, TX, United States. April 19, 2018 - April 21, 2018	SP-365
Bajaj, Manas; Hedberg, Thomas. "System Lifecycle Handler Spinning a Digital	
Thread for Manufacturing." Paper presented at 28th Annual INCOSE International	
Symposium, Washington, DC, United States. July 7, 2018 - July 12, 2018	SP-378
Anand, Dhananjay; Moyne, James; Tang, CheeYee; Tilbury, Dawn; Wang,	
Zheng. "Conflict-driven Hybrid Observer-based Anomaly Detection." Paper	
presented at American Control Conference, Milwaukee, WI, United States. July 2,	
2018 - July 6, 2018	SP-393
Bostelman, Roger; Foufou, Sebti; Hong, Tsai. "Modeling Performance	
Measurement of Mobile Manipulators." Paper presented at 7th Annual IEEE Int.	
Conf. on CYBER Technology in Automation, Control, and Intelligent Systems	
(IEEE-CYBER 2017), Honolulu, HI, United States. July 31, 2017 - August 4,	
2017	SP-401
Brundage, Michael; Hoffman, Michael; Moccozet, Sascha; Morris, Katherine;	
Sexton, Thurston. "Developing Maintenance Key Performance Indicators from	
Maintenance Work Order Data." Paper presented at 46th SME North American	
Manufacturing Research Conference (NAMRC) and the ASME Manufacturing	
Science and Engineering Conference (MSEC 2018), College Station, TX, United	
States. June 18, 2018 - June 22, 2018	SP-407
Qiao, Guixiu; Weiss, Brian. "Monitoring, Diagnostics, and Prognostics for Robot	
Tool Center Accuracy Degradation." Paper presented at ASME International	
Manufacturing Science and Engineering Conference, College Station, TX, United	ing Conference, College Station, TX, United
States. June 18, 2018 - June 22, 2018	SP-416
Klinger, Alexander; Weiss, Brian. "Robotic Work Cell Test Bed to Support	
Measurement Science for Monitoring, Diagnostics, and Prognostics ." Paper	
presented at ASME Manufacturing Science and Engineering Conference, College	
Station, TX, United States. June 18, 2018 - June 22, 2018	SP-425

Fox, Jason; Gnaupel-Herold, Thomas; Heigel, Jarred; Phan, Thien. "Experimental	
Investigation of Residual Stress and its Impact on Machining in Hybrid Additive/	
Subtractive Manufacturing." Paper presented at 46th SME North American	
Manufacturing Research Conference, NAMRC 46, College Station, TX, United	
States. June 18, 2018 - June 22, 2018	;
Brignoli, Riccardo; Brown, J; Domanski, Piotr; Kazakov, Andrei; McLinden,	
Mark. "Low-GWP Refrigerants: Options and Issues." Paper presented at Int.	
Conference on Cryogenics and Refrigeration 2018, April. 12-14, 2018, Shanghai.,	
Shanghai, China. April 12, 2018 - April 14, 2018	,
Hedberg, Thomas; Helu, Moneer; Sprock, Timothy. "A Standards and Technology	
Roadmap for Scalable Distributed Manufacturing Systems." Paper presented at	
2018 ASME Manufacturing Science and Engineering Conference (MSEC 2018),	
College Station, TX, United States. June 18, 2018 - June 22, 2018)
Gomez, Michael; Heigel, Jarred; Schmitz, Tony. "Force modeling for hybrid	
manufacturing." Paper presented at 46th SME North American Manufacturing	
Research Conference, NAMRC 46, College Station, TX, United States. June 18,	
2018 - June 22, 2018	;
Kim, Hyojin; Ng, Lisa; Nguyen, Khiem; Payne, William. "Thermal Performance	
Evaluation of a Small Duct High Velocity System Using NIST's Net-Zero Energy	
Residential Test Facility." Paper presented at Indoor Air 2018, Philadelphia, PA,	
United States. July 22, 2018 - July 27, 2018)
Burns, Martin; Roth, Thomas. "A Gateway to Easily Integrate Simulation	
Platforms for Co-Simulation of Cyber-Physical Systems." Paper presented at 2018	
Workshop on Modeling and Simulation of Cyber-Physical Energy Systems, Porto,	
Portugal. April 10, 2018 - April 10, 2018)
Babushok, Valeri; Linteris, Gregory. "Predicted Burning Velocities of C1 and C2	
Hydrofluorocarbon Refrigerant Flames with Air." Paper presented at Spring	
Technical Meeting? Eastern States Section of the Combustion Institute, Station	
College, PA, United States. March 4, 2018 - March 7, 2018 SP-476)
Sauca, Ana; Seif, Mina; Zhang, Chao. "Stability of Steel Structures at Elevated	
Temperatures: A Hybrid Fire Testing Approach." Paper presented at Annual	
Stability Conference / NASCC: The Steel Conference, Baltimore, MD, United	

States. April 10, 2018 - April 13, 2018	SP-482
Gong, Mengyan; Poppendieck, Dustin. "Calibration of solid-phase microextraction	
for measuring indoor gas phase SVOC concentration." Paper presented at Indoor	
Air Conference, Philadelphia, PA, United States. July 22, 2018 - July 27, 2018	SP-501
Bagchi, Shelly; Marvel, Jeremy. "Towards Augmented Reality Interfaces for	
Human-Robot Interaction in Manufacturing Environments." Paper presented at	
Workshop on Virtual, Augmented and Mixed Reality for Human-Robot	
Interaction, Chicago, IL, United States. March 5, 2018 - March 5, 2018	SP-503
Hoehler, Matthew; Smith, Christopher. "Application of Digital Image Correlation	
to Structures in Fire." Paper presented at Society of Experimental Mechanics	
Annual Meeting, Greenville, SC, United States. June 4, 2018 - June 7, 2018	SP-506
Manzello, Samuel; Naruse, Tomohiro; Suzuki, Sayaka. "Study on Firebrand	
Generation From Roofing Assembly Combustion." Paper presented at Japan	
Association for Fire Science and Engineering, , Japan. May 29, 2018 - May 30,	
2018	SP-509
Manzello, Samuel; Suzuki, Sayaka. "Experimental Study on Ignitions of Thatched	
Roofing Assemblies Exposed From Firebrands." Paper presented at Japan	
Association for Fire Science and Engineering, , Japan. May 29, 2018 - May 30,	
2018	SP-511
Haapala, Karl; Morris, Katherine; Shankar Raman, Arvind. "Towards a standards-	
based methodology for extending manufacturing process models for sustainability	
assessment." Paper presented at 46th SME North American Manufacturing	
Research Conference (NAMRC) and the ASME Manufacturing Science and	
Engineering Conference (MSEC 2018), College Station, TX, United States. June	
18, 2018 - June 22, 2018	SP-513
Bhattacharyya, Shuvra; Candell, Richard; Hany, Mohamed; Jeng, Ging; Li,	
Honglei; Liu, Yanzhou; Liu, Yongkang. "Model-based Cosimulation for Industrial	
Wireless Networks." Paper presented at 14th IEEE International Workshop on	
Factory Communication Systems, Imperia, Italy. June 12, 2018 - June 15, 2018	SP-523
Andres Valiente, Blanca; Hoehler, Matthew. "Influence of Fire on the Shear	

Capacity of Steel-Sheathed Cold-Formed Steel Framed Shear Walls." Paper

presented at SiF18 Ã,â " The 10th International Conference on Structures in Fire, Belfast, United Kingdom. June 6, 2018 - June 8, 2018
Brandon, Daniel; Bundy, Matthew; Hoehler, Matthew; Kimball, Amanda;
Lafrance, Pier-Simon; Ostman, Birgit; Su, Joseph. "Fire Safety Challenges of Tall
Wood Buildings: Large-Scale Cross Laminated Timber Compartment Fire Tests."
Paper presented at SiF18 Ã, â " The 10th International Conference on Structures in
Fire, Belfast, United Kingdom. June 6, 2018 - June 8, 2018 SP-541
Chernovsky, Artur; Sauca, Ana; Seif, Mina; Zhang, Chao. "Hybrid Fire Testing of
a Single Degree-of-Freedom Linear System." Paper presented at SiF18 Ã,â " The
10th International Conference on Structures in Fire, Belfast, United Kingdom. June
6, 2018 - June 8, 2018
Babushok, Valeri; Linteris, Gregory. "Numerically-Predicted Velocities of C1 and
C2 Hydrofluorocarbon Refrigerant Flames with Air." Paper presented at 17th
International Refrigeration and Air Conditioning Conference at Purdue, West
Lafayette, IN, United States. July 9, 2018 - July 12, 2018 SP-557
Bundy, Matthew; Choe, Lisa; Gross, John; Grosshandler, William; Hoehler,
Matthew; Ramesh, Selvarajah; Seif, Mina. "Fire Performance of Long-Span
Composite Beams with Gravity Connections." Paper presented at SiF18 - The 10th
International Conference on Structures in Fire, Belfast, United Kingdom. June 6,
2018 - June 8, 2018
Brandon, Daniel; Hoehler, Matthew; Kimball, Amanda; Ostman, Birgit; Schmid,
Joachim; Su, Joseph. "Experimental Fire-Simulator For Post-Flashover
Compartment Fires." Paper presented at SiF18 - The 10th International Conference
on Structures in Fire, Belfast, United Kingdom. June 6, 2018 - June 8, 2018 SP-575
Bell, Ian; Linteris, Gregory; McLinden, Mark. "An Empirical Model for
Refrigerant Flammability Based on Molecular Structure and Thermodynamics."
Paper presented at 17th International Refrigeration and Air Conditioning
Conference at Purdue, West Lafayette, IN, United States. July 9, 2018 - July 12,
2018
Kiritsis, Dimitris; Kulvatunyou, Boonserm; Smith, Barry; Wallace, Evan; Will,
Chris. "The Industrial Ontologies Foundry Proof-of-Concept Project." Paper

presented at APMS 2018 International Conference Advances in Production

Management Systems (APMS 2018), Seoul, Republic of Korea. August 25, 2018 - August 30, 2018
Benmohamed, Lotfi; Candell, Richard; Hany, Mohamed; Liu, Yongkang. "
Dimensioning Wireless Use Cases in Industrial Internet of Things." Paper
presented at 14th IEEE International Workshop on Factory Communication
Systems, Imperia, Italy. June 12, 2018 - June 15, 2018
Jameson, Jordan; Vogl, Gregory. "Comparative Analysis of Bearing Health
Monitoring Methods for Machine Tool Linear Axes." Paper presented at Society
for Machinery Failure Prevention Technology 2018 Annual Conference, Virginia
Beach, VA, United States. May 15, 2018 - May 17, 2018SP-604
Choi, Sangsu; Kulvatunyou, Boonserm; Wuest, Thorsten. "Towards a Platform for
Smart Manufacturing Improvement Planning." Paper presented at APMS 2018
International Conference Advances in Production Management Systems (APMS
2018), Seoul, Republic of Korea. August 26, 2018 - August 30, 2018
Stutzman, Paul. "Reference Materials and Proficiency Testing for Validating X-
Ray Powder Diffraction Measurements of Cement Composition." Paper presented
at 20th International Conference of Building Materials, Weimar, Germany.
September 12, 2018 - September 14, 2018 SP-628
Heigel, Jarred; Whitenton, Eric. "Measurement of Thermal Processing Variability
in Powder Bed Fusion." Paper presented at 2018 ASPE and euspen Summer
Topical Meeting - Advancing Precision in Additive Manufacturing, Berkeley, CA,
United States. July 22, 2018 - July 25, 2018
Evans, Christopher; Fox, Jason; Kim, Felix; Reese, Zachary. "Complementary Use
of Optical Metrology and X-Ray Computed Tomography for Surface Finish and
Defect Detection in Laser Powder Bed Fusion Additive Manufacturing." Paper
presented at 2018 ASPE and euspen Summer Topical Meeting - Advancing
Precision in Additive Manufacturing, Berkeley, CA, United States. July 22, 2018 -
July 25, 2018
Evans, Christopher; Fox, Jason; Reese, Zachary; Taylor, John. "Evolution of
Cooling Length in Parts Created Through Laser Powder Bed Fusion Additive
Manufacturing." Paper presented at 2018 ASPE and euspen Summer Topical
Meeting - Advancing Precision in Additive Manufacturing, Berkeley, CA, United

States. July 22, 2018 - July 25, 2018
Evans, Christopher; Fox, Jason; Kim, Felix; Reese, Zachary; Taylor, John. "Effect
of Subsurface Defects on the Surface Topography of Additively Manufactured
Components." Paper presented at 2018 ASPE and euspen Summer Topical
Meeting - Advancing Precision in Additive Manufacturing, Berkeley, CA, United
States. July 22, 2018 - July 25, 2018
Klinger, Alexander; Weiss, Brian. "Examining Workcell Kinematic Chains to
Identify Sources of Positioning Degradation." Paper presented at Annual
Conference of the Prognostics and Health Management Society 2018,
Philadelphia, PA, United States. September 24, 2018 - September 27, 2018SP-66-
Bruns, Morgan; Cleary, Thomas; Gilbert, Stanley; Peacock, Richard; Reneke,
Paul. "Development of Tools to Evaluate the Impact of New Residential Fire
Protection Technology." Paper presented at SUPDET 2018, Cary, NC, United
States. September 11, 2018 - September 13, 2018

COMPUTATIONAL ENVIRONMENT FOR MODELING AND ENHANCING COMMUNITY RESILIENCE: INTRODUCING THE CENTER FOR RISK-BASED COMMUNITY RESILIENCE PLANNING

John W. van de Lindt^{1,*}, Bruce R. Ellingwood², Therese McAllister³, Paolo Gardoni⁴, Daniel T. Cox⁵, Harvey Cutler⁶, and Walter Gillis Peacock⁷

¹Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523-1372, USA.

²Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523-1372, USA.

³Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA

⁴Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

⁵School of Civil and Constructional Engineering, Oregon State University, Graf Hall, Corvallis, OR 97331, USA

⁶Department of Economics, Colorado State University, Fort Collins, CO 80523-1771, USA

⁷Hazard Reduction and Recovery Center and Department of Landscape Architecture and Urban Planning, Texas A & M University, 1372 TAMU, College Station, TX 77843, USA

ABSTRACT

The resilience of a community is defined as its ability to prepare for, withstand, recover from and adapt to the effects of natural or human-caused disasters, and depends on the performance of the built environment and on supporting social, economic and public institutions that are essential for immediate response and long-term recovery and adaptation. The performance of the built environment generally is governed by codes, standards, and regulations, which are applicable to individual facilities and residences, are based on different performance criteria, and do not account for the interdependence of buildings, transportation, utilities and other infrastructure sectors. The National Institute of Standards and Technology recently awarded a new Center of Excellence (NIST-CoE) for Risk-Based Community Resilience Planning, which is headquartered at Colorado State University and involves nine additional universities. Research in this Center is focusing on three major research thrusts: (1) developing the NIST-Community Resilience Modeling Environment known as NIST-CORE, thereby enabling alternative strategies to enhance community resilience to be measured quantitatively; (2) developing a standardized data ontology, robust data architecture and data management tools in support of NIST-CORE; and (3) performing a comprehensive set of hindcasts on disasters to validate the data architecture and NIST-CORE.

KEYWORDS: Community resilience; hazards; investment optimization; post-disaster recovery; resilience performance metrics; risk-informed decision.

INTRODUCTION AND MOTIVATION

Disaster resilience has been defined many ways, but one major commonality exists in virtually all definitions: the ability to rebound following a shock or major disruption. Presidential Policy Directive (PPD) 8: National Preparedness (2011) defines resilience as the ability to "adapt to changing conditions and withstand and rapidly recover from disruption due to emergencies." This definition was expanded in Presidential Policy Directive (PPD) 21 (2013) which defines resilience as the ability to "prepare for and adapt to changing conditions and withstand and recover rapidly from disruptions." The ability of a community to recover from a disaster is a function of many factors including pre-event planning, preparation, mitigation, infrastructure type, complexity, condition, individual and collective experiences prior to the disaster and the ability to mobilize resources

afterward. Prior to the disaster, a community may contain a heterogeneous mix of populations and institutions of varying degrees of vulnerability. Following the occurrence of an extreme event, there is a rapid drop in community functional capacity, followed by a period of response and recovery, leading to a "new normal" state for the community. For example, the most vulnerable communities may have a diminished "new normal," in which the community functions at a level that is below where they were prior to the event because they were only able to build back based on pre-impact or diminished capabilities; an example of this is the City of New Orleans, Louisiana, USA shown flooded in Fig 1a following Hurricane Katrina in 2005, particularly in the lower 9th ward. Conversely, communities may achieve a "new normal," in which their community functions at a higher level following the event, in what has been termed a "build back better" scenario. An example of this is Greensburg, Kansas, USA following the 2007 EF5 (> 320 kph winds) tornado, in which 95 % of the city was destroyed as shown in Fig 1b.



Figure 1 (a) New Orleans, LA during flooding following hurricane Katrina in 2005 (Ed Levine, NOAA); (b) Greensburg, KS following the EF 5 tornado in 2007 which destroyed 95 % of the city (Greg Henshall, FEMA)

The performance of the built environment in the United States, which is a key factor in community resilience, is largely determined by codes and standards, which are applicable to individual facilities and residential structures, and have the primary objective of preserving life safety under severe events. Current codes do not address the period of recovery following an event. Moreover, design of interdependent transportation systems, utilities and communication systems currently is based on different performance criteria. As a consequence, there currently is no assurance that all systems required for community resilience will perform at a necessary or consistent level following an extreme event. Uncertainties in the demands placed on a community by extreme natural hazards and in the response of community infrastructure to withstand those demands are very large. Furthermore, science-based measurement tools to evaluate performance and resilience at the community scale, fully integrated supporting databases, and risk-informed decision frameworks to support optimal life-cycle technical and social policies aimed at enhancing community resilience do not exist. The NIST-CoE will focus on creating these science-based measurement tools over the next five years. These tools will allow researchers and planners to assess strategies quantitatively for optimizing resiliency subject to a community's technical, economic and social constraints. The risk-informed decision framework and unique set of science-based measurement tools will help communities to better assess, visualize and potentially achieve resilience goals while managing lifecycle costs, thus making it possible to establish, for the first time, a business case agreed upon by a communities multiple constituencies, for achieving community resilience.

OVERVIEW OF THE CENTER FOR RISK-BASED COMMUNITY RESILIENCE PLANNING

The NIST-CoE at Colorado State University, funded by the National Institute of Standards and Technology (NIST), will develop systems-level models and databases that will provide the technology for enhancing community resilience in a research and development program. In addition to faculty at CSU, the team members include noted experts in resilience from the California Polytechnic University Pomona, Oregon State University, Rice University, Texas A&M University, Texas A&M University-Kingsville, the University of Illinois at Urbana-Champaign, the University of Oklahoma, the University of South Alabama, and the University of Washington. The Center's research program is organized along three major thrust areas.

Thrust 1 will develop a multidisciplinary computational environment with integrated supporting databases, known as NIST-CORE (NIST-Community Resilience Modeling Environment), that will enable the factors (and their inter-relationships) that determine community resilience to be better understood. This environment is being implemented as a risk-informed decision framework that permits the effectiveness of alternative strategies for addressing resilience to be measured quantitatively. A distinctive feature of Thrust 1 is its emphasis on multiple hazards and inter-dependent physical systems, which may have significant cascading effects. This will include consistent handling of and propagation of uncertainties throughout the analysis. Furthermore, nontechnical systems that are essential for the recovery and vitality of a community - housing, businesses, healthcare, schools, and civic organizations - are being integrated into NIST-CORE, creating a nexus between social and technological infrastructure networks intended to narrow the gap between engineering and social science aspects of resilience planning and to facilitate risk communication among stakeholders. Finally, optimization strategies for enhancing community resilience involving advanced intelligent decision algorithms are being integrated, and are based on performance metrics such as time to recovery, number of mortalities, number of morbidities, and other metrics to be identified.

Thrust 2 will produce a CoE standardized data ontology, a robust data architecture, and effective data management. This includes tools to support the computational environment developed in Thrust 1 and to permit databases from stakeholders representing multiple domains of engineering and social sciences to be integrated to the decision process. The data architecture and data management tools in NIST-CORE will be customized to help users incorporate, manage, query, visualize and share data.

Finally, Thrust 3 will validate the resilience data architecture through a series of testbeds that stress the process of data collection, its integration into the computational modeling environment, and decision algorithms. The NIST-CoE research thrusts seek to align with those of NIST, with synergies and opportunities for collaborative research, staff exchanges, development of partnerships with the private sector, and for participating in collaborative field studies that can be used to validate and improve the computational environment and its supporting databases.

THRUST 1 - COMPUTATIONAL MODELING ENVIRONMENT FOR COMMUNITY RESILIENCE: DEVELOPMENT OF NIST-CORE

The NIST-CORE architecture is depicted in Fig 2. When fully developed, NIST-CORE will have the capability of computing proposed resiliency assessment measures at the user-specified community level (described to the right in Fig 2) and utilizing the data architecture (summarized at the bottom of Fig 2), thereby producing performance metrics, performance requirements, and risk-informed decision tools for community resilience. This section addresses some of the immediate challenges that Center investigators are addressing.



Figure 2 The architecture of the NIST Community Resilience (NIST-CORE) Modelling Environment

Hazards and Climate Change

The treatment of hazards, including earthquake, earthquake aftershock, coastal and riverine flooding, tornado including wind and debris, hurricane wind and debris, hurricane surge and wave, wildfires and their interface with urban areas, tsunami, and landslide, is a major part of the NIST-COE. The risk formulation for some

individual hazards is reasonably mature; for other hazards, there is a need for improved characterization of intensity and damage potential for the computational resilience modeling environment. For example, wind speeds resulting from tornadoes are not well developed but have been recently updated by, for example, Standohar-Alfano and van de Lindt (2014) by combining earlier methodologies. Relative to seismic and wind hazards, the hazards due to surge, waves, and water-borne debris for hurricanes and severe coastal storms have been identified as a critical research gap (NIST 2014), not only for characterizing the probable inundation levels and wave conditions but also for identifying the joint distributions of wind and wave effects for multi-hazards scenarios. Moreover, recent research (Tamiczek et al 2014) following Hurricane Ike has shown that water depth alone is often a poor predictor of damage to structures. Each of these cases requires special consideration with regard to the specific multi-hazard models for: a) multi-hazard characterization, b) modeling infrastructure fragilities when subjected to concurrent or time lag events; and c) restoration modeling.

In addition to consideration of hazards, the NIST-CoE will include methods to estimate the effect of climate change on hazards when applying them within NIST-CORE. During the past three decades, evidence has grown that global climate change may affect both the frequency and severity of extreme events resulting from natural hazards (e.g., IPCC 2013). The severity of extreme climatic events is characterized by changes in the climate variables such as wind speed, precipitation, temperature or flooding. However, practically all risk assessments for civil infrastructure to date have been based on the assumption that the hazard-imposed demands on the system can be modeled as stationary random processes. This assumption becomes untenable when the effects of global climate change are considered (Lee and Ellingwood 2013). Stochastic models of natural hazard demands from hurricanes, surge and coastal flooding that incorporate nonstationarity due to changes in sea surface temperatures and other manifestations of climate change are also being developed.

Modelling the Built Environment

Five key components of the built environment are being modeled, including buildings, transportation infrastructure, water and wastewater networks, energy systems, and telecommunication networks. Performance objectives for individual buildings are required to enable a building inventory to support community resilience goals (McAllister 2013). For example, the SPUR Program in San Francisco (Poland 2011) established a set of performance objectives for buildings (e.g., hospitals, school, emergency operation centers, residences, businesses) in different performance categories exposed to different earthquake intensities (i.e. routine, expected, and extreme) and with different functionality needs (i.e., immediate recovery within 72 hrs., short-term recovery within 60 days, and long-term reconstruction).

While the SPUR Program provides an example of building performance criteria linked to community resilience planning for one hazard (earthquake), more general metrics, criteria and guidelines for building performance that are based on measurement science are required for disaster planning for communities with different hazard exposures, social needs and resources. Resilience assessment of buildings, as implemented by structural engineers for extreme events, has focused on minimizing direct and indirect losses through enhanced system robustness as well as more effective recovery strategies (e.g., Bruneau, et al. 2003).

Transportation infrastructure is comprised of a number of multi-modal systems, such as bridge and roadway networks, heavy and light rail, airports and port and maritime systems. Common themes for characterizing transportation system resilience include the need to assess component reliability under multiple hazards, model connectivity within the network, evaluate the recovery trajectory and quantify the associated impacts on community functions. A significant body of work exists on modeling the fragility of transportation infrastructure to aging, with emphasis placed on resilience assessment of bridge networks (e.g., Bocchini & Frangopol 2013), but with only limited work on other modes of transportation. In the transportation network developed within the NIST-CoE, temporal effects (e.g., aging, time lag in multiple events, sequence of restoration) will be captured along with spatial considerations across a regionally distributed network (e.g., damage correlations).

Water systems are regarded as a "lifeline" infrastructure crucial for minimizing the societal impact of extreme events. These networks are large, distributed interdependent systems and their failure can be significant, since residential, commercial, industrial, and other users depend on the network: supply (ground and surface), transmission, treatment, pumping, and storage.

Electric power delivery systems are comprised of generation, transmission and distribution subsystems and are central to the proper functioning and operation of most other infrastructure and again critical for social systems in general and business survival in particular. Thus, they have been examined in greater detail than have natural gas, liquid fuel, or geothermal systems, and models of resilience for electric power delivery networks have been

developed. Most of these models focus on the resilience of the physical systems comprised of generation facilities, transmission towers, substations and transformers, poles and towers that comprise the bulk of the distribution lines.

Innovative approaches to modeling resilience and interdependency include agent-based modeling, input-output models, mathematical models and game theory. Physics-based models (as opposed to regressive models) for assessing resilience and reliability of electric power delivery include system-level fragility functions based upon multiple hazards, expanded system reliability indices as defined by IEEE 1366 (2012); post-event restoration models; functionality and inoperability models with associated metrics of "vulnerability", "rapidity" and "recovery" and mechanical analog single degree of freedom system models used to characterize the properties of the underlying system and to determine how these properties may be used minimize outage times (e.g., Reed et al. 2009).

Modelling Economic and Social Systems and Networks

Economic models can inform a portfolio of risk management investment decisions, including mitigation, warning and evacuation systems, as well as recovery and reconstruction efforts (Gilbert 2010). In short, these models can show how investing or failing to invest in resilience-enhancing interventions generate direct and cascading effects in economic systems that can alter a regional (or community) economy's long-term trajectory. Two complementary economic impact modeling approaches are being utilized within the CoE to estimate the direct and multiplier effects of assorted disaster shocks. First, applied econometric models examine community resilience across a variety of economic outcomes (e.g., gross regional product, government revenue, employment growth, wages, births and deaths, and migration flows) over hazard events of varying intensity (i.e., fatalities, injuries, destruction of capital stock), given varying regional attributes and levels of investment in hazard resilience. Specifically, the objective is to quantify the economic impacts of disasters in the US over the last 50 years to determine the level of resilience that has been achieved across regions. Second, computable general equilibrium (CGE) models capture economic dynamics during and after major shocks, analyzing how hazard losses manifest themselves in the economy of a region through industry-specific losses in capital stock, damage to critical infrastructure, and human capital stock deterioration through out-migration (Rose 2009). These models will be integrated into the intelligent decision algorithms for optimizing investments in community resilience.

The impacts of disasters on human populations (mortality, morbidity, and psychological) and social systems (disruption, displacement, failure, change), while set into motion by the particular disaster agent, are nevertheless heavily influenced by pre-existing inequalities related to physical vulnerabilities in the built and natural environment and social vulnerabilities generated within economic and social systems. Linking these physical and social vulnerabilities is key to understanding community disaster resilience. Socially vulnerable populations within communities that are residing in hazard-prone locations and vulnerable structures before a disaster often are most likely to experience disproportionate losses, higher damage rates, and housing losses (Highfield et al. 2014). Loss of housing resulting in short- and long-term displacement, jeopardizes employment, business survival, food security, educational attainment, and access to transportation and local support networks, thus placing households in even more precarious and difficult recovery trajectories (Peacock et al 2014). Variations in socio-economics results in clusters of households and businesses facing similar direct and cascading effects, leading to increased probabilities of local business disruption and subsequent failure. The ability to model cascading effects requires that these preexisting social vulnerabilities of populations and social systems characterized in terms of age, race and ethnicity, gender, income/poverty, including relative access to potentially scarce resources such as health care, food, education, employment and housing will be modeled (Van Zandt et al 2012). The way in which hazards losses manifest and cascade among these populations and social systems, such as the dependencies between housing loss, population displacement and local business and neighborhood disruption (Xiao and Van Zandt, 2013) is being modeled for synthesis into NIST-CORE.

Treatment of Interdependency, Aging Infrastructure, and Uncertainties

Community resilience is highly affected by the mutually interdependent nature of buildings, transportation, water and wastewater, power, and communications (Reed 2009). Interdependency is defined herein as "the multi or bi-directional reliance of an asset, system, network, or collection thereof, within or across sectors, on input, interaction, or other requirement from other sources in order to function properly" (CIKR 2008). Most available approaches to modeling interdependency only support a portion of the simulation (i.e., response phase, failure cascading phase, or recovery phase) involved in the resilience analysis (Ouyang 2014). Within the NIST-CoE a combination of methods to achieve a uniform interdependency modeling framework in both time

and geographic scale for the overall resilience analysis in NIST-CORE will be used. Data obtained from remote sensing, satellite tracking, and geographical information systems at various scales will be integrated into NIST-CORE, which will include a combination of topological approaches, network flow models, and empirical methods. The data-rich history which currently exists (and will be continually updated and housed within NIST-CORE) in empirical methods and statistical learning will inform the topological representation and thereby improve the accuracy of the interdependencies captured by the model.

Physical infrastructure facilities may deteriorate due to exposure to extreme conditions (e.g., excessive loading or harsh environments), weathering, routine use and accidents. The past two decades have seen considerable research on time-dependent reliability of deteriorating structures (e.g., Mori and Ellingwood 1993; Frangopol, et al. 2004). To model the life-cycle of a facility, including time to failure under random occurrences of loads or number of loads until failure occurs (e.g., operational and extreme loads), stochastic models of the deterioration processes and the dependencies between these processes will be included in NIST-CORE. The models will utilize a novel framework that accounts for 1) both shocks and a gradual deterioration process; 2) the effect of deterioration on both capacity of and demand on infrastructure components; and 3) the possibility of both types of failures. The stochastic framework will be used in a novel renewal theory-based life-cycle analysis (RTLCA) model for deteriorating infrastructure components, which accounts for both functional and ultimate failures.

Significant uncertainties exist in every phase of community resilience evaluation - in the condition of the built environment prior to the occurrence of the extreme event, in the loss of functionality immediately after the event, and in the time and degree of post-disaster recovery. Essential components of uncertainty analysis include physical system-based uncertainties that mainly affect the community robustness (e.g., stemming from performance of individual facilities, building portfolios and infrastructure systems as a whole) and socioeconomic uncertainties that affect the extent of societal impact and community recovery (e.g., those associated with social, economic and political institutions within the community, human responses and available external resources.) Such socioeconomic uncertainties will be considered (Murphy et al 2011).

Resilience assessment methods developed within the last decade do not address the spatial and temporal correlations in the response of building portfolios and infrastructure systems as an integral part of the community resilience; nor do they address the uncertainties in the recovery trajectory, which depend on the residual strength of key community institutions following the disaster. NIST-CORE will incorporate simulation-based algorithms for propagating uncertainties in individual and competing natural hazards, in facility response and capacity, and in structural modeling through the risk analysis. The uncertainties in the effect of aging on capacity and demand and the effect of climate change on the intensity/frequency of natural hazards will also be addressed. Coupling and linkage of engineering and social science models of displacement and housing recovery, for example, introduce additional issues in the modeling of uncertainties that will be addressed within NIST-CORE.

THRUST 2 - DATA STANDARDIZATION, USER REOUIREMENTS, AND MANAGEMENT TOOLS

Thrust 2 will develop a standardized data ontology, a robust data architecture and effective data management tools to support the computational environment. Thrust 1 encompasses multiple domains of science and it is therefore critical to first understand all user needs prior to data standardization. Existing ontology, data types and data formats for communities and stakeholders will be integrated into the data standards for NIST-CORE. Finally, the data architecture and data management tools will be customized to help users ingest, manage, query, visualize and share data effectively.

The existing ontology, data types, and data formats identified for the communities and stakeholders involved will be reviewed and integrated into the standards for NIST-CoE. For example, an existing earthquake loss modeling environment may have approximately 200 data types and about 10 data formats for seismic risk assessment; thus it is envisioned that as many as 500 to more than 1000 data types may be required in NIST-CORE. These will be reviewed and integrated according to the user requirements. This will utilize the extension of GML (Geospatial Markup Language) to define the data types and metadata of the datasets.

The core component of the data architecture will be the NIST-CORE data middleware layer, the place where the processing/curation of data occurs. In other words, the raw data, such as hazards, fragility curves for infrastructure, structure inventory, utility network, socio-economic data, etc., will be curated according to the data ontology, data types and data formats developed as described above. This layer has four components: 1) metadata extraction, 2) format conversion, 3) user curation, and 4) data publishing. The metadata extraction component automatically captures/processes the metadata from the raw data according to the data ontology. The

format conversion provides the conversion service of the data format with information about data loss from the conversion. The user curation component then allows users to finalize the curation process with extracted metadata. Finally, the curated data will be published to the NDS. The curated data from the middleware layer will be stored in distributed data repositories. The data repositories will have secure access control allowing user(s) to have private repositories, if needed for security reasons.

Defining Resiliency Baselines, Resilience Metrics for Recovery, and Performance Improvements in Resiliency

A critical element in the impact-recovery model is the recognition that resiliency planning and actions can have direct consequences on physical impacts and social consequences, as well as influencing the nature and speed of recovery. The adoption of comprehensive mitigation policies (e.g., building codes and effective land-use planning) can reduce physical vulnerabilities and exposure, ensuring that buildings are kept out of high hazard areas or that buildings constructed there are better able to meet potential risks. Recovery planning can help ensure that rebuilding and repair lessens preexisting vulnerabilities, expedites debris removal, ensures capitalization of post disaster public sector activities, addresses vulnerable population needs and lessens postdisaster risks. The NIST-CORE environment, as a comprehensive community decision support tool, will address broad based resilience planning activities by helping communities understand 1) their baseline levels of resilience in relation to existing vulnerabilities and potential impacts, 2) associated recovery outcomes, and 3) resilience enhancements through the adoption and implementation of policies during recovery or as a function of resilience planning.

To capture direct and cascading impacts, the linkage between direct physical damage to the built environment (buildings and infrastructure) and broader socio-economic impacts will be refined to capture and specify damage to housing in its various forms (owner vs rental, single family, duplexes, multi-family and concomitant loss of housing units), to businesses, and to critical facilities. The focus is initially translating direct physical damage to structures and infrastructure systems to the associated disruptions caused by dependencies of social systems on the built environment. Additionally, social impacts and cascading effects for a set of hazards of varying levels and intensities will be modeled. Resilience metrics will assess impacts in the form of the loss of different forms of housing, loss and disruption of various forms of businesses, critical facilities (health, and institutional entities (hospitals, education, childcare, etc.), These in turn will be employed to model further social consequences related to population dislocations, food and health security, etc. The consequences of disaster impact for each sector cannot be defined based on a single metric (damage state(s), life safety, economic loss, etc.) or value due to differences in initial conditions, community expectations, and unequal consequences. Thus, within the NIST-CoE, an impact-performance matrix will be developed for building, infrastructure systems, and social systems, for a number of hazards.

While recovery is still one of the most understudied areas in disaster research (NRC 2006; Peacock et al. 2008; McAllister 2013), recent years have seen new research on long-term recovery and modeling approaches based on empirical and expert knowledge (e.g., Chang et al. 2014; Peacock et al, 2014). Drawing on these findings and research undertaken by the NIST team, the CoE is developing a series of algorithms that will model recovery trajectories for infrastructure, housing, businesses, and populations. They will consider not only technical aspects of repair, replacement, lifecycle and interdependence, but also broader socio-economic factors and policies that can influence recovery. The results will provide recovery-timing assessment for various sectors and sub-populations, the overall community, and a reassessment of vulnerabilities and risk given the new post disaster end-state. A goal will be to parameterize models assessing recovery under scenarios such as expedited recovery without policy changes or policy driven recovery based on recovery and mitigation planning policies.

As part of comprehensive resiliency planning, activities stakeholders may undertake performance improvement analysis as part of a recovery scenario or simply to reflect the adoption and implementation of new resiliency policies by their community or sector. The NIST-CORE environment will model performance changes (both increases and decreases) given changes or modifications in codes, standards, and policies relative to the baseline resiliency impact-performance matrix. It will be critical that the modeling environment be flexible to adapt to user, stakeholder, and community defined scenarios to allow communities to consider alternative solutions and various scenarios, such as business as usual, progressive intervention, or some other scenario. Furthermore, a long-term goal will be forecasting scenarios capturing what-if-outcomes given specific policy changes and howto scenarios whereby stakeholders establish a resilience goal (i.e., reduced damage probabilities or recovery rates) and the model will offer optimal resiliency changes (mitigation investments, policies, etc.) needed to reach goals. This will allow users to understand the effect of changing one performance metric in the grand

scheme and allow performance metrics to be refined and selected. To test this approach, numerous simulations and sensitivity studies will be performed.

THRUST 3 - SENSITIVITY, VALIDATIONS, AND FIELD STUDIES

Investment Optimization and Decision Algorithms

Considerable research has been conducted on the topological analysis and optimal flow of single networks (Newman et al. 2006). On the other hand, the mathematical modeling and optimization of interdependent networks - a fundamental component of community resilience -is relatively new. A key distinction between the two paradigms is that a network component identified as minor importance from the perspective of an isolated network may be critical when the network is considered from an interdependent framework. Failures of links or nodes in interdependent networks may lead to cascading effects and system-wide failure. Variation in the failed links or nodes will affect the speed at which a community recovers from a disaster as well as the extent of recovery. The size and complexity of a community-level model require novel algorithmic approaches that produce accurate solutions efficiently. Intelligent search algorithms based on meta-heuristics are being developed and applied to effectively search the solution space (defined by classical modeling techniques), identify critical interdependent components, develop strategies to reduce time to recovery, and determine nearoptimal network designs which improve community resilience. These algorithms are being incorporated into NIST-CORE.

Engineering and economic investments to improve community resilience are being evaluated against a no-action (laissez faire) benchmark. More specific, investment expenditures meant to boost resilience to hazard shocks involve reallocation of resources away from normal expenditures in the economy. The net short-run economic impact of reallocating resources is negative. However, insofar as these investments are resiliency-enhancing, they will function to mute initial losses (robustness) and speed return to pre-shock equilibrium (rapidity), thus enhancing resourcefulness going forward (Bruneau et al 2003). The negative short-term costs of investments, both direct and indirect (from system efficiency losses), will be compared to the long-term losses averted by community resilience for various investment options as part of this optimization.

To summarize, NIST-CORE will model how a natural hazard impacts the functioning of commercial and residential buildings, transportation, utilities (water), telecommunications and energy. A Computable General Equilibrium (CGE) model will be used to estimate the economic impact of all these effects. The CGE model can estimate the initial downturn in the economy and also estimate the economic resilience of the economy depending on the resilience of the five engineering scenarios. Alternative mitigation policies that can reduce the impact of the hazard with a focus on the costs and benefits of each policy will be developed.

Infrastructure Evaluations

NIST-CORE will be evaluated for individual infrastructure systems (e.g., buildings, transportation), coupled infrastructure systems (e.g., power-water, power-buildings-water), and the fully coupled model (all components including economics and social structures). Observations from Hurricane Ike will be used to validate the individual system models. Hurricane Ike is particularly suitable for this because damages were confined to a relatively small region (Houston, Galveston). For initial validation of the fully coupled system, observations and data from both the Northridge earthquake and Hurricane Ike will be used.

NIST-CORE Architecture and Ontology Validation Studies

Data for Hurricanes Ike and the Northridge earthquake will be used to perform beta testing approximately every 4 to 6 months as the full NIST-CORE architecture is being developed during Year 2 and into Year 3 of the Center. This will allow a systematic validation of each of the integration steps for databases, open source environment communication, search algorithms, and investment optimization. Validation of the NIST-CORE computational environment architecture and forecasts of risk and community resilience based on available computational models and supporting databases will contain significant unavoidable uncertainties. In any application of NIST-CORE (validation or forecasting), the question of paramount importance is whether the models and databases embedded (or interfaced) in it are sufficient to assess infrastructure risk or forecast community resilience with a level of accuracy (or confidence) sufficient for rational decision-making. Within NIST-CORE this critical research challenge is being addressed by structuring the decision analysis module within the environment within a machine learning shell that will enable NIST-CORE to "learn" from the data

provided to it, recognizing specific patterns or capturing statistical characteristics during the course of the resilience forecast, and advising the analyst or the decision-maker as to whether the forecasts are likely to be useful.

Field Studies

Field studies will be conducted to enable validation of NIST-CORE for individual systems, interdependencies and cascading effects if present, and the fully integrated data architecture. Mid-term (30 days to 6 months) and long-term field studies will help validate NIST-CORE accuracy in computing recovery trajectory. Data will be collected to characterize the hazard, damage to infrastructure and social systems, and recovery by the community.

ACKNOWLEDGMENTS

The Center for Risk-Based Community Resilience Planning is a NIST-funded Center of Excellence (NIST-CoE); the NIST-CoE is funded through a cooperative agreement between the U.S. National Institute of Science and Technology and Colorado State University (NIST Financial Assistance Award Number: 70NANB15H044). The views expressed in this paper are those of the authors, and may not represent the official position of the National Institute of Standards and Technology or the US Department of Commerce. The authors acknowledge contributions from the entire NIST-CoE team and NIST research collaborators who are listed at: http://resilience.colostate.edu/index.shtml

REFERENCES

- Bocchini, P., and Frangopol, D. M. (2013). "Optimal resilience- and cost-based post-disaster intervention prioritization for bridges along a highway segment." J. Bridge Eng., BE.1943-5592.0000201, 1-13.
- Bruneau, M., S. Chang, R. Eguchi, G. Lee, T. O'Rourke, A.M. Reinhorn, M. Shinozuka, K., Tierney, W. Wallace, and D.V. Winterfelt. (2003). "A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities." Earthquake Spectra, 19 (4); 733-752
- Chang, S.E., T. McDaniels, J. Foz, R.Dhariwal and H. Longstaff. (2014). Toward Disaster-Resilient Cities: Characterizing Resilience of Infrastructure Systems with Expert Judgments. Risk Analysis, 34(3):416-34
- Frangopol, D.M., J.J. Kallen and J.M Van Noortwijk (2004), "Probabilistic models for life-cycle performance of deteriorating structures: review and future directions," Prog. Struct. Engrg. and Mat. 6(4):197-212.
- Gilbert, S.W. (2010). "Disaster Resilience: A Guide to the Literature." NIST Special Publication 1117, National Institute for Standards and Technology, Gaithersburg, MD.
- Highfield, W., W.G. Peacock, and S. Van Zandt, (2014). "Mitigation Planning: Why Hazard Exposure, Structural Vulnerability, and Social Vulnerability Matter." Journal of Planning Education & Research. 34(3):287-300.
- IEEE (2012). IEEE 1366-2012; IEEE Guide for Electric Power Distribution Reliability Indices.
- IPCC (2013). Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.
- McAllister, T.P. (2013). "Developing Guidelines and Standards for Disaster Resilience of the Built Environment: A Research Needs Assessment." NIST Technical Note 1795, National Institute for Standards and Technology, Gaithersburg, MD.
- Mori, Y. and Ellingwood, B. (1993). "Reliability-based service life assessment of aging concrete structures.": J. Struct. Engr. ASCE 119(5):1600-1621.
- Murphy, C., Gardoni, P., and Harris, C.E., (2011). "Classification and moral evaluation of uncertainties in engineering modeling," Science and Engineering Ethics, 17 (3), 553-570.
- Newman, M. E. J., Barabsi, A.-L. & Watts, D. J. eds. The Structure and Dynamics of Networks (Princeton Univ. Press, 2006). NAE (2010). Disaster Resilience: A National Imperative. The National Academies, Washington D.C.
- NIST (2014) Measurement Science R&D Roadmap for Windstorm and Coastal Inundation Impact Reduction, NIST GCR 14-973-13.
- NRC, 2006. Facing Hazards and Disasters: Understanding Human Dimensions, National Research Council. The National Academies Press, Washington D.C.
- Ouyang, Min (2014). "Review on modeling and simulation of interdependent critical infrastructure systems". Reliability Engineering and System Safety 121 (2014) 43-60.
- Peacock, W.G., H. Kunreuther, W.H. Hooke, S.L. Cutter, S.E. Chang, and P.R. Berke. 2008. Toward a

Resiliency and Vulnerability Observatory Network: RAVON. Final Report NSF Grant SES-08311115. Hazard Reduction and Recovery Center, Texas A&M University.

Peacock, Walter Gillis, Shannon Van Zandt, Yang Zhang, and Wesley Highfield. 2014. Inequities in Long-term Housing Recovery After Disasters. Journal of the American Planning Association. 14(4): DOI: 10.1080/01944363.2014.980440

Poland, C.D. (2013). "SPUR Resilient City Goals, Roundtable on Standards for Disaster Resilience for Buildings and Infrastructure Systems." September 26, 2011.

Presidential Policy Directive/ PPD 8 (2011). "National Preparedness." www.whitehouse.gov. Presidential Policy Directive/ PPD 21 (2013). "Critical Infrastructure Security and Resilience." www.whitehouse.gov.

Rose, A. "Economic Resilience to Disasters." Community and Regional Resilience Institute (CARRI) Report 8, November 2009.

Reed, D.A., K.C. Kapur and R.D. Christie (2009). "Methodology for Assessing the Resilience of Networked Infrastructure." IEEE Systems Journal, Vol. 3, No. 2, pp. 174-180.

Tamiczek, T., A. Kennedy, S. Rogers, (2014) "Collapse Limit State Fragilities of Wood-Framed Residences from Storm Surge and Waves during Hurricane Ike," JWPCOE, 140, 43 - 55.

Van Zandt, S., W.G. Peacock, D. Henry, H. Grover, W. Highfield, and S. Brody. 2012. Mapping Social Vulnerability to Enhance Housing and Neighborhood Resilience. Housing Policy Debate, 22(1):29-55.

Xiao, Y., and S. Van Zandt. 2013 "Building Community Resiliency: Spatial Links between Household and Business Post-disaster Return." Urban Studies, 49(11):2523-2542.

Considerations and Recommendations for Data Availability for Data Analytics for Manufacturing

Don Libes, Seungjun Shin, Jungyub Woo Engineering Laboratory National Institute of Standards and Technology Gaithersburg, Maryland, 20899 USA {libes,seungjun.shin,jungyub.woo}@nist.gov

Abstract— Data analytics is increasingly becoming recognized as a valuable set of tools and techniques for improving performance in the manufacturing enterprise. However, data analytics requires data and a lack of useful and usable data has become an impediment to research in data analytics. In this paper, we describe issues that would help aid data availability including data quality, reliability, efficiency, and formats specific to data analytics in manufacturing. To encourage data availability, we present recommendations and requirements to guide future data contributions. We also describe the need for data for challenge problems in data analytics. A better understanding of these needs, recommendations, and requirements may improve the ability of researchers and other practitioners to improve research and more rapidly deploy data analytics in manufacturing.

Keywords- big data; challenge problems; data analytics; data quality; requirements; smart manufacturing

L INTRODUCTION

Data analytics for the purpose of improving performance in the manufacturing enterprise is becoming increasingly attractive for manufacturers. This is motivated by several factors. First, "big data" characteristics (increasing volume, velocity, veracity, variety, and complexity) from models, sensors, etc., enables new approaches (e.g., algorithmic, visualization) that produce novel conclusions. These novel conclusions lead to a second factor -- an increasingly competitive environment. Manufacturers who do not use data analytics will not be competitive with manufacturers who do [1]. Competition can drive prices down, increasing costs for resources and additional regulation provides an additional factor for finding savings wherever possible. Thus, data analytics can be expected to play a critical role in smart manufacturing in the future.

While papers and books describe the promise and some successes of data analytics, few real-world descriptions of the difficulties, challenges, and pitfalls exist [2]. Data analytics is not a cookbook science. Granted, one can find plenty of papers describing success; one can find papers describing methodologies and models; and one can find papers comparing algorithms. However, in our literature search, we could not find papers specific enough to guide anyone through the complete process of deploying data analytics. Each enterprise traditionally makes its own

decisions about what data are important and how these data should be stored and analyzed.

Complexities exist for many reasons such as uniqueness and diversity of manufacturing enterprises and goals, lack of standards or in some cases, multiple standards without compatibility. Work on these areas is hindered by a lack of widely-recognized challenge problems in data analytics for manufacturing applications. These problems must be meaningful and find wide acceptance by researchers across many fields and industries such that progress is possible in reasonable timeframes. Underpinning all of these is a corresponding need for real data. By its very essence, data analytics cannot function without data. For the same reason, research into data analytics also requires data.

The National Institute of Standards and Technology (NIST) is researching data analytics to support U.S. manufacturing. However, NIST is not a manufacturer and has no direct access to the amounts and types of big data that forms the basis for most of our research into data analytics [3][4]. NIST has therefore had to rely on industrial partners for data. Unfortunately, these data are often problematic in a number of ways which in turn impedes the research. In our own work, we have faced significant issues dealing with contributed data and had to expend significant time resolving these issues due to the lack of relevant guidelines for data acquisition, types of data, and the volume of data. This is the motivation and source for much of this paper.

This paper discusses the complexities of producing usable manufacturing data. The paper also provides recommendations on what can be done to 1) produce more data that will help researchers (such as NIST and academia) of data analytics in smart manufacturing, and 2) help manufacturers that want to supply their data to researchers. We describe the types and qualities of data needed. We describe other attributes of significance. We also describe the reliance that many researchers have on industry data, why this reliance presents difficulties, and suggestions on how these difficulties can be ameliorated. At the very least, we need to understand how critical data is to our current state of the art of data analytics. In that sense, we believe that our thoughts, observations, and recommendations in this paper may be helpful as guidance and toward stimulating further discussion

While our own interest stems from manufacturing issues, that is not intended to limit applicability of this paper.

Libes, Donald; Shin, Seungjun; Woo, Jungyub. "Considerations and Recommendations for Data Availability for Data Analytics for Manufacturing." Paper presented at 2015 IEEE International Conference on Big Data (IEEE Big Data 2015), Santa Clara, CA, United States. October 29, 2015 -

Readers may find that much of our discussion is widely applicable to other domains.

П DATA AVAILABILITY ISSUES AND RECOMMENDATIONS

Deciding what data to collect, how to collect, and how much to collect is a significant challenge. Such decisions include the source of data and their type, location, and frequency. Each of these may have additional complexities. For example, locations may change manually or automatically. Sensors may be mobile - for example, ontransport devices or in-process parts. It may be of value to measure all parts or just a percentage and frequency may be regularly periodic or event-driven. Each of these issues has further ramifications. For example, using data from multiple sources can be difficult when the sources are unsynchronized or use different output formats. This "data fusion" cuts across many of the issues we describe.

This section of the paper describes issues and potential recommendations that we are considering.

Α. Data Use Restrictions

We recognize that there may be restrictions on the way certain data are used. While this may not be true in an experimental test environment, at a certain point, we want to reflect real world concerns. Restrictions arise in several ways that are further described here.

1) Proprietary Data

Significant amounts of data exist but are inaccessible to researchers as the data are proprietary. Many companies collect such data but only make it available internally or to a limited set of partners with confidentiality agreements.

The problem of proprietary data includes both generated (for example, from sensors) and intrinsic data. For example, the manufacturer of a milling machine may decline to provide data describing speed and torque relationships. Possibly, the manufacturer may not even have the kind of comprehensive data that researchers want. Manufacturers may have models and other descriptions of their products to the extent needed for their own manufacturing purposes but may be unwilling to release the data.

The unwillingness, for example, of machine tool manufacturers to share data is understandable. Models and performance data can be used by competitors to improve their own manufacturing processes and products. While the very largest buyers of such equipment may have the leverage to access (or can afford to pay for) proprietary information, it is inaccessible to the majority - especially to mid- and smallsize manufacturers.

Without this proprietary information, researchers are left to make sample parts and carry out onerous experiments and measurements, all for limited purposes. With limited resources, many assumptions must be made since these experiments involve physical operations rather than virtual simulation. For this reason, we are left with researchers proposing equipment and other manufacturing decisions that are far from optimal despite the possibility of significant improvements with access to proprietary data.

Although we could recommend that equipment manufacturers make as much of their data as possible available to enable and encourage researchers, this statement is too vague. It may be more useful for them to, for example, make available complete data for a machine that is no longer sold but nonetheless widely used. Similarly, manufacturers can release data for manufactured products or process plans that are no longer used or critically proprietary.

2) Data Security

To thwart corporate espionage or even inadvertent data leaks, data at all levels must be protected. Traditionally, any user or application behind a firewall can access any and all data. For enhanced security, it is becoming common to compartmentalize access on a need-to-know basis [5]. This not only requires justification and requests, but also time limits. This can be problematic since need-to-know and ondemand philosophies conflict with the ability of algorithms that expect open access to all data. In part, this is because the algorithms themselves may not know in advance what data they need to explore.

This problem is exacerbated by the use and number of suppliers, shippers, and other external parties in the supply chain. These parties will likely not allow access to their internal data except in extremely limited ways. This significantly limits the types of, and the extent of both data discovery and analytics that can take place. In short, the traditional benefits of using third parties to gain efficiency and cost advantages are offset by these data-access issues and the less-than-optimal decisions that result.

Infrastructure deemed critical refers to elements crucial to such issues as national economic security, economic viability, and public safety. These may include both longterm and short-term issues that arise from deliberate attacks or disruptions from natural events [6]. One important strategy to provide security is to reduce risk by minimizing access. From a security perspective, the more that access is restricted, the more likely it will achieve the goals of reducing risk and defending infrastructure. Restrictive access thwarts data analytics where the more expansive and ready access to data, the more likely it will be possible to detect previously unknown patterns and conclusions from massive quantities of seemingly unrelated data.

Encryption may be used to shield data from inappropriate consumers. This introduces drawbacks such as extra processing which could be difficult at low levels with limited processing power. Encrypted data also generally take more space both during transmission and when resident. Finally, encryption provides yet another opportunity for data leakage as encryption keys provide another attack surface [7]. However, encryption may be useful when sharing data in a scenario that does not require real-time access. Many encryption variants exist. For example, homomorphic encryption allows computations directly on encrypted data. Decryption can be done at a later time in a secure environment and the result will match the results that would have occurred had the data analytics been done on the unencrypted data [8].

Data producers may find it useful to digitally sign data as proof of provenance and that the data have not been changed.

Libes, Donald; Shin, Seungjun; Woo, Jungyub. "Considerations and Recommendations for Data Availability for Data Analytics for Manufacturing." Paper presented at 2015 IEEE International Conference on Big Data (IEEE Big Data 2015), Santa Clara, CA, United States. October 29, 2015 -

This latter can be useful as researchers want to be able to replicate results of others or want to assert improved results on the same body of data. Digital signatures can also be used with public-key encryption so that only partners with corresponding private keys can read the data.

Security-related recommendations for improving access to data do not generalize widely. We can only raise awareness of the tradeoffs.

B. Non-Existent or Missing Data

In some areas, data may not exist for a variety of reasons. Examples: Data may be too expensive to capture - installing sensors or buying models may cost more than the expected benefit. It may be impossible to capture data that are internal to a sealed system. A KPI (Key Performance Indicator) may be unavailable if it is too new or vaguely defined. Sensors, communications, or a database may have failed in a run that is too expensive to repeat.

Such non-existence of data, while frustrating, occurs in real scenarios. Sensors don't always work. Databases don't always record. For a variety of other reasons, it may be impossible to obtain data.

Workarounds include the use of simulated data, reference data, and reasoning techniques [9][10][11]. Simulated data are generated by models and algorithms rather than measured in situ. Reference data are any data acquired second-hand. Reference data can derive from a wide-variety of sources (design goals, for instance) as well as measured. However, while reference data are ideal in theory, it is rare to find different manufacturers who have identical independent variables in their manufacturing systems except in very simple processing. Even manufacturing operations of commodity items generally vary widely among manufacturers. Lastly, reasoning techniques such as Bayesian networks may provide suitable workarounds to address missing data.

C. Data Timing and Synchronization

Synchronization of data is a significant problem in several ways. An enterprise typically has many producers of information. To make use of such information from multiple sources, it is necessary to understand when that information was created. This is more difficult than simply timestamping data from a central time server.

Consider a sensor on a machine tool that is reporting force data. It is necessary to understand how this correlates with machining operations. Operations can be viewed as a hierarchy with a process plan at a high level and machine instructions describing speed and position at a low level.

There are two significant issues: synchrony of sensor timing, and synchrony of processing steps.

First, timing chips, while inexpensive, are not a part of many sensors [12]. The resulting data are unsynchronized or require retroactive synchronization to deal with relative timestamps. Absolute timing adds cost. While central-time service distribution is achievable, adding this functionality to all sensors increases costs. This is especially problematic because there will typically be hundreds of sensors in a work cell, and many thousands throughout a factory. Relative synchronization is cheaper but is non-trivial. A priori synchronization is not trivial nor is adding timestamps at a later stage (for example by an MTConnect-aware server that sends the results back to the database), which generally adds latency resulting in inconsistent time lags.

Second, processing steps are typically hidden inside the processing elements of a machine tool, and can be inaccessible to the external world. This makes it difficult to correlate these processing steps to sensor measurements. Even if these steps are known (for example from experimentation), there can be indefinite latency since operations and algorithms inside commercial controllers are proprietary and not intended to be examined nor guaranteed. This latency introduces mismatches between low-level data, making certain types of data analytics impossible or introducing a significant amount of uncertainty.

Most of what is available today are unsynchronized sensor data. We believe that many sensor manufacturers or integrators simply don't view this type of synchronization as a requirement. Hence it remains a challenge for researchers.

Recommending that all sensors and machine tool internals include absolute timing circuitry is simple but unrealistic. Instead, we recommend that such information be tagged at the lowest level possible with absolute timing information with millisecond precision or, if possible, submillisecond. Timestamps must be in a format that is directly machine processable such as ISO (International Organization for Standardization) 8601 [13]. Timings that are known or likely to be affected by lag should be tagged as such. This should include a measurement or estimate and probability of the lag.

D. Data Frequency

During manufacturing, there is a tradeoff over the frequency with which data are collected. While it is intuitive that additional data can lead to detection of additional patterns, increased accuracy, and the potential for otherwise undiscoverable insights, additional data also have costs.

Increased data means increased cost of processing time, data storage, access, network traffic, and database contention. Synchrony (see above) and the cost and accuracy of interpolation are also a concern as is algorithm performance. While an increase in data can lead to better results, algorithms will invariably take more time. Data mining algorithms already use heuristics to trade off exponential performance for suboptimal results so unnecessary data only leads to further deteriorating performance. This is especially problematic when it impedes the ability to draw conclusions and to make adjustments to processes in real-time.

One argument for collecting more data is that it is simpler. No knowledge is needed on how to pass judgment on data. So immediately after having taken and handed off a measurement, the sensor begins the cycle again with no waiting for any kind of clock cycle. Of course throttling may be induced indirectly. For example, the network may be incapable of absorbing data at the rate at which the sensor sends it. The database may not be fast enough. The architecture stack may induce throttling at many different

levels. Whether this throttling occurs directly (planned throttling) or indirectly is a significant concern. For example, if a database is thrashing due to the amount of data it receives or from demands by the analytics algorithms, gaps that appear between batches in the data may be large, causing poor performance of algorithms - as if less data were collected than actually was [14]. Understanding and managing these tradeoffs is beset by subtle choices and risks.

During research or pre-production, the problem of excess data is ameliorated to an extent. For example, real-time performance may be unnecessary. However, badly performing algorithms are still at risk with even minor increases in data size. And data mining by its very nature involves exploring data for unexpected results. Therefore, more data are better. However, not all data are of equal value. Being able to "help" algorithms by withholding data is a challenge. A related issue is that some algorithms perform worse with additional data. For example, for certain machine-learning algorithms, increasing the training set size can produce outputs that are inferior to those of smaller training set sizes. While this phenomenon is data dependent, it is a significant concern since this runs counter to expected outcomes that suggest more data are better [15]. This arises with other algorithms as well such as overfitting in neural network training. And, in practical terms, the world is moving to essentially unlimited data so this question of data size arises frequently.

The goal of replacing unlimited big data with limited ideal data is often called the "smart data" problem. However, it remains an open problem [16]. For now, we are stuck requiring, generating, storing, and analyzing "big data."

To assist with the problem of frequency, we recommend that data publishers produce two sets of data in situations where frequency is an issue. One set should be a complete set with all measurements. A second set should be limited to a more realistically representative amount that would normally be available in production. Additional sets may be provided as supplements but should be marked appropriately - for example, indicating extra sensors are used that would not normally be available.

E. Data Formats, Standards, and Specifications

Enterprises sit atop a vast collection of disparate data, often produced by a multitude of heterogeneous sensors, and often ultimately stored in files formatted according to a variety of standards, with varying degrees of compliance. Significant amounts of data may follow no standards whatsoever. Using standard specifications such as XML (eXtensible Markup Language) and JSON (JavaScript Object Notation) can help [17]. However, problems such as underspecification can remain that leave ambiguities. For example, using an XML attribute called "time" means little if there is no definition for how the string is to be interpreted -absolute with respect to UTC (Coordinated Universal Time) or local time zone? Relative to what -- start of a process or something else?

While standards can be helpful, they are not panaceas. For example, equipment and software from different vendors may use different standards. Data, while still standardcompliant, can lose fidelity during interchange. Standards frequently have different levels of compliance that users may choose from. Even highly specific standards do not guarantee data that are usable. For example, the machine tool standard, MTConnect, covers only one direction of communication; so, correlation to commands may not be present or need reconstruction with timing uncertainties. Equally important, MTConnect does not cover all possible types of data that a machine tool can generate [18]. For example, MTConnect defines a fixed set of statistics for DataItems. Kurtosis, a measure of peakedness relative to a normal distribution, is in the set. Skewness, a measure of symmetry, is not in the set.

All of these choices have reasons for existence. For example, different vendors may have different reasons for their choices. These can include historical issues, expense in tracking developing standards, and interactions with other software. For example, the choice of OWL (Web Ontology Language) variants depends on how much need there is for expressiveness - the breadth of concepts that can be represented. But greater expressiveness brings with it a loss of computational guarantees [19].

Data standards may be descriptive (describing practices) or prescriptive (defining practices). Each carries with it downsides. For example, descriptive standards may prevent the use of innovative techniques that are too new to be incorporated in standards while prescriptive standards may be ignored when better technology solutions are discovered. These dilemmas are particularly apparent in rapidly changing and highly-competitive fields. To allow variations and technological advances, some standards intentionally leave areas of ambiguity with a resulting ambiguity in the data.

Some standards and specifications have been created to help with the issues mentioned in this section, such as PMML (Predictive Model Markup Language) and PFA (Portable Format for Analytics) [20][21]. For example, PMML and PFA can specify time offsets or transformations such as normalization.

We recommend that well-recognized standards be used at every level of data formatting and description. Metadata should be supplied to describe data formats and meaning. Proprietary and ad hoc standards should be avoided except when necessary. Standards and specifications such as PMML and PFA should be used instead of ad hoc programming language solutions to deal with likely conversion issues.

F. Data Uncertainty and Reliability

Enterprise data can have all the characteristics common to any collection of big data. These include uncertainty, reliability, and accuracy, among others. It is important that data collections address these types of characteristics in a realistic way.

Some of these characteristics should be dealt with by providing the appropriate metadata (data describing data). For example, uncertainty that is quantified should be expressed in metadata included with the data. In some cases, however, uncertainty may be unknown a priori and it is the task of the data analyst to deduce the uncertainty during the

Libes, Donald; Shin, Seungjun; Woo, Jungyub. "Considerations and Recommendations for Data Availability for Data Analytics for Manufacturing." Paper presented at 2015 IEEE International Conference on Big Data (IEEE Big Data 2015), Santa Clara, CA, United States. October 29, 2015 -

analytics process. As in the former case, the latter possibility can and should be expressed in the metadata.

In the general sense, problematic data presents several choices. It is tempting to cull data before publishing. It may be similarly tempting to artificially induce problematic data whether programmatically or physically - such as by using an intentionally faulty or broken sensor. No matter which of these approaches are taken, metadata must indicate what is known about the data.

When there has been modification of data (removal, normalization, etc.), we recommend that multiple data sets be made available corresponding to the original and the modified data.

G. Data Access, Storage, and Processing

Data access and availability impact data analytics. In even simple cases, data is structurally changed when it is distributed to researchers. For example, distribution traditionally has meant serialization. However, this raises issues of how de-serialized data are intended to be stored and accessed. Alternatively, researchers might have direct access to the database of a producer, perhaps because the producer is generating information in real-time or, more commonly, because the data set size is extremely large and researchers only need access to a relatively small portion.

Our estimates of a small manufacturing shop scenario with a modest number of machine tools using Hadoop/Hive technology could generate daily data of 1 TB (25 MB/sec) based on 500 K/day sensor readings at 50 KB/reading with database overhead, data normalization, and other issues. With replication, compression, and caching for 250 workdays/year, this approaches 1 PB required to retain a year's worth of data. Our estimates may soon be seen as conservative given that the cost to store data continues to decrease and the ability to generate data continues to increase.

Data may be converted to an entirely different database type by, in essence, changing the solution. Data storage choices are significant since different database technologies have different strengths and weaknesses. For example, highly-regular relational data are optimally stored in a relational database. In contrast, data that are predominantly hierarchical such as geospatial (e.g., factory floor map) incur a significant extra expense when stored relationally. Another example is that a fixed data schema can significantly improve the performance of data mining by reducing the time to access data. In contrast, more agile databases have the flexibility to store new data types but only by increasing the expense of access.

A related problem is that some enterprises have limited database expertise with the result that inferior database solutions are employed. For example, data may not be normalized so data are redundant, contain inconsistencies, lack connections, etc. While this might not be a problem for ad hoc in-house use, distributing such databases causes significant and unnecessary extra effort and frustration.

More sophisticated enterprises will use multiple databases. Different machines, different caching, different bandwidths all affect availability. For example, joins within a

database are generally easy; joins across databases on multiple machine often incur significant expenses. However, multiple databases are a more realistic portrayal of enterprise data since they can simplify the problems of rapidly changing applications and data analytics demands. This contrasts with traditional data distribution using a single file, however large.

We mentioned earlier that databases might process data before storage. As processing elements become increasingly powerful, it is common to find databases at the lower levels of the enterprise with sophisticated operations and storage capacities. Such low-level databases can be used to address certain problems such as avoiding network overload by only sending up significant results rather than raw data. We have hypothesized that stored programs may also be used to solve security issues noted earlier (see Proprietary Data) in that processing takes place inside of a controlled proprietary environment so that results are returned without exposing raw data [21]. However, we have not seen any evidence of this in practice and remain skeptical of it as a workable technique. Nonetheless, the use of low-level processing in general can have the inadvertent effect of making data availability less likely.

Multiple researchers who maintain the same data in different ways may be solving different problems. Therefore, we recommend that data sets identify how they are nominally intended to be accessed and, if necessary, stored. Researchers who opt to use different techniques should state those differences when releasing results so it is clearer when comparisons are being made using different access and storage technology as well as other aspects of the research.

H. Data Preparation Time and Other Resources

Assuming a manufacturer is supportive of releasing data, the time and effort required can be substantial. The potential for unaffordability can be a reason to prevent the release of data.

We have already mentioned many categories of issues. Each of these takes time. For example, a manufacturer must take the time to decide what is proprietary and what is not and how to separate them. Decisions may change due to multiple influences. For example, sensor measurements that may initially seem distributable may "give away" too much information if taken at a much higher frequency or with greater accuracy.

Having engineers devote limited time to selecting database fields, formatting and packaging data, finding a suitable repository, keeping it updated, and creating data descriptions, can be difficult to justify if it does not immediately feed back to the corporate bottom line. Similarly, legal requirements and concerns may force the use of additional resources, for example, for non-disclosure agreements that may vary across consumers and data sets.

We recommend that, before embarking on data distribution projects, manufacturers create a budget that incorporates all aspects of time and resources that will be needed. This not only protects a company from unwelcome surprises but is more likely to lead to a careful and useful release of data that can be sustained in the future.

Libes, Donald; Shin, Seungjun; Woo, Jungyub. "Considerations and Recommendations for Data Availability for Data Analytics for Manufacturing." Paper presented at 2015 IEEE International Conference on Big Data (IEEE Big Data 2015), Santa Clara, CA, United States. October 29, 2015 -

Ш CATEGORIES OF DATA

A. Application Domains

It is important to understand the categories of data that exist in an enterprise so that data may be properly acquired and used for the purpose of data analytics. We are in the process of enumerating a comprehensive list of application domains and needs of manufacturing. This section describes a few examples to show the general idea.

1) Supply-Chain Management Data

Supply-chain data include product types, scheduling, risks, cost accounting, efficiency, sustainability, and other factors. Many of these may include choices, for example between multiple suppliers or products and tradeoffs. Singlesupplier, single-product data sets are useful but only for very limited purposes.

Access to these kinds of data would allow investigations into problems such as maximizing responsiveness to change, minimizing inventory and energy use, and balancing cost vs critical dependencies.

2) Production Scheduling Data

Production scheduling data include performance criteria, product data, work cell models, and resource descriptions. Data should include a variety of generated schedules under stated conditions (for example, fixed maximum total time or minimizing total time needed) and computed outputs (such as energy used, time expended, and other KPIs).

Access to these kinds of data would allow investigations into production efficiency, load leveling, dynamic rescheduling, and trade-offs such as inventory reduction vs total production throughput.

3) Process Planning Data

Process planning data include product data, resource descriptions, and fabrication techniques. Data should include a variety of process plans under stated conditions (for example, fixed maximum total time or minimizing total time needed) and output (such as energy used, time expended, and other KPIs).

Access to these kinds of data would allow investigations into work cell efficiency and trade-offs such as energy vs manufacturing time vs throughput.

4) Machine Tool Data

Machine tool data include tool models, energy, reliability, and tool wear. Data should describe a variety of raw materials, processes, tools, and outputs such as energy use, time expended and other KPIs).

Access to these kinds of data would allow investigations into comparisons of different processes, machines, and tools, as well as experiments trading off various factors such as material cost, cutting speed, tool wear, and mean time to failure.

B. Types of Data

Separately from the preceding domain categories, there are different types of data characterized, not by the application domain, but by their type. Each has significant concerns.

1) Key Performance Indicators

KPIs are a type of metric designed to guide decisionmaking, for sustainability or profitability for example. However, use of common KPIs can still not be taken for granted. For example, in a survey of manufacturers, 36.8 % of respondents did not use KPIs [23]. Reasons included "Unsure of what data to measure" (20.7 %) and "Not convinced that measuring KPIs adds value" (24.1 %). Even KPIs that intuitively sound useful may not be regarded so by many people. For example, less than half of respondents who use KPIs in the Schenck survey agreed that it was worth such straightforward metrics for equipment utilization (42.6 %) and downtime (40.4 %).

ISO 22400 defines standard key performance indicators for manufacturing operations management [24]. While we do not discourage collection and use of other KPIs and the merits of existing and new KPIs are a regular source of debate, we encourage the use of standard methods for KPIs collection where they exist.

2) Human Interaction

While minimizing manual activity is desirable due to tradeoffs of cost, consistency, and reliability, some human activity will remain. Unfortunately, each type of data collected about human activity comes with different and potentially expensive tradeoffs. For example, keyboard interaction is easily and inexpensively tracked, even to individual keystroke values and timing. In contrast, observing what a human is looking at, while trackable, is expensive. Eye movements can be recorded, screen videos can be saved, and the two can be synchronized. Other types of interactions are even more expensive ranging from shortterm indicators (e.g., muscle effort, joint stress) to long-term indicators (e.g., endurance, concentration).

The justification of collecting such data may seem difficult for the very reason that such interactions are necessary in the first place. Namely, the interaction cannot be inexpensively and reliably automated in the first place. Whether it is human judgment, human vision, or some other type of interaction that is seemingly simple for humans but hard for machines, these types of issues make these data among the least easy to deal with.

At the same time, data from observing human interaction is intriguing. While high-level judgments may remain difficult, they have a significant potential to lead to more efficient human activity in the workplace. However, such observations are remarkably easy to overlook and likely to be misunderstood without data.

3) Data Models

Much of our focus on data collection has been on lowlevel data collected directly from sensors. However, understanding these data can be difficult or impossible without the underlying models on which the data are based.

Models are descriptions of designs, activities, and algorithms. They exist at every level of an enterprise - from supply chain to machine tool. Such models include a variety of information including requirements, architecture diagrams, descriptions, interaction and physical specifications. Use of models have several problems:

Models may contain proprietary information which is stripped before distribution.

Libes, Donald; Shin, Seungjun; Woo, Jungyub. "Considerations and Recommendations for Data Availability for Data Analytics for Manufacturing." Paper presented at 2015 IEEE International Conference on Big Data (IEEE Big Data 2015), Santa Clara, CA, United States. October 29, 2015 -

- Models may require integration with other models or it may be unclear how they relate to the data.
- Models may be more aspirational than accurate. Implementers may have deviated from the models.
- 4) High-level Data and Metadata

We have already mentioned metadata in several contexts earlier in the paper such as describing the format, timing, and source of data. However, metadata is a more general issue that deserves additional elaboration. As an example, many axioms or assumptions are metadata that should be documented.

Consider an investigation into energy use. While energy use may be readily available from power meters and fully recorded, hidden assumptions may not be. The price of energy may already have been considered when the enterprise was designed. A high-cost energy location will naturally have minimized the use of energy while a low-cost energy location will lead to data that is biased toward profligate energy use. Such assumptions may not make sense for researchers, depending on the focus of the research. As another example, consider a factory in a location where energy cost is time-sensitive. Time-sensitive energy pricing may come about from the power company or in-house sources such as solar power availability, which is only available during the day. A failure to record this metadata will lead to energy-use data that may look quite odd.

Other metadata omissions can be equally problematic. For example, consider a machine tool shared across production lines or one that is older and needs more frequent maintenance. These types of factors will influence the data. However, a lack of explanation will leave inexplicable impacts in the research conclusions. Similarly, data may reflect expenses related to machine tools due to requirements in the product line that are not evident in the data. Schedules (i.e., absences) related to trained personnel availability may also perturb data in similar ways.

5) Simulated Data

Some projects have addressed a lack of physical data by creating simulated sensor data complete with comparisons of simulated data to physical data. This is not necessarily easy since real-world artifacts will include errors for many reasons such as calibration drift or pseudo-random errors. These must be modeled (to some degree) for useful data. In some cases, simulated data may be a useful substitute. In others, simulated data may not be useful. Therefore, researchers must take care to appreciate the differences in requirements.

We believe there is a valid use for simulated data and encourage development in data analytics. However, such data must include appropriate metadata that explains its simulated nature and source. This may include assumptions, requirements, and source code for the simulated data generators.

IV. DATA REPOSITORIES

There is strong motivation to create "big data" repositories to hold contributions from industry as well as academia. Two such repositories exist already.

- CO2PE! (Cooperative Effort on Process Emissions in Manufacturing) is an initiative with a number of objectives related to improving manufacturing processes. CO2PE! proposes to "develop a methodology that allows to provide data in a format useful for inclusion in LCI (Life Cycle Inventory) databases." [25] CO2PE! also expects that its partners will contribute to LCI data "as required for systematic LCA studies, covering the production stage..." While more specialized to specific fields and problems, the focus of CO2PE! is a strong indicator supporting the focus of our own work described in this paper.
- ecoInvent is an LCI database representing "human activity and its exchanges with the environment and with other human activities." [26] These exchanges include energy and other resources. These datasets represent the higher levels of data of an enterprise and go all the way to national tracking of resources. Some of the factors we described earlier are addressed by ecoInvent such as formats and uncertainty, ecoInvent also addresses other factors such as revision control.

Since these are specialized repositories, it is likely that there will be additional repositories created. We would like to see repositories that have the following attributes:

- Automated. We anticipate contributions will occur by connecting to a website and uploading data along with metadata that describes scenario information. In the case of direct access, information must also be provided along with certain best-effort promises of availability and any bandwidth restrictions (such as "no more than 1 GB/day without prior approval"). Researcher access can occur by manual or automatic download using Webservices.
- Secure. Digital signatures must be used to ensure authorship and to ensure changes are trackable.
- "Lightly" curated. Given the ever-increasing span and complexity of scenarios and rapidly evolving technology, it may be difficult to judge the value of most submissions. Submissions that attract little interest initially may turn out to be of significant value at a later time. Except for the space consumed, submissions that are of no interest have no impact on others
- Generated data. While we seek contributions of real data, simulated data are also useful. Ideally, simulated data will include sources to the data generator with generation parameters so that distribution size can be minimized and further experimentation is possible while allowing research conclusions to cite the parameters for the particular generation run.

One particularly desirable type of repository is for challenge-problem data. Challenge problems are problems widely recognized as fundamental or critical to advancing research in their fields. In the field of data analytics in smart manufacturing, there are many issues associated with these

Libes, Donald; Shin, Seungjun; Woo, Jungyub. "Considerations and Recommendations for Data Availability for Data Analytics for Manufacturing." Paper presented at 2015 IEEE International Conference on Big Data (IEEE Big Data 2015), Santa Clara, CA, United States. October 29, 2015 -

challenge problems including their description, availability, and relevance. Most importantly, challenge problems require data.

Challenge-problem data serve two purposes. First, such data are necessary for researchers who would otherwise have access to insufficient real data. Not all researchers have direct access to machine tools or work cells, for example. Second, having data common to all researchers addressing these problems allows their results to be evaluated and compared more easily. Without common data, trying to compare results is akin to comparing apples and oranges. It becomes impossible to do meaningful comparisons.

V. CONCLUSION

In this paper, we have described issues that presently hinder access to and usability of data. We have made some recommendations and observations with the intent of spurring more data contributions and providing guidance to make them more useful. In our own data analytics research using real manufacturing data, we have encountered many of these issues. Sometimes significant time is required for resolving problems. In the worst case, these data may not be usable at all. We hope that our experiences allow researchers to avoid these problems in the future.

We have also described the need for data for challenge problems. A better understanding of these needs, recommendations, and requirements may improve the ability of researchers and other practitioners to better study and more rapidly deploy data analytics in manufacturing. Lastly, we have described existing repositories and suggestions for future "big data" repositories for data analytics. We believe that all of these ideas will improve the opportunities for data analytics researchers and contributors.

VI. ACKNOWLEDGMENTS

Thanks to Sudarsan Rachuri, KC Morris, Al Jones, Sharon Kemmerer, Sandy Ressler, Guodong Shao, and several anonymous reviewers for their insightful critiques, suggestions, and fruitful discussions. Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

REFERENCES

- [1] L. Columbus, "84 % of Enterprises See Big Data Analytics Changing Their Industries' Competitive Landscapes In The Next Year," Forbes, Oct 19, 2014.
- D. Libes, T. Lee, "Challenges of Data Analytics in Smart [2] Manufacturing Systems," in draft.
- [3] Tata Consultancy Services, "Manufacturing: Big Data Benefits and Challenges," http://sites.tcs.com/big-data-study/manufacturing-bigdata-benefits-challenges, 2013.
- Tata Consultancy Services, "Emerging Big Returns on Big Data," [4] http://www.tcs.com/SiteCollectionDocuments/Trends_Study/TCS-Big-Data-Global-Trend-Study-2013.pdf, 2013.

- [5] K. Stouffer et al., "Guide to Industrial Control Systems (ICS) Security," NIST Special Publication 800-82, National Institute of Standards and Technology, Gaithersburg, Maryland, 2015.
- U.S. Dept of Homeland Security, National Infrastructure Protection [6] Plan (NIPP) 2013: Partnering for Critical Infrastructure Security and Resilience, 2013.
- P. Manadhata, An Attack Surface Metric, CMU-CS-08-152, Carnegie [7] Mellon University, Nov 2008.
- C. Gentry, "A Fully Homomorphic Encryption Scheme," [8] https://crypto.stanford.edu/craig/craig-thesis.pdf, Stanford University, September 2009
- R. Little, Statistical Analysis with Missing Data, ISBN 978-[9] 047118386, Wiley, 2002.
- [10] M. Karthika, J. Pearl, "On the Testability of Models with Missing Data," Proceedings of AISTAT-2014, 2014
- [11] K. Mohan, G. Van den Brock, A. Choi, J. Pearl, "An Efficient Method for Bayesian Network Parameter Learning from Incomplete Data," Causal Modeling and Machine Learning Workshop, ICML-2014 2014
- [12] C. Grimes, E. Dickey, and M. Pishko, Encyclopedia of Sensors, American Scientific Publishers, ISBN 1-58883-056-X, 2006.
- ISO, "Data Elements and Interchange Formats Information [13] Interchange - Representation of Dates and Times," ISO 8601:2004, http://www.iso.org/iso/catalogue_detail?csnumber=40874, 2004.
- [14] P. Denning, "Thrashing: Its Causes and Prevention," Proceedings AFIPS, Fall Joint Conference 33: 915-922, 1968.
- [15] W. Yousef, S. Kundu., "Learning Algorithms May Perform Worse with Increasing Training Set Size: Algorithm-Data Incompatibility,' Computational Statistics & Data Analysis, Elsevier, vol. 74(C), pages 181-197, 2014.
- [16] B. Rossi, "Is Big Data Dead? The Rise of Smart Data," Information Age, http://www.information-age.com/technology/informationmanagement/123458486/big-data-dead-rise-smart-data, Sep 23, 2014.
- [17] JSON.org, "JSON: The Fat-Free Alternative to XML," http://www.json.org/xml.html, undated.
- [18] P. Warndorf, "MTConnect Institute Releases Version 1.3.0 of the MTConnect Standard," http://mtconnect.org/recentnews/mtconnect®-institute-releases-version-130-of-the-mtconnectstandard.aspx, 2014.
- [19] World Wide Web Consortium, OWL Web Ontology Language Reference, http://www.w3.org/TR/owl-ref, 2004.
- [20] A. Guazzelli, "Predictive Analytics, Big Data, Hadoop, PMML," http://www.predictive-analytics.info/2009/04/pmml-data-preprocessing-primer.html, 2009.
- [21] J. Pivarski, "PFA: Portable Format for Analytics," http://scoringengine.org, 2015.
- [22] Y. Bernier, "Latency compensating methods in client/server in-game protocol design and optimization," Valve, Kirkland, WA, http://web.cs.wpi.edu/~claypool/courses/4513-B03/papers/games/bernier.pdf, July 7, 2015
- "Using [23] S. Schenk, Key Performance Indicators." http://www.schencksc.com/2013mfgsurvey, 2014.
- [24] ISO, "Automation Systems and Integration Key Performance Indicators (KPIs) for Manufacturing Operations Management - Part 2: 22400-2:2014, Definitions and Descriptions," ISO http://www.iso.org/iso/catalogue_detail.htm?csnumber=54497, Jan 1994
- [25] Group T, CO2PE! (Cooperative Effort on Process Emissions in Manufacturing) Home, http://www.co2pe.org, International University College Leuven, 2015.
- [26] B. Weidema et al, "Overview and methodology," ecoInvent Centre, http://www.ecoinvent.org/files/dataqualityguideline ecoinvent 3 201 30506.pdf, Swiss Centre for Life Cycle Inventories, 2013.

Libes, Donald; Shin, Seungjun; Woo, Jungyub. "Considerations and Recommendations for Data Availability for Data Analytics for Manufacturing." Paper presented at 2015 IEEE International Conference on Big Data (IEEE Big Data 2015), Santa Clara, CA, United States. October 29, 2015 -



COMMUNITY RESILIENCE OF LIFELINE SYSTEMS: SOCIETAL NEEDS AND PERFORMANCE ASSESSMENT

C. Rojahn⁽¹⁾, L. Johnson⁽²⁾, T. D. O'Rourke⁽³⁾, V. Cedillos⁽⁴⁾, T. P. McAllister⁽⁵⁾, S. L. McCabe⁽⁶⁾

⁽¹⁾ Director Emeritus, Applied Technology Council (ATC), crojahn@atcouncil.org

⁽²⁾ President, Laurie Johnson, Laurie Johnson Consulting, laurie@lauriejohnsonconsulting.com

⁽³⁾ Professor, Cornell University, tdo1@cornell.edu

⁽⁴⁾ Associate Director of Projects, Applied Technology Council (ATC), vcedillos@atcouncil.org

⁽⁵⁾ Leader, Community Resilience Group, National Institute of Standards and Technology (NIST), therese.mcallister@nist.gov

⁽⁶⁾ Leader, Earthquake Engineering Group, National Institute of Standards and Technology (NIST), steven.mccabe@nist.gov

Abstract

The concept of community resilience is a complex, multi-dimensional problem that relies on social science, engineering, earth sciences, economics, and other disciplines to improve the way communities prepare for, resist, respond to, and recover from disruptive events. Community resilience can break the cycle of destruction and recovery and reduce the impacts of earthquakes and other hazards. This requires community planning for recovery of function and setting of recovery goals prior to the occurrence of hazard events to minimize social and economic disruptions.

Investigations of lifeline system performance, interdependencies, and corresponding impacts on communities following earthquakes and other disasters around the world have highlighted the need for a new approach to improve the earthquake resilience of lifelines. In 2013, the National Institute of Standards and Technology (NIST) funded a project, led by the Applied Technology Council (ATC), that resulted in the NIST GCR 14-917-33 Report, *Earthquake Resilient Lifelines: NEHRP Research, Development, and Implementation Roadmap.* This report identified the need for assessing societal expectations of acceptable lifeline performance levels and recovery times at the community level as a high-priority research and development topic.

NIST funded a follow-on project in 2014 led by ATC to assess current societal expectations of acceptable lifeline performance levels and recovery timeframes, distinguishing those that are hazard independent and those that are specific to earthquakes (including tsunami), as well as other natural hazard events (ATC-126 Project). The team developed the NIST GCR 16-917-39 Report, *Critical Assessment of Lifeline System Performance: Understanding Societal Needs in Disaster Recovery*, which focuses on societal needs and interdependencies of six key lifelines: electric power, gas and liquid fuel, water, wastewater, telecommunications, and transportation networks.

The NIST GCR 16-917-39 Report identifies important gaps between expected lifeline system performance and the performance required to support adequately societal needs during and after a hazard event. Gaps were identified through an evaluation of performance and societal impacts during past events, as well as the assessment of key guidelines, standards and performance criteria that govern and shape the design, construction, operation and management of lifeline systems. The report identifies and discusses the social institutions and societal needs that should drive lifeline system performance levels and timeframes for recovery of function. Critical interdependencies, recommendations and needs particular to policy, modeling, research, and future trends for social needs and lifeline systems are included in the report. This report supports larger efforts, such as the Community Resilience Program and National Earthquake Hazards Reduction Program (NEHRP) at NIST, intended to help communities and lifeline owners and operators achieve a more integrated and consistent approach to resilience. The findings from this report, summarized in this paper, are intended to inform resilience efforts in the United States and other countries.

Keywords: Resilience; Lifelines; Social needs



1. Introduction

Resilience involves the ability of people and communities to adapt to changing conditions and withstand and rapidly recover from disruptions [1]. At the community level, this concept is complex and multi-dimensional, relying on contributions from social science, engineering, earth sciences, economics, and other disciplines to improve the ways communities prepare for, resist, respond to, and recover from disruptions resulting from either natural hazards or human causes. Resilience is intended to reduce the impact of hazards by restoring a community to normal function in a specified timeframe to minimize disruption, and in turn, reducing the duration and cost of recovery. This requires planning for recovery of physical, social, and economic functions and services prior to hazard events.

Disasters can substantially reduce a community's resilience by interfering with the operation of infrastructure, such as electric power, natural gas and liquid fuel, telecommunication, transportation, and water and wastewater systems (otherwise known within the earthquake engineering profession as "lifelines"). Lifelines are often distributed over large geographic regions and have numerous interdependent levels of operation, making them especially exposed to the impacts of earthquakes, hurricanes, and other hazards that affect broad areas. As a result, lifelines can experience distress and malfunction at many locations which, in turn, can impede response and recovery. Lifeline system failures and disruptions can displace people, disrupt social and economic institutions, and in the worst cases lead to death and long-term negative societal consequences.

In September 2014, the Applied Technology Council (ATC) commenced a task order project under National Institute of Standards and Technology (NIST) Contract SB1341-13-CQ-0009 to assess current societal expectations of acceptable lifeline performance levels and restoration timeframes that are informed by the phases of response and recovery, distinguishing those that are hazard independent and those that are specific for seismic (including tsunami), wind (including hurricane and tornado), flood, snow/ice, and wildfire hazard events.

The study was conducted as part of the NIST Community Resilience Program, a multi-year, multi-faceted program that is part of a broader disaster resilience effort that includes the oversight of the National Earthquake Hazards Reduction Program (NEHRP), which was established by the U.S. Congress with its passage of the Earthquake Hazards Reduction Act of 1977 [2]. NIST has served as the lead NEHRP agency since 2006, and has conducted numerous projects to meet NEHRP goals, including the effort described in this paper as well as the development by ATC of the NIST GCR 14-917-33 Report, *Earthquake Resilient Lifelines: NEHRP Research, Development, and Implementation Roadmap* [3], which identifies a wide range of needed studies and guides investments made by NIST and other NEHRP agencies in generating national performance and restoration goals and accompanying guidelines, manuals, and standards for lifeline systems and components.

Element I of the *Earthquake Resilient Lifelines Roadmap* identifies the priority research, development and implementation topics necessary for establishing national lifeline system performance and restoration goals. Topic 2 of Element I of the *Roadmap* specifically calls for a national assessment of societal expectations of acceptable lifeline performance levels and restoration times as a high-priority and a necessary part of the development of resilient design and construction goals for interdependent lifeline components, systems, and the communities those systems serve. While this roadmap specifically addresses resilience issues in the context of earthquakes, those priority research and development topics, including Topic 2, are valid for other hazards as well.

This paper includes consideration of the social institutions and societal needs that should drive lifeline system performance levels and recovery timeframes. It also provided an assessment of current system guidelines/standards and performance criteria. The assessment identifies deficits that could be potentially addressed through better awareness and definition of community requirements and goals as part of the overall Community Resilience Planning process [4].

The following sections discuss the project purpose, project development approach, principal findings, future considerations, and recommendations resulting from the study.


16th World Conference on Earthquake, 16WCEE 2017 Santiago Chile, January 9th to 13th 2017

The primary purpose of this study was to assess current societal expectations of acceptable lifeline performance levels and recovery timeframes that are informed by the phases of response and recovery, distinguishing those that are hazard independent and those that are specific for seismic (including tsunami), wind (including hurricane and tornado), flood, snow/ice, and wildfire hazard events. An additional goal of the study was to identify gaps between the desired and anticipated performance of lifeline systems to enable communities and lifeline system operators to set priorities and define strategies to reduce risks and improve lifeline resilience. In addition, this study was intended to provide a technical foundation for first-generation systems-based models that can be used to analyze community resilience and account for interdependencies among lifeline systems and the social and economic functions that they support.

The assessment described in this study is also intended to support the *Community Resilience Planning Guide for Buildings and Infrastructure Systems* [4]. The *Guide* defines a six-step voluntary process for engaging stakeholders and representatives to define key social and economic needs and community functions; develop an inventory of the community's building stock and infrastructure systems; establish performance goals for buildings and infrastructure to serve key social and economic needs over a range of different hazard types; identify and evaluate gaps between the desired and anticipated performance of buildings and infrastructure; and set priorities and define strategies to reduce risks and improve the resilience of buildings and infrastructure systems in their community. The *Guide* recommends that communities identify and plan for the prevailing hazard types (e.g., wind, earthquake, inundation, fire, snow or rain, and technological or human-caused) that may have a significant effect on their built environment. It also recommends consideration of a range of hazard levels defined as: *routine* (above the design level and sometimes referred to as the maximum considered event).

3. Project Approach

Over the last 43 years the Applied Technology Council has employed a staffing model that consists of: (1) ATC staff specialists in earthquake and natural hazard mitigation, who have lead responsibility for defining the project work effort, identifying and engaging project consultants, carrying out needed quality control efforts, ensuring that the project products incorporate the intended scope, and managing all financial matters; (2) highly qualified technical consultants from private design practice and the research community, who are engaged to conduct detailed technical research and development work required of the project; and (3) leading specialists from structural engineering practice, the academic/research community, and the regulatory community who are engaged to serve on an "blue-ribbon" review panel to ensure that the project work products are technically accurate, complete, and reflect a broad spectrum of professional opinion.

Project staffing included an ATC Project Manager and Associate Project Manager, a Project Technical Committee consisting of a Project Director (an urban planning consultant), Project Co-Director (an academic pipeline system specialist), and six academic and private-sector consultants (a social scientist, an electric power system specialist, a telecommunication system specialist, a transportation system specialist, a water/wastewater system specialist, and a researcher specializing in lifeline system interdependencies), a blue ribbon Project Review Panel consisting of six academic and private-sector consultants (a researcher specializing in risk and reliability, an electric power system specialist, a natural gas and liquid fuel system specialist, a telecommunication system specialist, a transportation system specialist, a telecommunication system specialist, a natural gas and liquid fuel system specializing in structural, earthquake and tsunami engineering).

The Project Technical Committee reviewed and summarized key lifelines guidelines, standards, and current performance criteria, and identified critical social considerations and system interdependencies. Those and subsequent efforts helped form a set of overarching considerations and project findings, which were then refined through several rounds of review by the Project Review Panel and NIST sponsors.

The culminating effort was the development of the NIST GCR 16-917-39 Report, *Critical Assessment of Lifeline System Performance: Understanding Societal Needs in Disaster Recovery* [5], which contains:

3



- Detailed analyses of a broad range of societal considerations;
- Lifeline assessments and reviews of standards, guidelines, and performance criteria for electric power, natural gas and liquid fuel, telecommunication, transportation, water and wastewater systems;
- A review and analysis of available information on lifeline interdependency issues; and
- Findings, conclusions and recommendations that identify needed developments in lifeline codes, standards, and guidelines; needed research; modeling opportunities; and needs related to lifeline system operations and operational design.

4. Principal Findings

The principal findings of the study address the current state of lifeline codes, standards, guidelines, and performance requirements in the United States; overarching societal considerations; critical interdependencies; disaster lessons (short-, intermediate-, and long-term recovery); key gaps and deficits; and future trends and considerations.

4.1 Current State of Lifeline Codes, Standards, Guidelines, and Performance Requirements

A survey was conducted of the codes, standards, guidelines and manuals that govern the design, construction and performance of key electric power, natural gas and liquid fuel, telecommunication, transportation, water and wastewater lifeline systems and system components. These documents were developed in a variety of ways to achieve functionality and safety, and to address system performance and reliability. They vary considerably from system to system and represent various levels of consensus—most typically among operators, regulators, and engineering experts. Codes for interdependent systems are also starting to emerge, particularly in the interoperability of automated electrical substations and telecommunication systems.

Some commonalities and trends have emerged from this survey, which are relevant for improvements in community-scale resilience. In particular, there are gaps in codes and standards for the infrastructure systems reviewed. Codes, standards, guidelines and manuals tend to emphasize component-level rather than system-wide performance. They tend to cover minimum levels of safety or performance (e.g., for extreme loading conditions) of components as opposed to system response and levels of service. Most address day-to-day operations and do not cover the range of hazard types considered in this study, particularly the extreme, low-probability, high-consequence events.

The processes by which lifeline codes, standards, and guidelines are developed tend to be reactive and based upon past information. This may result in significant time lags before the latest information on hazards is incorporated into updates. Furthermore, some standards and guidelines have not necessarily been developed following rigorous and systematic processes.

The performance requirements within lifelines codes, standards, and guidelines focus mainly on engineering for system design and construction and operational concerns, taking into account different, but more limited sets of societal considerations. Life safety, public health, emergency response, critical service provision, property and monetary loss prevention, and environmental protection are some of the most prevalent societal considerations identified in the performance requirements established by standards and codes for lifeline systems. In particular, failures in some lifeline systems (e.g., natural gas, liquid fuel, and transportation) and some lifeline service outages (e.g., power, telecommunication (911), natural gas, water, and wastewater) that can contribute to mortality and morbidity are given the greatest emphasis in codes and standards.

For electric power, gas and liquid fuel systems, certain measures of system performance are routinely assessed and even required to be reported to regulators. These measures generally address outages during normal operations and operations and often exclude disruption caused by hazard events, which can distort the true costs and impacts of system disruption and restoration. For example, electric power providers routinely report System Average Interruption Frequency Index (SAIFI), System Average Interruption Duration Index (CAIDI), Customer Average Interruption Duration Index (CAIDI), and Customer Average Interruption Frequency Index (CAIFI)— which generally exclude data during disruptions due to hazard events.



Most system performance measures are not directly informed directly by or linked to societal expectations, nor do they differentiate between customers or consider differential hardships imposed on society from varying durations of outage. The direct and indirect costs to customers will likely vary with the duration of system disruptions (i.e., mere inconvenience in the first few hours to severe hardship after weeks of lost service). One notable exception is the service level agreement (SLA) for telecommunication systems. SLAs establish thresholds for customer service and outages that trigger contractual penalties to the system providers. However, at present, disruptions caused by hazards such as earthquakes, hurricanes, and man-made threats are generally excluded from such agreements.

The lifeline assessments also identify a number of examples of system restoration guidelines that are generally informed by an array of societal expectations that include: public safety, mobility, business continuity, and environmental protection. System restoration times and guidelines tend to be governed by scenarios often derived from recent disaster experience or the largest disruption in the region. They may not take a robust view of the range of potential hazard types or levels of hazard severity that could affect the system.

The degree to which lifeline system codes, standards and guidelines address the provision of substitutions, temporary solutions and accessibility services (e.g., bottled water and portable toilets) for lost services also varies considerably by system. Such substitutions and temporary solutions (both within and between systems as well as those undertaken by society) can significantly affect performance outcomes and societal expectations, impacts and recovery following disruption.

Codes and standards are not the only influence on lifeline system design; factors such as voluntary measures to exceed standards, injunctions of regulatory authorities to improve performance in hazard events, and other risk management practices also influence the resilience of lifeline systems. Also noteworthy is the considerable variability in lifeline system regulation and enforcement by system, hazard, and geography (i.e., national vs. state, state-to-state, and locally). Natural gas, liquid fuel, and electric power all have regulatory data requirements focused on safety and reliability. In particular, for gas and liquid fuel pipelines, there is a nationally-consistent regulatory framework defined by federal legislation. Federal regulations also guide telecommunication systems, whereas, few states now regulate Internet-based telecommunications. However, while most US jurisdictions adopt the latest codes and standards and some even add more stringent requirements to them, other jurisdictions may not adopt them in their entirety or may reduce some requirements. Even if codes and standards are adopted, their effectiveness may be jeopardized by poor enforcement during planning, design and construction of infrastructure components. This disparity in code adoption and enforcement can have significant consequences for community resilience.

4.2 Overarching Societal Considerations

This study was challenging in part because of the almost total lack of empirical data on what members of the general public regard as acceptable lifeline performance following disruptions. Without such data, the study team had to make inferences about those expectations. Gauging societal expectations was also challenging because U.S. society is highly diverse. Expectations can vary across individuals, households, and businesses as a function of a number of factors, including: vulnerability and resilience characteristics; geographic location; hazard (including event severity, probability and duration); lifeline type; available information on the impacts of disruption; prior experience and knowledge about service disruptions; levels of resulting losses; public perceptions of the trustworthiness and competence of service providers; and the availability of substitutes and contingencies that can compensate for lifeline outages.

Public expectations and tolerances for lifeline service disruptions are dynamic and likely to be shaped by both risk perception and risk communication. Factors affecting risk perception include prior experience with hazards and lifeline outages, substitutability and dependency on lost services, and available information on the impacts of disruption. Other things being equal, disruptions may be tolerated for longer periods in severe and catastrophic events than in less serious ones, because the public will be more willing to accept that extreme hazard events are more difficult for service providers to anticipate and mitigate. Public confidence and the past performance of lifeline service providers (both during routine operations and prior hazard events) can also influence expectations and tolerances.



Public perceptions and expectations are also shaped by communications regarding the risks associated with past disasters and lifeline outages. Lifeline service providers should work with emergency management, public safety, and other governmental agencies to ensure that risk communication messages reach and are understood by the affected public. Additionally, there is evidence to suggest that societal expectations and tolerances may be changing, as social and economic activity becomes more dependent upon highly reliable service provision, particularly with regard to electric power and telecommunication systems.

Using terminology from vulnerability science and resilience research, various segments of the population and sectors of the economy are differentially exposed, differentially sensitive, and differentially adaptable to lifeline service disruptions. Because vulnerability and resilience vary as a function of exposure, sensitivity and adaptive capacity, risks associated with lifeline service disruptions are not borne equally by all members of society but are imposed disproportionately on already vulnerable social and economic groups.

One approach to assessing likely societal expectations is to look closer at how lifeline performance and disruption can have deleterious effects on what society members value most. This approach is consistent with the NIST *Community Resilience Planning Guide for Buildings and Infrastructure Systems*, which uses Maslow's "hierarchy of needs" framework to prioritize different building and infrastructure systems in communities. For this reason, human health and safety, the functionality of health-care systems and economic well-being are priority issues that this study has focused on in exploring societal considerations regarding performance.

The NIST *Guide* also distinguishes among short-term (days), intermediate-term (weeks to months) and long-term (months to years) response and recovery needs that planning activities must take into account. While recognizing that all three time periods are important, this study has emphasized societal considerations for the short-term period and immediate impacts, disruption, and responses, again with respect to human health and the economy. The rationale for this more limiting discussion is that: (1) life safety and health are overriding values for communities, with economic activity following as a community priority since it provides the means through which a wide range of other institutions and community needs are satisfied; (2) although the consequences of lifeline disruption can cascade over time, impacts such as deaths, injuries, and business interruption losses are generally most acute during the immediate post-impact period; and (3) practically speaking, more research exists on shorter-term than on longer-term disruptions.

4.3 Lifeline Interdependencies

This study also looked critically at the interactions among different lifeline systems during normal operations as well as restoration after hazard-related events. Dependent and interdependent relationships among lifeline systems have evolved over time with various systems and technology advances expanding and linking systems together. Although they provide efficiencies and other benefits during normal operation, they also introduce unique and largely unknown behaviors in hazard-related events and system disruptions.

Because of interdependencies, some system failures can also trigger significant cascading failures and performance impacts on multiple lifeline systems, causing a number of additional direct and indirect societal impacts. The physical proximity and colocation of lifelines can provide for enhanced efficiencies but also increase societal risks for cascading failures, complex interactions in restoration, as well as risks posed by multi-hazard effects. Collocated lifelines may also require a cross-system, cross-organizational and integrated approach to planning that is difficult to implement. Within systems there can also be choke points that amplify interdependencies. Interdependencies also exist between lifeline systems and community-level processes.

Rinaldi et al. [6] offer a general classification of interdependency mechanisms as physical, geographic, cyber, and logical, emphasizing that they are not necessarily mutually exclusive. Physical interdependence exists when the performance of a given network depends on the outputs of other networks. Geographic interdependence emerges from the colocation and close proximity of systems. Cyber interdependence occurs when the interdependence between two networks is based on shared information (e.g., the "smart grid" which relies on telemetry and situational awareness data). Finally, there is logical interdependence when systems are interconnected through channels different from those previously discussed, including human decisions related to restoration prioritization among systems.

16th World Conference on Earthquake, 16WCEE 2017 Santiago Chile, January 9th to 13th 2017



The body of research literature on interdependent critical infrastructure has grown over the past 20 years with risk assessment and system management dominating the themes of studies, followed by interdependency effects and cascading failures. Resilience-related studies of interdependent critical infrastructure are increasing but the volume of literature is still limited by comparison.

The literature emphasizes studies of electric power, water, transportation, and telecommunication systems in varying combinations. Studies involving natural gas systems and supervisory control and data acquisition (SCADA) technologies are becoming more prominent; while studies involving economic and social considerations, such as public health, have decreased in recent years. There is a distinct lack of research on cyber networks, network security and cyber interactions with other lifelines.

There is a growing body of research on system modeling pertaining to lifeline system interdependencies that can be leveraged to improve resilience across lifeline systems. However, the vast majority of lifeline interdependency studies and tools tend to focus on the technical aspects of systems and the impacts of hazards on them. More emphasis on risk reduction, system restoration, and societal impacts is needed.

There is also a lack of unified performance and restoration goals across multiple and interdependent systems. Rigorous and readily implementable theories and methods for the study of lifeline system interdependencies are similarly lacking.

The lifeline assessments developed in this study reveal a number of critical dependencies and interdependencies across lifeline systems. Virtually every lifeline system depends on electric power and telecommunications for system control and monitoring. All lifeline systems also depend on fuel and transportation, particularly for service restoration and system repairs. Fuel is a critical contingency for power when outages occur. Water is critical for cooling in the generation processes for electric power. Water also helps with pollution control, and supports other infrastructure, such as natural gas and liquid fuel systems.

The colocation of multiple systems also creates physical interdependencies that increase the likelihood that failure in one system can damage and interrupt others. Some prime examples exist with electric power and telecommunication infrastructure, and transportation elements that also serve as corridors for buried infrastructure (e.g., pipes and conduits for water, wastewater, telecommunications, power and natural gas). However, because lifeline service providers often do not share system information, many areas of physical colocation are unknown.

Some interdependencies are also increasing with time, such as the expanding role that telecommunications and electric power have in monitoring and remote control of lifeline systems, as well as household management, personal choices for renewable power, and community planning for decentralized energy.

4.4 Disaster Lessons

In general, there is a substantial body of literature on the social and economic impacts of lifeline service disruptions, spanning studies on actual events—both hazard and non-hazard related—and scenario-based and probabilistic loss projections. The lifeline assessments found great variability in both the quantity and quality of data available for lifeline systems, hazard types and severity, and across the various phases of disaster-related response and recovery. Most studies of disaster related impacts on lifeline systems are event-specific; systematic, multi-event studies are generally rare. Also, in general, there is more information and a more complete understanding of the societal impacts and restoration patterns of short-term rather than longer-term disruptions.

For electric power, most major and widespread outages have been caused by storms or other weather events. There is also more data on weather-related events than other hazards and more data on electric power outages in major hazard-related events, particularly with regard to technical aspects of component and system failure and restoration. Common information for electric power performance in hazard events includes: standard industry measures, peak number of customers without service, and time to restore service to all (or nearly all) customers.

Gas and liquid fuel production, transmission and distribution systems are susceptible to damage due to all forms of hazard events. Loss of power to oil refineries and pipeline pump stations during high wind and coastal

7



inundation events will result in loss of production and transmission, causing a lack of fuel supply for businesses and consumers. The high concentration of fuel refineries and major transmission lines along the central Gulf Coast and Eastern seaboard results in increased exposure to hurricanes. Ground faulting and liquefaction caused by earthquakes can lead to rupture of gas and fuel pipelines, though appropriate design measures can mitigate these effects as demonstrated by the good performance of the Trans-Alaska Pipeline System during the 2002 Denali Fault earthquake.

Transportation systems are susceptible to damage or disruption during earthquakes, tsunamis and other forms of coastal inundation or riverine flooding. Common earthquake damage includes bridge failures and landslides that can hamper emergency response, particularly to remote communities. Scour induced by coastal inundation and riverine flooding can also result in bridge, rail and roadway failures. Weather events, involving wind, snow and ice, can cause disruptions but physical damage tends to be more limited. Consequently, data on transportation performance in earthquakes, tsunamis, coastal inundation and riverine flooding are more readily available.

Water and wastewater systems are also more susceptible to damage or disruptions during earthquakes, tsunamis and other forms of inundation. Disruptions to water supplies can have serious impacts on fire-fighting capacity and sanitation that can result in serious public health, safety and economic consequences. Damage to wastewater treatment systems can lead to discharge of untreated sewage into waterways and the ocean, resulting in major environmental and public health concerns. The provision of accessibility services for these two systems is also essential for community-level resilience post-disaster.

In terms of recovery timeframes, disaster experiences show that most lifeline system outages generally last from hours to weeks (short- to intermediate-term recovery). In the most severe cases, outages can last for months or years. The long-term outages are associated with the most destructive events when critical and/or multiple components of lifeline systems that are time consuming to replace, such as buildings, bridges, piping, and essential equipment, must be reconstructed or replaced to restore system operability.

4.5 Key Gaps and Deficits: Codes, Standards, Performance Requirements, and Societal Considerations

Overall, this assessment found uneven study and treatment of societal considerations in the codes and standards governing lifeline systems performance for different hazards. Lifeline codes, standards, guidelines and manuals of practice are largely associated with the performance and safe operation of components rather than system response and levels of service. Even less common are efforts to consider interoperability issues across various lifeline systems (upstream and downstream interdependencies). In a few instances, system owners or operators have established system-level performance objectives or targets for a limited number of hazards. Most guidance does not address the range of hazard types considered in this study, particularly the low-probability/high-consequence events. In general, more is known about the lifeline impacts and performance objectives for earthquakes and wind-related events than for other hazards.

Strategies to enhance resilience will invariably require difficult investment decisions in the face of limited resources. To make rational decisions, lifeline owners and operators need to consider the tradeoff between risk reduction and financial investment. Quantitative uncertainty and risk analysis therefore need to be an integral part of performance requirements and criteria. In addition, risk communication needs to be an essential part of the overall dialogue, including public messaging.

Also, there is little information on societal expectations of system performance during routine, design, and extreme hazard events for most lifeline systems. Some potential "proxies" and informants for societal expectations of lifeline performance and recovery timeframes are available, including outage reporting thresholds and criteria, post-event reviews of utility performance, regulatory changes made following disruptions, societal and economic losses, and state and local energy assurance plans. Societal expectations may also be indirectly reflected in industry practices for service restoration following outages.

Currently there is also a lack of systematic research or accepted methods for quantifying or even bounding the gap between the disaster capabilities of lifeline systems and the societal expectations of their performance. This study identified some important examples for how the gap between prescriptive performance and societal

8



expectations can be filled, which include: community planning programs and stakeholder engagement in risk reduction planning and practices, as well as high quality system design, detailed risk assessments coupled with appropriate emergency planning, system monitoring, well-communicated emergency response procedures, and robust contingency planning and design.

4.6 Future Considerations and Trends

This study identifies a number of considerations outside its scope as well as important trends that are likely to shape future lifeline system design and performance and societal expectations. Societal expectations for lifeline system performance are changing, with decreasing societal tolerance for outages in general and lower thresholds for outage durations. Increasing interdependencies among lifeline systems, urbanization and the growing inventory of aging and deteriorating infrastructure increase the risk of system damage and disruption in future hazard-related events as well as the potential for cascading effects.

Future work both in research and applications related to community resilience needs to consider the following issues:

- Increasing social vulnerabilities due to an aging population, urbanization, and U.S. policies that are increasingly enabling disabled persons to remain in their homes where they are especially vulnerable to the loss of lifeline services;
- Economic vulnerabilities caused by the ongoing shift from a manufacturing to an information economy, growth of "cloud" based information technology, rising use of "just-in-time" strategies in inventory control and maintenance, and data storage, e-commerce and the expansion of the financial service sector, increases in telecommuting; and cost-cutting and efficiency measures in business and inventory management;
- Increasing vulnerabilities to distant lifeline disruptions caused by the continued rise of globalization, supply chains, and the interdependencies inherent in modern manufacturing and finance;
- Dependency of all lifeline systems on electric power and telecommunications for system control and monitoring;
- Increasing availability of renewable and distributed energy resources, grid modernization, decentralized power management, energy storage, and changes in the traditional structure of the electric power industry;
- Increasing physical and social vulnerability caused by the effects of climate change; and
- Lifeline system performance issues caused by theft of essential equipment (e.g., copper wire used in telecommunication systems), cyber-attacks, and other security threats.

Finally, the study cautions that unanticipated societal impacts will likely present themselves in future hazard events due to existing lifeline system vulnerabilities that remain poorly understood, absence of guidance on hazard resilient construction and installation practices, unknown and unfavorable states of repair, issues with system functional capacity, and ongoing changes in technology.

5. Recommendations

Recommendations resulting from this study are organized to identify the needs associated with: (a) codes, standards, and guidelines; (b) research; (c) modeling; and (d) systems operations. These needs target lifeline operators, practitioners, regulators, and researchers concerned about societal considerations in hazard-related system performance and potentially involved in lifelines standards and code development related activities. The following set of overarching recommendations was derived from the individual assessments of the study team. They were then discussed and prioritized with the Project Review Panel. All recommendations shown below were deemed high priority. Additional recommendations are provided in Chapters 2 through 8 of the NIST GCR 16-917-39 Report, *Critical Assessment of Lifeline System Performance: Understanding Societal Needs in Disaster Recovery.*



5.1 Codes, Standards, and Guidelines

This study revealed critical gaps in the codes, standards, and guidelines that govern the design, construction, operation, and performance of various lifeline systems and system components. Ten recommendations are offered to address needs related to lifeline codes, standards, and guidelines. The priority rankings reflect organizational and framework needs, available information, new knowledge needs, guidelines and standards development needs, and scoping breadth, with recommendations that pertain to broad issues and improving community resilience considered higher priorities than recommendations for specific lifelines.

- A1. Identify or establish an organization and process for advocating, harmonizing and unifying the consensus procedures for lifeline guidelines and standards development.
- A2. Develop more consistent terminology for lifeline standards.
- A3. Develop an up-to-date and complete suite of codes, standards, and guidelines for all lifeline systems to reflect the current state of practice, knowledge, and performance requirements.
- A4. Develop a methodology to combine component-based design criteria into system level performance targets.
- A5. Develop lifeline system performance requirements that relate to community resilience and better reflect societal considerations.
- A6. Develop consensus-based guidelines and standards for the design of new lifelines and the retrofit of existing lifelines that reflect community resilience performance requirements and societal considerations.
- A7. Develop guidelines to inform the design, interoperability, and upkeep of lifeline system dependencies.
- A8. Reduce inconsistencies in the compendium of codes and standards that guide design, construction and resilience of the built environment, such as fire codes, building codes, and lifelines codes, standards, and guidelines.
- A9. Develop consistent policy and standards on accessing information and databases about critical infrastructure systems that is coordinated with Department of Homeland Security critical infrastructure activities.
- A10. Provide updated guidance for evaluating gas and liquid fuel pipeline and facility response to seismic hazards, floods, coastal storms, and tsunami-related inundation.

5.2 Research

The study identified a number of gaps in data and knowledge necessary to improve the fundamental understanding of acceptable lifeline performance. Fifteen recommendations with respect to systematic study and research needs are offered. All are considered high priority, with those encompassing all lifelines and broader topics listed before those related to only one lifeline.

- B1. Gather information on and systematically study the relationships between service disruptions, and societal impacts and expectations to better understand lifeline system performance.
- B2. Develop and conduct a targeted research program to assess societal expectations associated with lifeline system performance.
- B3. Systematically study and compare the array of design approaches and methods for addressing societally-based performance requirements within current codes, standards and guidelines for lifeline systems.
- B4. Investigate the differential vulnerability among social groups to lifeline system outages.



- B5. Systematically collect and review various "proxies" and secondary evidence for societal expectations of lifeline performance and restoration timeframes.
- B6. Assess the various lifeline performance programs and practices for public safety and develop guidance on their application to other critical lifelines, including multiple, interdependent systems and collocated facilities.
- B7. Conduct research on needed service restoration times, including how system operability as a performance metric supports community resilience.
- B8. Study lifeline system operator organizational issues and how they affect community-scale lifeline performance and resilience planning.
- B9. Enhance the understanding of infrastructure-related failures and cascading effects resulting from low-probability/high-consequence events.
- B10. Develop post-event data collection protocols to assess lifeline system recovery and restoration timeframes and improve the understanding of restoration processes across individual and interdependent lifeline systems.
- B11. Develop tools to identify interdependent infrastructure systems and services along with their restoration criteria.
- B12. Establish procedures to quantify hazards for spatially distributed systems.
- B13. Enhance the understanding of lifeline system supply sources and end-point facilities and their role in system performance, restoration, and community and regional recovery with the goal of improving databases and modeling of such sources and facilities.
- B14. Perform studies on changes in water demand considering an array of hazards as well as seasonal and longer-term climate variability, like drought.
- B15. Improve knowledge, databases and modeling for the impact of widespread flooding and storm damage on regional fuel supplies.

5.3 Modeling

There is a growing body of system modeling for lifeline systems and their interdependencies that can be leveraged to improve resilience across lifeline systems, but there are also notable limitations in scope, outputs, integration, and validation that need to be addressed. Three modeling related recommendations are offered. All are considered high priority.

- C1. Aggregate the existing suite of infrastructure modeling tools and create a user-friendly interface so communities can properly assess their lifeline-related system performance and restoration risks, including uncertainty.
- C2. Develop first-generation models and practical tools to analyze community resilience that account for lifeline system dependencies and interdependencies.
- C3. Improve numerical modeling of lifeline systems, with emphasis on validation of models, developing the most effective simulation procedures, and applications in real systems.

5.4 Lifeline System Operations

The study also identifies a number of needs related to lifeline system operations and operational design. These too must be addressed to improve community resilience and bridge the gap between the post-event capabilities of lifeline systems and the societal expectations of their performance and restoration. Five recommendations are offered. All are considered high priority.



- D1. Develop a process for major utilities to conduct self-assessments of their preparedness for various natural hazard events, as a basis for prioritizing improvement to system robustness and post-event response.
- D2. Develop guidance for lifeline service providers on how to engage and collaborate with communities, including emergency management agencies and other key community institutions, in developing resilience strategies and preparing system restoration and contingency plans.
- D3. Develop guidance for local planning (e.g., for fuel delivery to emergency responders and critical infrastructure).
- D4. Develop guidance for lifeline service providers to evaluate the effects of system component failures, both in isolation and in combination, and considering upstream and downstream dependencies.
- D5. Design protocols for lifeline service providers, working with emergency management and other community institutions, to communicate to the public the likely impacts of different hazard events on service provision and disruption.

6. Acknowledgements

The authors of this paper, who led the development of this study (L. Johnson, Project Director; T. D. O'Rourke, Project Co-Director; C. Rojahn, Project Manager; V. Cedillos, Associate Project Manager; T. P. McAllister, NIST Sponsor; and S. L. McCabe, NIST Contracting Officer's Representative), gratefully acknowledge the significant contributions of the members of the Project Technical Committee, which consisted of Stephanie Chang (University of British Columbia), Craig A. Davis (Los Angeles Department of Water and Power), Leonardo Dueñas-Osorio (Rice University), Ian Robertson (University of Hawaii at Manoa), Henning Schulzrinne (Columbia University), and Kathleen Tierney (Natural Hazards Center, University of Colorado), and the Project Review Panel, which consisted of Bruce Ellingwood (Colorado State University), Timothy J. Lomax (Texas A&M Transportation Institute), Douglas J. Nyman (D. J. Nyman & Associates), Dennis Ostrom (City of Santa Clarita, California), Jon Peha (Carnegie Mellon University), and Kent Yu (SEFT Consulting Group and ATC Board Representative).

7. References

- [1] Executive Office of the President (2011): *Presidential Policy Directive/PPD-8, National Preparedness*, The White House, Washington, D.C.
- [2] USA Public Law 95-124 (1977): Earthquake Hazards Reduction Act of 1977.
- [3] NIST (2014): Earthquake-Resilient Lifelines: NEHRP Research, Development and Implementation Roadmap, NIST GCR 14-917-33 Report, prepared by the NEHRP Consultants Joint Venture, a partnership of the Applied Technology Council and Consortium of Universities for Research in Earthquake Engineering, for the National Institute of Standards and Technology, Gaithersburg, Maryland.
- [4] NIST (2015): Community Resilience Planning Guide for Buildings and Infrastructure Systems, Vols. I and II, NIST Special Publication 1190, National Institute of Standards and Technology, Gaithersburg, Maryland.
- [5] NIST (2016): Critical Assessment of Lifeline System Performance: Understanding Societal Needs in Disaster Recovery, NIST GCR 16-917-39 Report, prepared by the Applied Technology Council for the National Institute of Standards and Technology, Gaithersburg, Maryland.
- [6] Rinaldi SM, Peerenboom JP, and Kelly TK (2001): Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems Magazine*, 21 (6) 11-25, Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

Issues in Synthetic Data Generation for Advanced Manufacturing

Don Libes National Institute of Standards and Technology Gaithersburg, MD, USA don.libes@nist.gov

David Lechevalier Le2i, Université de Bourgogne, Dijon, France david_lechevalier@etu.u-bourgogne.fr

Abstract— To have any chance of application in real world, advanced manufacturing research in data analytics needs to explore and prove itself with real-world manufacturing data. Limited access to real-world data largely contrasts with the need for data of varied types and larger quantity for research. Use of virtual data is a promising approach to make up for the lack of access. This paper explores the issues, identifies challenges, and suggests requirements and desirable features in the generation of virtual data. These issues, requirements, and features can be used by researchers to build virtual data scientists while avoiding known or potential problems. This, in turn, will lead to better requirements and features in future virtual data generators.

Keywords- virtual data; synthetic data; data generation; smart manufacturing

I. INTRODUCTION

Much of the research in advanced manufacturing involves the creation of models and simulation-based experimentation. Simulation leads to significantly faster results than physical experiments that require use of real machine tools and materials in a physical shop floor [1]. Such simulations and models require data for carrying out those experiments. These data can be representative of a large number of sources such as machine tools, robots, suppliers, etc. The data types are also quite varied such as material density and strength, machine tool wear and energy usage.

The concept of simulated data has been referred to in multiple ways in the literature. Many people use terms such as artificial data, virtual data, generated data, fake data, and synthetic data to all mean the same thing. In contrast, people denote data generated by physical machines as real data, physically-generated data, or live data. We (the paper authors) use both sets of terms interchangeably in speech but prefer synthetic data and real data in the written word. The term live data is occasionally used to emphasize physicallygenerated data that is being consumed within a small window of its generation. However, this can be confusing so we generally avoid that term or make clear in context what we mean. Sanjay Jain The George Washington University Washington, DC, USA jain@email.gwu.edu

A. Background

Synthetic data generation has received significant interest in recent years across a number of fields. A recent review [2] of the literature related to synthetic data identifies use in multiple fields including economics, urban planning, transportation planning, cyber security, weather forecasting, and bioinformatics. Interestingly this review didn't report any activity in manufacturing. The paper focused on synthetic data generation for learning analytics in the education environment and identifies some of the same issues that are discussed in relation to advanced manufacturing in different situations such as machinelearning training, or extension of real data set. Similarly, [3] focused on synthetic data generation for Internet of Things (IoT) environment to address some of the same challenges including limited access to real world data discussed here.

There have been papers that tangentially touch on synthetic data generation but with a primary focus on simulation. For example, [4] describes the European Virtual Factory Framework (VFF), an interoperability framework for factory modeling. The VFF includes a Virtual Factory Data Model (VFDM) for common representation of factory objects to evaluate performance of production systems and a Virtual Factory Manager to manage a shared data repository. Support for the external simulation tool, Arena, demonstrates the ability of data generation and potential for improved data interoperability [5].

The National Institute of Standards and Technology (NIST) has two efforts specifically in the area of data generation for manufacturing. The first is the STEP2M Simulator which simulates machine monitoring data, given process plans [6]. The process plans are Step-compliant data interface for Numeric Controls (STEP-NC) conforming [7] and the resulting generated data are presented using MTConnect [8]. A second project defined a virtual machining model that simulates a 2-axis turning machine [9]. This model uses STEP-NC [10] commands and material property as inputs, and uses equations to compute time and power consumption depending on the tool path. The simulation generates machine-monitoring data in MTConnect format usable for other simulations. In [11], the authors demonstrated the usability of these data with a 3-axis milling operation. They

integrated the earlier work into an agent-based model to provide capabilities to create a shop floor model [12]. More complex shop floor models could be built combining different agent-based models representing different operations. Such a combination would lead to the generation of data at the machine level but also at the shop floor level where data could be aggregated.

II. WHY IT IS SO HARD TO OBTAIN REAL DATA

There are many reasons why it can be difficult or impossible to get sufficient real data both by type and quantity. Libes et al describe how significant amounts of data exist but are nonetheless unusable [13]. For example:

- Data may be proprietary preventing access.
- Data may have timestamps but are not synchronized preventing data joining.
- Data frequency may be insufficient.
- Data may be formatted inadequately leaving ambiguity.
- Data may have inexplicable gaps.
- Data may have been generated with different underlying goals.
- Metadata may be inadequate or non-standard resulting in semantic confusion.
- Accuracy may be undocumented.
- Data provenance may be suspect (modified without documentation) or unknown.
- Data may be too costly to obtain.

We lack the space to explore all of these in detail but we will consider one as an example: data formatting.

Enterprises sit atop a vast collection of disparate data, likely produced by a multitude of heterogeneous sensors, and often ultimately stored in files formatted according to a variety of standards, with varying degrees of compliance. Significant amounts of data may follow no standards whatsoever. Using standard specifications such as XML (eXtensible Markup Language) and JSON (JavaScript Object Notation) can help [14]. However, problems such as under specification can remain and leave ambiguities. For example, using an XML attribute called "time" means little if there is no definition for how the string is to be interpreted: absolute with respect to UTC (Coordinated Universal Time) or local time zone? Is the time relative to the start of a process or something else?

While standards can be helpful, they are not panaceas. For example, equipment and software from different vendors may use different standards. Data, while still standardcompliant, can lose fidelity during interchange. Standards frequently have different levels of compliance that users may choose. Even highly specific standards do not guarantee data are usable. For example, the machine tool standard, MTConnect, covers only one direction of communication; so, correlation to commands may not be present or need reconstruction with timing uncertainties. Equally important, MTConnect does not cover all possible types of data that a machine tool can generate. For example, MTConnect defines a fixed set of statistics for DataItems. Kurtosis, a measure of peakedness relative to a normal distribution, is in the set. Skewness, a measure of symmetry, is not in the set.

All of these choices of standards and specifications have reasons for existence. For example, different vendors may have different reasons for their choices. These can include historical issues, expense in tracking developing standards, and interactions with other software. For example, the choice of OWL (Web Ontology Language) variants depends on how much need there is for expressiveness – the breadth of concepts that can be represented [15]. But greater expressiveness brings with it a loss of computational guarantees [16].

Data standards may be descriptive (describing practices) or prescriptive (defining practices). Each carries with it downsides. For example, descriptive standards may prevent the use of innovative techniques that are too new to be incorporated in standards while prescriptive standards may be ignored when better technology solutions are discovered. These dilemmas are particularly apparent in rapidly changing and highly-competitive fields. To allow variations and technological advances, some standards intentionally leave areas of ambiguity with a resulting ambiguity in the data.

III. USES OF SYNTHETIC DATA

Development and maintenance of data analytics applications, models, and real factories can all make use of synthetic data, albeit in different ways. For example, data analytics applications can use synthetic data to test that training algorithms perform adequately. Factories can also use the data to experiment with proposed changes. For instance, it is not possible to test a replacement machine tool before it is purchased and installed; however, data synthesized by a machine tool simulator can allow the factory to be modeled and tested as if the machine tool was present. Similarly, policy or algorithm changes can be tested before deployment in a real factory. Lastly, synthetic data can be a resource for test suites that exercise the full range of states of a system – including both normal operation and error conditions.

In short, synthesized data can be used throughout the lifecycle of a factory – from initial brainstorming to development to maintenance. We focus on several uses of synthetic data in this section.

A. Machine-Learning Training

Machine learning models are not explicitly programmed or based on a physical model but primarily by analyzing data. For example, neural networks are trained on data and typically produce networks that make no attempts to model the physics. There is no use of equations that correspond to performance of any real machines. An appropriate amount of data is necessary to make sure that machine learning algorithms well perform.

While defining the right amount of data is one area of research, it is equally important to understand the limitation of synthetic data in this context. For instance, a neural network that is trained on synthetic data is unlikely to provide a better model than the model used to generate the synthetic data in the first place. For this reason, users (e.g., neural networks) of machine-learning algorithms should track the provenance of any generated data used to train those algorithms. Without such provenance, mistaken assumptions may arise over the quality of decisions generated by machine-learning algorithms.

B. Verification

The American Society of Mechanical Engineers (ASME) defines verification as the process of determining that a computational model accurately represents the underlying mathematical equations and their solution [17]. During development, verification is used to ensure that data analytics applications and models meet their design requirements and specifications. While direct methods are available in limited ways for verification, data can be used in verification, in some cases as input to be consumed and in some cases as output to be tested against.

Being able to generate indefinite amounts of data allows more extensive testing than would otherwise be possible through limited reference data produced by physical machines. In addition, parameters can be changed to generate different synthetic data that would be very expensive or impossible to obtain by using physical machines.

C. Validation

ASME defines validation as the process of determining the degree to which a computational model is an accurate representation of the real world from the perspective of the intended uses of the model [17]. Validation confirms that applications or models match the needs of the customer. Data analytics applications and models should be validated – ideally, during development and continuing through the lifecycle. Data are useful in this validation process, particularly in cases when physically realized applications, models, and physical instances will not yet be available. Having synthetic data – both input and output – as if the enterprise existed, can be used to test that requirements are valid. Throughout the life of applications, models, and real factories, synthetic data can be used to validate that proposed changes continue to result in valid results.

D. Optimization

Optimization is the process of improving algorithms or applications to produce the best results, and improve the system under study. (This term is often mistakenly used for the process of "improvement" which is more practical while "optimality" is aspirational.) Optimization is frequently performed by simulation.

Synthesized data can be used for optimization, such as in factory models. These models are not algorithmic machinelearning models (see earlier) but rather simulations of factories (or subsets) that can be tested for performance optimality. For example, work cell and machine tool arrangements of a factory floor can be simulated and tested prior to deployment in order to select the best layout. Since such simulations have no real counterpart, synthesized data can be used in the simulation. For example, part flow, machining time, machine breakdown, etc., can be used during simulation runs to produce and measure more realistic performance metrics.

E. Augmenting Real Data

For many reasons, real data can be missing or insufficient. For example, sensors may be incapable of collecting data quickly enough. Many techniques are possible to replace missing data values. For example, simple mathematical interpolation can produce data that fill gaps and reflects existing data. However, interpolation can distort data. This may sound counterintuitive but means that interpolated data give the mistaken impression that the data are smoother than it actually is.

Analytic applications can automatically fill in missing data values by essentially running algorithms in reverse producing data that is more representative of real data. For example, once a neural network has been trained on a data stream, the neural network can be used to produce more data. However, by its very nature, such synthetic data will not have any impact on analytics rendering it of no intrinsic value except perhaps for pro forma purposes such as creating more complete visualizations.

Missing data can be used during development (for example, for optimization) but is also useful when a factory is in production. In that case, there may be instances or periods when data are unavailable whether due to a broken sensor or communications problem. In this case, missing data can be replaced with synthetic data in real time.

IV. CHALLENGES FOR DATA GENERATION AND COLLECTION

While synthetic data have many uses, challenges exist, many of which may not be obvious until encountered. Thus, we describe some of most significant challenges in data generation.

A. Quantity Sufficiency

Algorithms for synthesizing data require some minimum amount of realistic data although the data may be of a different type. For example, generation of energy data may require a machine specification and material data. Some algorithms may generate data by using data similar to what is desired – for example, by increasing or decreasing the variance of an existing dataset.

Data analytics algorithms generally increase accuracy with more data when that data cover more of the descriptive space to be analyzed. Of course, more data can slow down algorithms worsening time performance. More data can also be unnecessary, essentially providing no new information. More data can even introduce artifacts that are irrelevant and hurt model performance. However, sufficient quantities of training data can be significant, for example during walkforward testing while training neural networks. Knowledge of the quantity required for training can be as useful as the choice of the data itself and which data are more likely to be seen in the real system. Defining the right quantity might be possible by calculating the squared error of the data analytics model while training and testing to define if it is over or under fitted. Defining the right quantity of data before training is, however, a more complex issue.

B. Timing

Closely related to quantity sufficiency are issues of timing. These issues can be complex and be impacted by many parts of an enterprise but sensor behavior is the simplest place to explore such issues.

Sensors operate in two modes. Some sensors run free meaning that as soon as they have finished reporting data, they begin collecting or sending additional data. They are constantly busy. This can be useful for algorithms that want to consume as much data as possible. Other sensors may be synchronized to an internal or external source. Such sensors artificially discretize readings. Being able to synthetically simulate such data sources can be useful for properly modeling real-world factories.

Generators may need to be capable of generating both types of data sources with the ability to generate data at arbitrarily large rates.

C. Dynamic Generation

Data analytics developers often do not need or want to store all data in advance of the need. The amount of data needed may be too large to store. Generating data dynamically can also better reflect the way data are consumed as well as leaving open the possibility for feedback that can change the data generation process. The converse is also true. Consider data coherency. Data coherency can be disrupted by updating data at the same time it is being consumed but without synchronization. By avoiding the storage of large amounts of synthetic data, data coherency is maintained with lower costs (e.g., locking protocol overhead). For these reasons, it can be useful to generate and immediately consume data dynamically whenever possible.

On the other hand, storage of large data sets may be useful for certain types of analysis. When time and space are not significant factors, data can be evaluated more thoroughly. Resulting models that are relatively static or are already optimal do not need to be continuously or frequently retrained if there is no benefit to doing so.

1) Feedback and Control

Real-time performance applications make decisions that are fed back to the system thereby affecting its performance. Naturally, changing the outcome of such decisions changes the performance of the system which in turn changes any performance data produced by the system that would ordinarily be fed back to the analytics applications.

It is desirable for data generators to be able to incorporate feedback in order to control future data generation. This is not necessarily easy as feedback control can introduce nontrivial synchronization issues. For example, lock-step synchronization is typically not necessary as there are lags in real factories when data-driven decisions affect controllers and similarly how quickly sensors can return feedback. Determining at what levels this is accounted for (e.g., generators, models, simulations) and with what accuracy can be challenging.

2) Speed

For physical data generation, data are generated at whatever speed the producers can generate it. For data analytics applications, that speed is generally not of interest. A data analytics developer is not consuming real data directly from a shop floor except in rare circumstances. Developers almost always use models or simulations that can run faster than real-time. For that reason, when data are generated dynamically, it is desirable to have data generators that can run faster than real-time, preferably as fast as the simulation itself.

In some contexts, data generation may need to be slowed down to real time to ensure that applications are able to provide feedback in a timely manner to impact factory performance. For example, if batch dispatching decisions occur every minute and an analytic application promises to improve system performance through better dispatching, the application should be able to execute and respond within sub-minute intervals. The capability of such an application to perform well in a real system should be tested with a manufacturing simulation that ensures real time performance.

D. Data Hiding / Suppression

For a variety of reasons, it can be useful to suppress or hide (i.e., to not use) synthetic data during analysis (or development and testing of analysis). It may be desirable to hide details of the implementation as well. Some of these considerations are presented in this section.

1) Intentional hiding

For various reasons, more data may be available than is desired. Thus, it may be useful to hide some of the data. For example, the analytic engines may be incapable of consuming all the data, particularly if time limits are an issue. This is a concern with the need for real-time results. Data sampling might be a solution to achieve this task without modifying the dimensions represented in the data set.

2) Walk-forward testing

Walk-forward testing consists in training a machine learning model with a subset of the data, and testing the trained model with an unseen subset of the data. It is another example where data must be hidden – at least initially. This may be repeated on many quantities of unseen data in order to ensure that systems are not over-trained. The aim is to create a model that is not necessarily expert at recognizing only the training data but capable of recognizing data that is also likely to be produced in similar scenarios.

3) Filtered

Data may be hidden because it is withheld (filtered) for a variety of reasons. For example, data that is clearly wrong or exceeds certain bounds may be suppressed depending on the needs of the data analysis applications. An analytics application may be designed to run in a live system where data are guaranteed by database constraints as to its quality. During development, guarantees may be maintained when the generator is providing the data without any database filtering. Describing what should be filtered can be arbitrarily complex as there are many reasons for filtering and the reasons can be combined in complex ways. For instance, analytics software under development might be restricted from out-of-bounds data (previous example) as well as malfunctioning sensor readings that are within bounds but inaccurate.

4) Black box generators

A data generator or its underlying models and data may be treated as a black box. Hiding the implementation can prevent certain types of analytics that may give an unwarranted impression of a tool that is more universal than it is, simply because the implementer (or software) can "see" the implementation and use the implementation rather than the data as a basis for analytics.

Data hiding also prevents premature optimization as well as shortcuts that can overlook problematic data. For example, an optimizer that knows it will never see data outside a certain range may learn that it is not necessary to handle such data. When it is faced with such unexpected data, perhaps by a misbehaving process upstream, the optimizer is likely to behave inappropriately since it has not been trained on such data.

E. Data Quality

Sensors quantize data, have lags, fail, and have other issues. So, ideal data should never be expected from real enterprises. However, there is value in building ideal data generators. For example, certain types of algorithms require or perform much better when optimal goals are provided. This is the case with many non-heuristic approaches to nondeterministic polynomial (NP) problems such as identifying optimal routing [18]. Nonetheless, most interest for data analytics is in creating more realistic data. The following sections describe data quality issues in data generation.

1) Reliability

Physical systems can be unreliable so it is useful for synthetic data to be able to reflect that. This unreliability can be difficult to model as misbehavior can come about in so many ways. For example, sensors can behave erratically, communication can encounter interference, or power can be dirty. Each causes reliability issues.

Within each type of problem, there is a spectrum of unreliability. Reliability can also change over time, typically increasing but occasionally decreasing. In short, reliability is complex so generating realistic unreliable data is complex.

2) Accuracy

The accuracy of physical sensors must be accounted for by data generators. While an accuracy limit may suffice, it is more realistic to produce a range that models the physics of the sensor; however, this is often not possible. Instead a variety of ranges are used such as a Gaussian distribution, Poisson distribution, or Bayesian-based distribution.

3) Uncertainty

There are many sources of uncertainty. For example, machine-tool manufacturers themselves may not have a usable mathematical model for their products. Even if they do have a model, it may specifically omit aspects that are difficult or entirely unknowable, for instance. So for these and other reasons, physical device manufacturers often state only a range of accuracy leaving questions of distributions a difficult challenge between the designer of a generator which mimics that device and the user who configures the data generator. Quantifying and aggregating (epistemic and aleatory) uncertainty generated by different sources during the simulation is complex and requires to clearly identify and evaluate each source of uncertainty [19].

4) Adjusted

Sensor data can often be adjusted at intermediate processing nodes. For example, data can be joined at a network node that collects data from several sensors for each entry in a log. Adjustments can include normalization, scaling, and quantization. This means that analysis software may only see adjusted data rather than raw data. For this reason, data generators should be able to produce either raw data or data adjusted in a variety of ways.

F. High-Level Key Performance Indicators Without Low-Level Data

In many scenarios, the company is interested in highlevel (i.e., enterprise or shop floor level) Key Performance Indicators (KPIs) or data and not in low-level (i.e., process or machine level) KPIs or data. For example, machine tool measurements may not be significant to higher-level processes such as machine-tool inventory predictions despite the basis on machine tool information. Many analyses only use high-level KPIs as input such as finished product cost, inventory cost, and goals such as minimizing late orders. Once the KPIs are created, the low-level data are never again used. In such scenarios, the low-level data are not necessary if the KPIs can be created independently.

Data generators that provide high-level KPIs may not need to model the underlying physics if the KPIs are good enough. Of course, "good enough" is challenging to define. Generating KPIs in a top-down approach (for example, driven by organizational goals) may be harder than generating low-level data and doing a higher-level simulation to arrive at KPIs. Whether this is true depends on the fidelity requirements of the KPI-consuming applications [20].

G. Well-Described Scenarios

To generate data that are useful, scenarios must be identified to define the context in which the data applies. Finding and describing such scenarios can be difficult for several reasons. There are many variables that likely differ for every factory: products, machine tools, goals, and costs. These are almost always mixed. For example, a manufacturer may have a mix of machine tools and the number, types, and layout will differ from one manufacturer to another. Similarly, one manufacturer's goals are likely to differ from another. One manufacturer may have contracts with suppliers and utilities while another manufacturer will have different suppliers and other constraints. Machine tools will be of different ages and exhibit different performance characteristics. Where semi-automation is an issue, human factors will be unique as well.

To accommodate differences such as constraints or goals, data generator parameters can be adjusted; however, thought should be given to whether these parameters are intrinsic to data generation. For example, while both machine tool performance characteristics and goals (such as "use minimal power") will affect the generated data, the former is intrinsic to the generated data while the latter can be considered a dependent variable that is only meaningful to a higher level of control.

H. Manufacturing Levels

It is customary to organize a hierarchy of manufacturing at different levels of operation, e.g., models for machine tools, workcells, factories, enterprises. Additionally, customers and supply chains may also be modeled. Developing simulators or analytics applications may be specific to a level or may include multiple interacting levels such as model supply chain level and its interaction with the factory level [21].

Data at any one level are likely to have a strong dependence on other levels of a hierarchy. Obviously, a factory model depends greatly on the performance of the operations within. Less obviously, a machine tool depends on the goals of its workcells or even higher levels. For example, a machine tool may wear less by running at a slower speed which is only acceptable if the goals of the factory or workcell permit.

I. Data Type Complexity

Many different types of data exist in a real factory. For example:

- Material Data: costs/characteristics/physics of material, water, energy
- Process Data: task time, customer demand, production schedules
- Product Quality: geometry, structural integrity, performance
- Manufacturing equipment: efficiency, reliability, spare capacity
- Employee data: salary, hours worked, employee skills

It is possible to generate all of these but the more types of data generated, the more work is required. For practical purposes, many of these are unnecessary and only depend on the particular goals of the analysis. Not all goals can be achieved simultaneously as many will always conflict. For example, it is generally impossible to achieve both minimal energy usage and time. There is always a tradeoff. For the same reason, there is no minimally optimal data set. The scope of data sets varies based on the objectives of the analysis and the level of abstraction desired by the analyst.

J. Repeatability, Reproducibility, and Provenance

Repeatability, reproducibility, and provenance are closely related, all having to do with the confidence in the ability to recreate generated data.

1) Repeatability

Repeatability of data generation from a physical enterprise is difficult. For synthetic data generation, repeatability is generally straightforward. Generators must be capable of publishing and accepting random number seeds. Generator code must not require anything else that could change the data output from one run to another. The code itself must be published in such a way that the code remains the same so that others can rerun the same data generator with identical results. This can be ensured with a signing procedure such as providing the results of a cryptographic hash function.

2) Reproductibility

While repeatability is straightforward, reproducibility can be more difficult since underlying libraries and computer hardware can cause output differences despite identical highlevel code. Languages such as C are particularly notorious for this. These differences are not necessarily a bad thing but intended to give programmers more control over efficiency of an implementation. However, programmers are free to ignore (or may be unaware of) such subtleties which can lead to non-reproducibility. C is not alone. Many higher-level languages have subtle dependencies as well. In [22], the authors review correct approaches to assess the consistency of measuring process. These approaches can be similarly applied to data simulation to ensure consistent generation of data.

3) Provenance

Some data may not need reproducibility as long as it is suitable for its purpose and its provenance can be ascertained. Provenance through a digital signature provides a guarantee of the source of the data and that it has not changed. Only the signing of data and its metadata is necessary.

Closely related to provenance is traceability. While data may be provably shown to have come from one data supplier, it may be of no value if others cannot ever hope to build a factory that reproduces it.

K. Model Type Choice and Provenance

There are a variety of model types used for data generation. There may also be hybrid combinations of these models.

1) Physics-based Models

Physics-based models are based on equations that reflect our understanding of what physically happens in real life. In theory, physics-based models are the optimal way of modeling any manufacturing process since they capture all information and effectively produce our best understanding of a process.

However, these equations can be difficult to obtain, relying on, for example, a machine tool manufacturer which may have little incentive to provide specific types of equations or explanations. Alternatively, the equations may exist but fail for a variety of reasons to correlate with what is observable. For example, machine tool wear may be a factor in the equation but is effectively unmeasurable because it requires destructive testing or machine disassembly that voids a warranty.

For practical reasons, it is rarely the case that we truly have accurate equations that hold in all situations. For example, it would not be sensible to have quantum-level models when Newtonian models are sufficient for almost all purposes. Similarly, while physics-based models are often implemented using continuous simulation, the ability to achieve arbitrary levels of resolution and scale is generally overkill for most data generation needs.

The result is that physics-based models are always idealized and require adaptation to be used for realistic data generation.

2) Empirical Models

Empirical models are based on data, ideally from physical systems. These data are then used to build machinelearning models using techniques such as neural networks and Gaussian networks. Such empirical models are generally used with discrete simulation techniques although this is not mandatory.

The result is that empirical models are adequate to provide good data generation for some situations. However, for situations that they have not been trained for such as edge cases and other unusual events, empirical models can fail to generate suitable.

3) Special-purpose Models

Special-purpose models can be based on other techniques besides those based in statistics or physics. For instance, a data generator could be used to produce intentionally bad data specifically to test the behavior of analysis tools. It may suffice to use a stream of zeros or just open a file of random data. Similarly, an ideal data generator (see 4.5) can be used to test minimal fitness or conformance. Such data can be based on published or nominal specifications from a manufacturer with no regard to a statistical or physical model. Data could also be aspirational, referring to a goal for which no known algorithms can achieve but represent provable limits.

L. Integration, Interoperability & Standards

While it is important that data be meaningful, it is also important that data are in a form that allows it to be easily used.

1) Data Interchange Standards

It is desirable for data to be in a form that uses wellrecognized standards. A collection of standards, often considered as a stack or hierarchy, may be used together although rarely is the delineation clean. For example, standards such as XML or JSON may be used for low-level while International syntactic formatting Standards Organization (ISO) 10303 for product manufacturing information and ISO 22400-2:2014 (KPIs) apply to higherlevel manufacturing semantics [23], [24], [25]. Many standards exist to address other concerns. For example, CMSD (Core Manufacturing Simulation Data) and CSPI (COTS (Commercial Off-The-Shelf) Simulation Package Interoperability) are standards for facilitating the use of simulation models such as shop floor configurations [26], [27]. Standards such as MTConnect and Open Platform Communications Unified Architecture (OPC-UA) may be used to exchange manufacturing information from machine tools while Representational State Transfer (REST) and Simple Object Access Protocol (SOAP) are examples of standards for carrying out communications [28],[29].

Data generators must be able to produce data in a form that either conforms to the relevant standards or is readily adaptable to them.

2) Plug-in interoperability

Data generators should be able to serve as plugins to other systems such as domain-specific testbeds and model design software that may provide other services such as model transformation or optimization. By using standards, plug-in capability can be more easily supported and generators can more likely be incorporated into additional systems. In the reverse sense, generators should also be able to use models based on a plugin architecture. For example, a data generator only capable of generating the results of a milling machine using aluminum would see limited use. Allowing the plugging in of models of tools and materials and parameterizing factors such as feed rates, a generator would be much more useful, leading to more of a widely applicable generator.

V. SUMMARY AND CONCLUDING NOTES

Limited access to real-world data is a significant impediment to advanced manufacturing. Use of virtual data is a promising approach to make up for the lack of access. We have presented issues, challenges, and desirable features in the generation of virtual data. These can be used by researchers to build virtual data generators and gain experience that will provide data to data scientists while avoiding known or potential problems. This, in turn, will lead to better requirements and features in future virtual data generators.

Three areas for future research and development are of particular interest and deserve increased attention and effort.

A. Test Data Repositories

Many people experimenting with data analytics would benefit from repositories of both real data and synthetic data. Such repositories would allow multiple academic researchers or commercial companies to be confident that they are using the same data in creating and testing software to deal with what are intended to be common scenarios. It is also desirable to provide configuration data to reproduce the raw data in the repositories so that it can be reproduced as well as modified.

Data repositories should also have other aspects such as areas for algorithms and models that have been proposed to address data sets in the same repository. Ideally, documentation areas and discussion forums would be helpful as well. For example, observations or questions about particular data or data configurations would enable others to make progress by more easily re-using earlier results.

Repositories have been established that incorporate some of these ideas. For example, Bosch has created a challenge that includes measurement data produced from production lines [30]. One example challenge is to "predict which parts will fail quality control." The Bosch data sets and competitions are hosted on Kaggle, a service for general data science challenges [31]. Another example is the NIST Smart Manufacturing Systems Test Bed which makes available data from a manufacturing facility which resembles a small manufacturing shop [32]. Sets of data can be downloaded or queried. In addition, dynamic data streams can be monitored using MTConnect.

B. Standards

Standards development is already an area of intense interest. However, there are gaps in standards that would facilitate data generation and publication of synthetic data. For example, PFA (Portable Format for Analytics) is a useful specification in which to express data analytics [33]; however, while a PFA-enabled host can generate data usable to other software, PFA lacks the ability to control the seeding of its random number generators which limits its flexibility and repeatability. More work is needed on standards to better support data generation.

DISCLAIMER

No approval or endorsement of any commercial product by the National Institute of Standards and Technology (NIST) is intended or implied. Certain commercial software systems are identified in this paper to facilitate understanding. Such identification does not imply that these software systems are necessarily the best available for the purpose.

AKNOWLEDGMENTS

David Lechevalier's work on this effort was supported by National Institute of Standards and Technology's Foreign Guest Researcher Program.

REFERENCES

- M. Saadoun and V. Sandoval, "Virtual Manufacturing and its Implications." Virtual Reality and Prototyping, 1999.
- [2] A. M. Berg, S. T. Mol, G. Kismihok, and N. Sclater, "The Role of a Reference Synthetic Data Generator within the Field of Learning Analytics," J. Learn. Anal., vol. 3, no. 1, pp. 107–128, 2016.
- [3] J. W. Anderson, K. E. Kennedy, L. B. Ngo, A. Luckow, and A. W. Apon, "Synthetic data generation for the internet of things," presented

at the 2014 IEEE International Conference on Big Data, 2014, pp. 171–176.

- [4] W. Terkaj and M. Urgo, "Virtual factory data model to support performance evaluation of production systems," in Proceedings of OSEMA 2012 workshop, Graz, Austria, 2012, pp. 24–27.
- [5] "Arena Simulation Software." Rockwell Automation, 2017.
- [6] G. Shao, S. Jain, and S.-J. Shin, "Data Analytics Using Simulation for Smart Manufacturing," in Proceedings of the 2014 Winter Simulation Conference, Savannah, GA, 2014.
- [7] M. Albert, "STEP NC The End of G-Codes?," Mod. Mach. Shop, Mar. 2006.
- [8] P. Warndorf, "MTConnect Institute Releases Version 1.3.0 of the MTConnect Standard." 2014.
- [9] S. Jain, D. Lechevalier, J. Woo, and S. Seung-Jun, "Towards a virtual factory prototype," Winter Simulation Conference (WSC), 2015, pp. 2207–2218.
- [10] "ISO 10303-238 (2007) Industrial automation systems and integration - Product data representation and exchange - Part 238: Application protocol: Application interpreted model for computerized numerical controllers." Geneva: International Organization for Standardization, May-2007.
- [11] D. Lechevalier, S.-J. Shin, S. Rachuri, S. Foufou, Y. T. Lee, and A. Bouras, "Simulating a virtual machining model in an agent-based model for advanced analytics," Journal of Intelligent Manufacturing, Sep. 2017.
- [12] Jain, Sanjay, and David Lechevalier. "Standards based generation of a virtual factory model." In Proceedings of the 2016 Winter Simulation Conference, pp. 2762-2773. IEEE Press, 2016.
- [13] D. Libes, S. Shin, and J. Woo, "Considerations and recommendations for data availability for data analytics for manufacturing," in 2015 IEEE International Conference on Big Data (Big Data), pp. 68–75, 2015.
- [14] "JSON: The Fat-Free Alternative to XML." JSON.org.
- [15] "OWL Web Ontology Language Reference." World Wide Web Consortium.
- [16] B. Motik, B. Grau, I. Horrocks, Z. Wu, A. Fokoue, and C. Lutz, "OWL 2 Web Ontology Language Profiles." W3C.
- [17] "Guide for verification and validation in computational solid mechanics." ASME, 2009.
- [18] M. R. Garey and D. S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness. New York: W.H. Freeman, 1979.
- [19] Oberkampf, William L., Sharon M. DeLand, Brian M. Rutherford, Kathleen V. Diegert, and Kenneth F. Alvin. "Error and uncertainty in modeling and simulation." Reliability Engineering & System Safety 75, no. 3: pp. 333-357, 2002
- [20] D. Kibira, K. C. Morris, and S. Kumaraguru, "Methods and Tools for Performance Assurance of Smart Manufacturing Systems," J. Res. Natl. Inst. Stand. Technol., vol. 121, pp. 287–318, 2016.
- [21] S. Jain, E. Lindskog, J. Andersson, and B. Johansson, "A Hierarchical Approach for Evaluating Energy Trade-offs in Supply Chains," Int. J. Prod. Econ., vol. 146, no. 2, pp. 411–422, 2013.
- [22] Watson, P. F., and A. Petrie. "Method agreement analysis: a review of correct methodology." Theriogenology 73, no. 9: pp. 1167-1179, 2010.
- [23] A. Katzenbach, S. Handschuh, and S. Vettermann, "JT Format (ISO 14306) and AP 242 (ISO 10303): The Step to the Next Generation Collaborative Product Creation," NEW PROLAMAT, pp. 41–52, Jan. 2013.
- [24] "ISO 10303-1:1994 Industrial automation systems and integration --Product data representation and exchange -- Part 1: Overview and fundamental principles." ISO, Dec-1994.
- [25] "ISO 22400-2:2014 Automation systems and integration -- Key performance indicators (KPIs) for manufacturing operations management -- Part 2: Definitions and descriptions." ISO, Jan-2014.

- [26] Y.-T. T. Lee, "A Journey in Standard Development: The Core Manufacturing Simulation Data (CMSD) Information Model," J. Res. Natl. Inst. Stand. Technol., vol. 120 (2015).
- [27] "SISO-STD-006-2010 Standard for Commercial-off-the-shelf Simulation Package Interoperability Reference Models." Simulation Interoperability Standards Organization (SISO), Inc., Mar-2010.
- [28] W. Mahnke and S.-H. Leitner, "OPC Unified Architecture The future standard for communication and information modeling in automation," ABB Rev., vol. 3/2009, pp. 56–61, Mar. 2009.
- [29] J. Mueller, "Understanding SOAP and REST Basics And Differences." Smartbear.com, Jan-2013.
- [30] "Bosch Production Line Performance." Bosch, Aug-2016.
- [31] "Kaggle: Your Home for Data Science." Kaggle, Inc.
- [32] T. Hedberg and M. Helu, "Smart Manufacturing Systems (SMS) Test Bed." National Institute of Standards and Technology.
- [33] J. Pivarski, "Portable Format for Analytics: moving models to production." KDnuggets, Jan-2016.57670.

Smart Manufacturing Systems based on Cyber-physical **Manufacturing Services (CPMS)**

Yan Lu* and Feng Ju**

*National Institute of Standards and Technology, Gaithersburg, MD 20899 USA (Tel: 301-975-8228; e-mail: yan.lu@nist.gov) **School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ 85281 USA (Tel: 480-965-1746; e-mail: fengju@asu.edu)

Abstract:

Future manufacturing is becoming "smart" – capable of agilely adapting to a wide variety of changing conditions. This requires production plants, supply chains and logistic systems to be flexible in design and reconfigurable "on the fly" to respond quickly to customer needs, production uncertainties, and market changes. Service-Oriented Architecture (SOA) provides a promising platform to achieve such manufacturing agility. It has proven effective for business process adaptation. When combined with the emerging Internet of Things (IoT) technology and the concept of cyber-physical production systems, it is expected to similarly revolutionize real-time manufacturing systems. This paper proposes a new concept of cyberphysical manufacturing services (CPMS) for service-oriented smart manufacturing systems. In addition, we propose a modeling framework that provides appropriate conceptual models for developing and describing CPMS and enabling their composition. Specifically, the modeling framework separates service provision models from service request models and proposes the use of standardized functional taxonomies and a reference ontology to facilitate the mediation between service requests and service consumptions. A 3D-printing use case serves as an example implementation of an SOA-based smart manufacturing system based on our proposed modeling framework.

Keywords: Service-oriented architecture, cyber-physical manufacturing service, real-time, smart manufacturing system, 3D printing, additive manufacturing

1. INTRODUCTION

Historically, manufacturing systems have been designed along relatively rigid hierarchical architectures, the scopes of which are limited to the arrangement and operation of resources directly related to production. The ISA95 model1 classifies manufacturing functions into five logically separated layers. In a typical manufacturing system, lower level systems (0, 1, and 2) are called Operational Technology (OT)² functions, and include hardware and software that monitor and control physical processes. Higher-level functions (3 and 4) are considered Information Technology (IT) domain functions covering manufacturing operations management (MOM), enterprise resource planning (ERP), and more. Because manufacturing IT and OT applications are managed by different organizations within enterprises, their integration is difficult. As a result, traditional manufacturing systems are slow in responding to market or supply chain changes and are very vulnerable to malfunctions of main subsystems. Furthermore, the rigid architecture of traditional manufacturing systems also leads to the difficulty of reusing or retrofitting existing manufacturing capability. Manufacturers tend to replace old assets with new, incurring huge capital investment costs and expending considerable engineering effort.

In contrast, the emerging trend of smart manufacturing (SM) transforms manufacturing systems into agile and efficient ecosystems. SM integrates the latest information and communication technologies (ICT) into manufacturing systems to enable real-time response to changing demands and conditions in the factory, in the supply network, and in customer needs. In this new paradigm, the Internet of Things (IoT), the digital factory, and cloud computing technology play major roles in transforming the rigid hierarchical architecture into a flexible style (Kulvatunyou, (2016)). Lu et al. (2016b) propose a Service Oriented Architecture (SOA)-based approach to restructure manufacturing systems and make them smarter and more agile. SM systems based on SOA adopt modular design of both hardware and software and combine these elements into cyberphysical systems (CPS) that can then be made available as services for easy access and consistent reuse.

In analogy to the software-based computing services common in the IT world, CPS services are common units for SOAbased SM systems. In our research, we refer to cyber-physical

¹ https://isa-95.com/

² http://www.gartner.com/it-glossary/operational-technologyot/

manufacturing services (CPMS). The scope of CPMS covers both processing units in factories and logistic components in supply chains. The criticality of matching physical location, scheduling availability, and capability of a CPMS to job requirements in real time make the design of SOA-based smart manufacturing systems quite challenging.

This paper extends an SOA approach for integrating business applications to smart manufacturing system design based on the concept of CPMS. The main contribution of the paper is the description of the information models and methods that would enable easy development, usage, and dynamic composition of manufacturing and logistics services. We first present a literature review on the application of SOA to manufacturing systems. Then we introduce a definition of CPMS. Section 4 presents a CPMS modeling framework covering the design and application perspectives of CPMS-based manufacturing systems. In Section 5, we use a case study of an SOA-based real-time shop floor manufacturing system to illustrate the application of the CPMS modeling framework. Section 6 concludes the paper and discusses future work.

2. BACKGROUND

An SOA defines a collection of principles for distributed systems design. It relies on the integration of different system components that provide functionalities as loosely-coupled services over a network. These functionalities are packaged via standardized interfaces independent from the underlying implementation and location. SOA has been widely used to achieve adaptable business processes since 1980s (Krafzig et al. (2004)).

Web service technologies enable the most prevailing implementation of SOA³. XML (Extensible Markup Language)based languages like WSDL (Web Service Description Language), OWL-S (Semantic Markup for Web Services). are used to describe a service's interface, functionalities, and characteristics. The UDDI (Universal Description Discovery and Integration) specification depicts a way to build service repositories and specifies how services can be published by service providers and searched and discovered by service requesters. Composition of existing web services into more complex services can be done using BPEL (Business Process Execution Language). The services can be invoked based on the interface description in WSDL. SOAP (Simple Object Access Protocol) messages are used for such invocations. Besides standardsbased web service technologies, ESB (Enterprise Service Bus) is also considered as a de facto method for SOA⁴.

In the manufacturing domain, the concept of SOA has been practically applied in enterprise level integration. At the shop-floor level, quite a few research projects have applied SOA concepts to improve manufacturing system flexibility and reconfigurability^{5, 6}. For example, agent-based distributed intelligent manufacturing systems and holonic manufacturing systems – were heavily studied in academia and industry before

⁶ https://cordis.europa.eu/esprit/src/22728.htm

the early 2000s (Marik (2003)). However, agent-based or holonic manufacturing systems do not exercise some capabilities of SOA systems. They don't specifically address composability issues and the standards necessary to integrate heterogonous agent systems did not exist. In the last several years, the SOA approach to automation was studied and developed with the aim of providing methods, models, and technologies to reduce automation engineering (Theorin et al. (2012, 2014), Vyatkin et al. (2014)). Other influential projects applying SOA to industrial automation include SIRENA7 and SCORDES8, each focusing on developing a web-service-based communication architecture for field devices integration. The projects resulted in Device Profile for Web Service standards9 and their solution stacks for industrial automation. Separate efforts to apply OPC UA (OLE for Process Control Unified Architecture)¹⁰ to shop floor integration also promote the SOA approach, such as the one demonstrated by SmartFactoryFL¹¹. Overall, these efforts took bottom-up approaches and generated results for specific domains (e.g., automation engineering). There is no generic framework to enable SOA methods to integrate functions of the entire manufacturing ecosystems including OT, IT, and supply chain logistic systems.

Automatic service discovery, identification, and orchestration (choreography) to accomplish tasks are very important capabilities for SOA adoption. In the web service area, tremendous effort has been put toward automatic web service integration based on semantics. Although the concepts of semantic web and semantic web services haven't achieved their vision since those concepts cannot scale to meet the general web demands (Hendler, (2011)), some of the results from earlier research projects are very insightful for a domain-specific implementation. For example, the Web Service Modeling Framework (WSMF) (Fensel et al, (2002)) proposed a modeling framework describing various web-service aspects to support strong decoupling of the system components and strong mediation of service integration. Figure 1 shows the core elements of the modeling framework.



Fig. 1. WSMF core elements for web service integration (Domingue et al. (2005))

As shown in Figure 1, the WSMF consists of four core elements: goal repositories define service requests; web services descriptions define various aspects of a web service; ontolo-

- 8 http://www.socrades.net/
- ⁹ http://docs.oasis-open.org/ws-dd/dpws/wsdd-dpws-1.1-spec.html
- ¹⁰ https://opcfoundation.org/about/opc-technologies/opc-ua/
- ¹¹ http://www.smartfactory-kl.de/

³ https://www.w3.org/TR/ws-arch/

⁴ https://en.wikipedia.org/wiki/Enterprise_service_bus

⁵ https:/ims.org

⁷ http://www.ws4d.org/projects/sirena/

gies provide the terminology domain knowledge; and mediators do semantic matching between service requests and services provisions and facilitate the invocation.

Following the semantic web approach, there have been several ontologies developed to formally model manufacturing systems and to enhance sharing and reuse of manufacturing knowledge. Examples of these ontologies include P2 Ontology (Diep et al. (2007)), PSL ¹², MASON (Lemaignan et al. (2006)). PSL was adopted by ISO as an international standard in ISO 18629.

Despite these efforts, manufacturing services based on CPS haven't yet been studied as far as practical integration of OT, IT, and supply chain logistic systems. Our work aims to provide methods, standards, and tools to develop and integrate CPS-based manufacturing services within the scope of a whole manufacturing ecosystem (Lu et al. (2016a)).

3. CPMS AND SOA-BASED SMART MANUFACTURING

The fusion of ICT technologies changes both the scope of manufacturing functions and the interactions among them. While IoT enables unified communications among manufacturing system components, cloud and mobile computing enable manufacturing functions once implemented at different levels of the hierarchy to now be available without even knowing where they are executed. Smart components and smart systems on the shop floor can run advanced analytics and simulations, and make decisions beyond the lower functions defined in ISA 95. To realize the vision of smart manufacturing where systems respond in real time to changing demands and conditions in the factory, in the supply network, and in customer needs, the classical manufacturing system architectural paradigm based on a hierarchical control model (Figure 2 (a)) must

evolve. A new paradigm based on distributed manufacturing services is starting to be adopted, as shown in Figure 2 (b).

In Figure 2 (b), SOA-based smart manufacturing systems are composed of loosely coupled service nodes. Manufacturing services are classified as two types: Cyber (computing) Service and Cyber-physical Manufacturing Services (CPMS).

Traditional software-based IT domain services are classified as Cyber Service nodes. These include enterprise functions such as ERP, Supply Chain Management, MOM functions, and engineering functions associated with Product Lifecycle Management. In addition, virtual factory software for modeling and simulation, data analytics and visualization tools, and other IoT nodes are also regarded as Cyber Service nodes in smart manufacturing systems. Integration of purely Cyber Service nodes is similar to today's SOA approach.

The other type of smart manufacturing service node is named CPMS and differentiates the nature of smart manufacturing systems from existing SOA systems. The formation of CPMS partly breaks the traditional hierarchical automation pyramid and turns the manufacturing ecosystem into a distributed architecture. Depicted as circles in Figure 2 (b), CPMS nodes can represent services provided by work stations, conveyor belts, robots, cells, shops, plants, factories and entire enterprises. To leverage safety and security requirements, level 0 and 1 components of the ISA model, which include real-time and safety critical functions, are encapsulated inside atomic services and stay untouched in this architecture. The atomic CPMS nodes and newly added IoT devices can be nested and composed upwards as CPMS that will be available for shopfloor-level distributed manufacturing, decisions at the MOM and enterprise levels, and even for global production planning



Fig.2. Concept of Cyber-Physical Manufacturing Services

¹² http://www.mel.nist.gov/psl/psl-ontology/psl_core.html

in a connected world. The real-time requirements of the composite services have to be maintained through the composition process. We define CPMS as follows:

CPMS are manufacturing services that include both cyber components with computing and network capabilities and physical components. In addition, the physical components should be able to either process material, store partially finished products, or transport materials and parts.

With the definition of CPMS, our proposed SOA-based SMS has additional requirements for research and implementation that can't be satisfied by existing frameworks. A modeling framework is proposed to help develop such systems. The details will be illustrated in the next section.

4. MODELING FRAMEWORK FOR CYBER-PHYSICAL MANUFACTURING SERVICES

Implementing the CPMS based on SOA is neither simple, trivial, nor quick. It might take many years to fully realize the strength of the new architecture. Specifically, several aspects are considered for designing successful CPS-based services for SMS to cover the complete spectrum of service representation, discovery, understanding/validation, negotiation, invocation, monitoring, fault handling, and service adaptation. In addition, for dynamic and automatic service composition and execution, one must decide:

- 1) What is the best functional decomposition of CPMS for maximal manufacturing system adaptability?
- 2) How to enable automated orchestration possible to enable dynamic reconfiguration?
- 3) How to assure non-functional requirements are met?

4) How to operate and evolve the services once deployed?

To address these challenges, a modeling framework is necessary for designing, developing, and operating CPMS-based SMS. The CPMS model framework should cover the modeling perspectives for automated composition. Specifically, we need to consider constraints based on the physical component (context-based) and the services it provides. In addition, to satisfy the needs for real-time decision making, the states of physical system-based services should be available for query. A methd defining state models for CPMS is needed. Manufacturing goals and decomposition are critical as well in correlating manufacturing activities with services.

Figure 3 depicts a CPMS-based service-modeling framework that extends the WSMF framework. Semantic technology is adopted to separate service consumers in a specific context from the services to be consumed. In consequence, we model the service consumption side and service provision side separately. Domain-specific and even enterprise-specific models are accepted by the modeling framework. To match the manufacturing service goals with services, we employ a manufacturing reference ontology that can be used by manufacturing service mediators for service registration, discovery, match, identification, optimization, orchestration, and invocation.



Fig. 3. Modelling framework for Cyber-Physical Manufacturing Services

4.1 CPMS Service Modeling

CPMS models should consider explicit capability representation, standard interfaces, grounding mechanisms, and nonfunctional features. The Dublin Core¹³ Metadata Element Set can be used to capture general service property sets. The granularity and type of CPMS can be defined and standardized for individual domains. CPMS capability is the key to service matching and it should be modeled to indicate the pre-conditions, assumptions, post-conditions, and effects of such capabilities. There are special requirements for defining CPMS- interfaces for engineering and runtime invocation. Quality of





services, including reliability (availability), real-time criticality, safety criticality, and security should be specified as well. Figure 4 shows a concept model of CPMS.

Common Cyber-Physical Manufacturing Services

Among CPMS's non-functional properties, Service classification plays an important role for necessary service discovery. The use of taxonomy to describe the type of manufacturing services is inevitable. In general, common CPMS can be divided into three main categories: Processing, Transporting, and Storage. Processing services work on materials or partially

¹³ http://dublincore.org/

finished products by changing their physical or chemical properties. Both transporting and storage services are logistical services to move material around or solely store it. Each of these services can be further classified as shown in Figure 5. For example, production services can be classified into batch production, continuous production, and discrete production. Discrete production includes Shaping services and Nonshaping services, etc. Additive manufacturing shapes material layer on layer. It can be classified further into seven categories based on ASTM F2792.

To provide an unambiguous identification of service classes and their relations, a standardized common service dictionary will be needed.



Fig. 5. An example of CPMS taxonomy

CPMS Capability Description

The taxonomy potentially defined by a common CPMS dictionary provides a hierarchy of concepts to describe service types. For accurate service match and composition, the capability of the service has to be explicitly modelled. The capabilities of a CPMS determine not only the range of material it can process, but also its ability to meet product manufacturing specifications, as well as capacity (throughput, failure rate) and availability requirements. In addition, the possible states of the capability and job queue information should be defined for run-time choreography and dynamic orchestration. For example, a capability may be identified as current, or may be identified for future times, as depicted in Figure 4. For every object property, an ontology can be associated for service matching provided by a meta-service.

CPMS Interface

Information about CPMS interfaces is necessary for fast reengineering and dynamic CPMS orchestration or choreographic interactions. For re-engineering with physical adaptation using plug-n-play, physical interfaces are specified. The SmartFactoryFL project has defined unified mechanical and electrical interfaces for multi-vendor production module integration (Weyer et al. (2015)). For online automated service discovery-composition-execution, standard informational interfaces are necessary. Similar to web service description, an input message and an output message should be defined for service invocation. The exact binding methods should be specified as well.

4.2 Manufacturing Activity/Goal Model

Service requests are made when conducting manufacturing activities. Traditionally, manufacturing activity models define both information flows and physical processes involved in activities. When manufacturing activity models are used to represent service requests, the concrete process representation is not necessary. Instead, only the goals of the activity are modelled. A goal can be described as an objective to transform material or a product from one state to another state. The goal metamodel defined in WSMF is applicable in CPMS-based smart manufacturing systems. A goal class has non-functional properties similar to those of web services, e.g., type, name, etc. Its key object properties are the material states they are associated with, and other goals it has dependency on. Many existing works and standards potentially useful for manufacturing activity modeling exist, including Icam DEFinition for Function Modeling 14, Unified Modeling Language 15, Business Process Model and Notation¹⁶. Manufacturing activity models (Barkmeyer et al. (1998), Feng et al. (2015), Kang et al. (2001)) and common manufacturing goals potentially reduce the modeling effort. In addition, Cossentino et al. (2013) describes a process and tools to model goals in a Belief-Desire-Intention (BDI) framework which can be applied here.

4.3 CPMS Reference Ontology

A CPMS Reference Ontology captures the knowledge base to facilitate CPMS mediation – service query, discovery, goal deposition, goal and service match, service composition, etc. An ontology is a formal explicit specification of a shared conceptualization. Usually, an ontology defines two essential aspects – real-world concepts and formal semantics – necessary for information to be processed by a computer. A CPMS reference ontology defines basic concepts in the goal/activity and service domain and their relationships among the set of concepts. In addition, it provides a set of axioms for reasoning necessary for goal-service matching.

Many ontologies exist in the manufacturing domain that describe product, process, and resources (PPR) (Nielsen (2003)), and these can be reused here. The concepts of product and process link manufacturing goals with services, while the resource definition adds context to manufacturing service models by introducing the concepts of time and location. On the basis of the reference ontology, the mapping knowledge is necessary for describing each pair of two functional taxonomies from the service consumption domain and from the service provision domain. The mapping knowledge should consist of mappings between the concepts in the two functional taxonomies and their structural differences based on the taxonomies.

4.4 CPMS Lifecycle Management

The CPMS modeling framework can be applied to global production network resource planning, supply chain planning,

¹⁴ http://www.idef.com/

¹⁵ http://www.uml.org/

¹⁶ http://www.bpmn.org/

shop-floor manufacturing re-engineering as well as on-the-fly production line reconfiguration. Because production facilities live much longer than the production period for a particular product, CPMS usually are designed to be adaptive. While CPMS type and interface tend to remain constant, service capability may change as equipment degrades and available capacity varies. Therefore, CPMS lifecycle management is very important for SOA-based SM systems. In addition to versioning information, service usage logs, and snapshots of performance history should be kept and reported by a service management system.

5. CASE STUDY – 3D PRINTER AS A MANUFACTURING SERVICE

To illustrate CPMS and its modeling framework, we describe a three-dimensional (3D) printing application scenario. We can imagine a "customized vehicle manufacturing" factory which provides 3D printing services for automotive components. This factory floor has all sorts of additive manufacturing machines, each of which can print parts from a range of feedstock materials and within a certain build volume. Big parts can be divided into several partitions and printed in parallel. Consequently, additional joining services may be required to produce final parts. By applying CPMS in the semantic framework, an order manager can invoke several 3D printing services, automated guided vehicle services, robot services, and welding services and aggregate them into new customer services in a semi-automatic way. In this scenario, we focus only on the description of a 3D printing service and illustrate it based on a concrete implementation.

Our implementation is specifically devoted to exploring the capabilities an OPC UA-compliant data modeling and industrial automation software, Status Enterprise, in the context of implementing a distributed network of manufacturing machines to accomplish a common service-oriented goal. To evaluate and illustrate the CPMS modeling framework, we developed a mock-up use case – a 3D printer as a manufacturing service. Using such a service-oriented implementation, the 3D printing service can be accessed over the network with explicit capability representation, standard interfaces, grounding mechanisms, and non-functional features.

5.1 System setup

The architecture is based on B-Scada Status Enterprise, which is a visualization/automation software designed for industry. It provides an OPC UA-compliant interface and can characterize asset information and communication among assets using data models. As shown in Figure 6, the data model exposes all the non-functional features for the 3D printer, including the basic equipment description (model, ID, manufacturer) and service statuses (e.g., Time Remaining, QueueLength, tool positions). Specifically, a queue is constructed to handle concurrent requests from multiple clients/customers. If a job is sent to a busy 3D printer, it will be stacked into the queue and the QueueLength will be increased by one accordingly. TimeRemaining property keeps track of the current job's progress and all the remaining jobs in the queue to provide a realtime estimate of the total remaining time to a new job requester.

5.2 Process Execution

To implement these functionalities for the service invocation, a C# program is written as an SOA interface to the OPC UA server designed to mimic the functionality of a 3D printer and used as a platform to communicate with other SOA machines. Three major threads in the program are taking new requests, executing existing requests, and updating the remaining time. The detailed process execution is laid out as follows.

Properties - Printer001 Objects/User View/My simulation/Printer001					
Color Background on Change 5 (seconds) Animation Duration					
Property Name	Value	Quality	Time Stamp		
Asset Operation Status	None	Good	12/9/2015 1:19:25 PM		
Command		Good	7/27/2016 10:47:34 AM		
Criticality	None	Good	12/9/2015 1:19:25 PM		
Criticality Index	None	Good	12/9/2015 1:19:25 PM		
Equipment Description		Good	7/7/2016 9:57:13 AM		
Equipment Level	None	Good	12/9/2015 1:19:25 PM		
Fixed Asset ID		Good	7/7/2016 9:57:13 AM		
FuncCallMade	False	Good	7/27/2016 10:47:34 AM		
ID	184	Good	7/7/2016 9:57:13 AM		
Manufacturer		Good	7/7/2016 9:57:13 AM		
Model		Good	7/7/2016 9:57:13 AM		
Output		Good	7/27/2016 10:47:34 AM		
QueueLength	2	Good	7/27/2016 10:47:37 AM		
Record Locked	False	Good	12/9/2015 1:19:25 PM		
Serial Number		Good	5/27/2016 2:47:00 PM		
Time Remaining	12:44:59	Good	7/27/2016 10:47:37 AM		
XLim	200	Good	7/1/2016 11:23:14 AM		
XPos	93.862	Good	7/27/2016 10:42:27 AM		
YLim	200	Good	7/1/2016 11:23:11 AM		
YPos	75.859	Good	7/27/2016 10:42:27 AM		
ZLim	200	Good	7/1/2016 11:23:09 AM		
ZPos	0.5	Good	7/27/2016 10:42:27 AM		

Fig. 6. Data model of the OPC UA server for 3D printer

First, the connection between the server and the 3D printer is established. Through an OPC UA data connector, the specific values from the printer are connected to specific fields on the server when setting up the data connector on the server. Then a thread for taking requests is started. This thread polls the server for a service request, queues it for processing, and resets the input fields. It runs in the background for the entire duration of the program. As each request is queued, the time estimation is determined and added to a separate queue, as well as to the current time estimation. After that, a separate thread is created to write the time remaining in the queue to the server. This thread automatically updates the total remaining time for all the jobs in the queue, including the one being processed. Each command is dequeued and executed in sequence, updating the relevant server values (queue length, output field). After each job is completed and dequeued, the time estimate is recalculated based on a queue of the time estimations.

5.3 Discussion

In moving such a CPMS modeling framework into the real manufacturing world, two major issues arise: one is the challenge of handling multiple concurrent client requests, and the other one is automatic device registration for multiple machines in a scaled-up network. In this mock-up use case, the developed platform proved capable of handling concurrent requests from multiple clients. In addition, each client can access information from the server without interference from other server-client connections. These consequently warrant further investigation into the synthesis of an SOA system based on Status Enterprise.

As for the automatic device and service registration in a machine network, we adopt the concept of Devices Profile for Web Services (DPWS), which can enable secure web service messaging, discovery, description, and eventing on resourceconstrained endpoints. Under the DPWS framework, a client can automatically detect other DPWS-enabled devices on a network, and discover and invoke the service functionality that each device or several devices can offer, according to their corresponding service types and service capabilities. Based on the CPMS taxonomy and ontology, manufacturing goals are decomposed into (or composed of) services that will be fulfilled by the DPWS devices automatically discovered in the network. Once an event is in progress, the devices are able to communicate through messaging and eventing functionalities to possibly reconfigure the system or adjust the process. Meanwhile, the service requester or customer can access the job information and get real-time feedback on the manufacturing progress.

6. CONCLUSIONS

With the introduction of smart devices embedded with more intelligence and predictive analytics, production units and logistic components are more accessible as services on a network. This paper described a service-oriented smart manufacturing architecture based on cyber-physical manufacturing services. Specifically, the paper proposed a semantic modeling framework that would enable easy development, usage, and dynamic composition of CPMS services. Under the framework, we identified capability gaps requiring further research effort in order to facilitate the adoption of smart manufacturing technology and the proposed architecture, including creation of standards for communication protocols, data model, knowledge representation, and CPMS characterization. Our future work will focus on cyber-manufacturing service models. Uses cases of applying CPMS to smart manufacturing systems will be studied and classified. Requirements on CPMS service models will be derived from the use cases. Existing manufacturing information modeling standards and ontologies will be used as bases to model CMPS capabilities, performance, interfaces and interaction methods.

DISCLAIMER

Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST.

REFERENCES

- Barkmeyer, E. J., (1998), SIMA Reference Architecture Part 1: Activity Models, NISTIR 5939.
- Cossentino, M., Dalle Nogare, D., Giancarlo, R., Lodato, C., Lopes, S., Ribino, P., Sabatucci, L., and Seidita, V., (2014). GIMT: A tool for ontology and goal modeling in BDI Multi-Agent Design, Workshop "Dagli Oggetti agli Agenti. Deen, S.M. (Ed.), Agent-Based Manufacturing -Advances in the Holonic Approach, Springer-Verlag, Heidelberg, Germany, 2003.
- Diep, D., Alexakos, C. and Wagner, T., (2007). An ontologybased interoperability framework for distributed manufacturing control, IEEE Conference on Emerging Technologies and Factory Automation, Patras, Greece.

- Domingue, J.; Roman, D. and Stollberg, M., (2005). Web Service Modeling Ontology (WSMO): an ontology for Semantic Web Services. W3C Workshop on Frameworks for Semantics in Web Services, 9-10 June 2005, Innsbruck, Austria
- Feng S., Witherell, P. W., Ameta, G., Kim D. B., (2015), Fundamental Requirements for Data Representations in Laser-based Powder Bed Fusion, Proceedings of the 2015 Manufacturing Science and Engineering Conference.
- Fensel, D. and Bussler, C., "The Web Service Modeling Framework WSMF", Electronic CommerceResearch and Applications, Vol. 1, Issue 2, summer 2002, Pages 113-137.
- Hendler, J., (2011). Why the Semantic Web will never work, 8th Extended Semantic Web Conference (ESWC), Heraklion. Greek.
- Kulvatunyou, B., Ivezic, N. and Srinivasan, V., "On Architecting and Composing Engineering Information Services to Enable Smart Manufacturing", J. Comput. Inf. Sci. Eng. 2016: 16(3).
- Krafzig, D., Bank, K. and Slama, D., "Enterprise SOA: Service-Oriented Architecture Best Practices", Prentice Hall, 2004
- Lemaignan, S., Siadat, A., Jean-Yves Dantan, J., and Semenenko, A., (2006). MASON: A Proposal For An Ontology Of Manufacturing Domain, IEEE Workshop on Distributed Intelligent Systems: Collective Intelligence and Its Applications, Prague, Czech Republic.
- Lu, Y., Morris, K. C., Frechette, S. P., (2016a) " Current Standards Landscape for Smart Manufacturing Systems ", **NISTIR 8107.**
- Lu, Y., Reddick, F., and Ivezic, N., (2016b). The Paradigm Shift in Smart Manufacturing System Architecture, Advances in Production Management Systems, Iguassu Falls, Brazil
- Marik, V., McFarlane, D., and Valckenaers, P., (Eds.), Holonic and Multi-Agent Systems for Manufacturing, Springer-Verlag, Heidelberg, Germany, 2003.
- Nielsen, J., (2003). Information Modeling of Manufacturing Processes: Information Requirements for Process Planning in a Concurrent Engineering Environment, Doctoral Dissertation, Royal Institute of Technology.
- Theorin, A., Ollinger, L., and Johnsson, L., (2012). Serviceoriented Process Control with Grafchart and the Devices Profile for Web Services, 4th IFAC Symposium on Information Control Problems in Manufacturing, Bucharest, Romania.
- Theorin, A., Ollinger, L., and Johnsson, L., (2014). Service Orchestration with OPC UA in a Graphical Control Language, 9th IEEE International Conference on Emerging technologies and Factory Automation, Barcelona, Spain
- Vyatkin, V., Dai, W., Dubinin, and V., Christensen, J. H., (2014). IEC 61499 Implementation of Service Oriented Architecture: Case Study, IEEE Conference on Robotics and Automation, Hong Kong.
- Weyer, S., Schmitt, M., Ohmer, M., and Gorecky, D., (2015). Towards Industry 4.0 - Standardization As the Crucial Challenge for Highly Modular, Multi-Vendor Production Systems, IFAC Symposium on Information Control in Manufacturing, Ottawa, Canada.

Kinetics Based Models to Describe Environmental Degradation of Polymeric Sealants

Donald Hunston and Christopher White National Institute of Standards and Technology Engineering Laboratory Gaithersburg, MD 20899-8615, USA donald.hunston@nist.gov

Introduction

Sealants represent an essential component of modern construction where they prevent unwanted moisture intrusion and subsequent water damage. Although current materials are good, studies have shown that a large percentage will fail within 20 years after installation [1-3]. Consequently, better techniques are needed to estimate/predict lifetimes to ensure timely repair and facilitate the development of improved materials. In this connection, the National Institute of Standards and Technology (NIST) established a government-industry consortium program involving most of the major sealant companies. The goal is to develop the ability to understand and model the degradation behavior of sealant materials. As part of this effort, a wide variety of different sealant chemistries and formulations were exposed to weathering both outdoors and indoors. Any changes in the sealant's properties induced by the exposure were monitored by measuring the stress relaxation modulus using a procedure developed in this program and adopted as ASTM C1735 [4,5].

To date, 18 different sealants have been examined in the program and significant degradation in modulus was observed for 9 of the materials. The most surprising result is that for 8 of the 9 sealants, the exposure produces a downward shift in the stress relaxation curve with relatively little change in shape. In simple elastomers, the rubbery modulus is proportional to the cross-link density in the network structure [6]. Sealants, however, have a complex network structure involving chemical cross-links, entanglements, attachments to various fillers, etc. Consequently, the modulus depends on what might be called an effective density of cross-links and junction points. The exposure induced decrease in modulus, therefore, would be related to a reduction in the effective cross-link density. Eventually, the exposure produces cracks in the sealant which progress to failure but the focus here is on the early phase of the degradation process where damage is at the molecular level.

Theory

As a first simple approach to modelling the behavior, the modulus, E, is assumed to be proportional to the average density of "effective" cross-links and junction

points, α

$$E = C \alpha \tag{1}$$

where *C* is a proportionality constant. Four different kinetic models were formulated to describe the rate of change of α during exposure. For the particular set of data that will be examined here, however, the best fit was obtained using a first order equation with the added stipulation that not all effective cross-link and junction points were susceptible to degradation. This situation could result from the variety of different cross-link types that are present or, when the degradation is driven by UV radiation, a limited penetration depth. To develop this model, the density of cross-link and junction points is first divided into that part which can be degraded, α , and that part which does not degrade, α_{∞} . As a result, eq. (1) is modified by replacing α with ($\alpha + \alpha_{\infty}$). The first order rate equation is

$$d\alpha / dt = -k\alpha \tag{2}$$

where k is the rate constant which may depend on temperature, humidity, strain history, and UV exposure. Integrating eq. (2), substituting into the modified version of eq. (1), and rearranging gives

$$E = (E_{\alpha} - E_{\alpha})e^{-kt} + E_{\alpha}$$
(3)

where E_o is the initial modulus and E_{∞} is the limiting value of modulus when all degradation is complete. To facilitate comparisons, the expression can be normalized with the initial value of the modulus: $R=E/E_o$ (similarly for R_o and R_{∞} , where $R_o = 1$)

$$\boldsymbol{R} = (\boldsymbol{R}_{a} - \boldsymbol{R}_{\infty})\boldsymbol{e}^{-kt} + \boldsymbol{R}_{\infty}$$
(4)

The analysis above assumes that the exposure conditions are constant, but in an actual application, the conditions are constantly changing. One approach to dealing with this is to divide the exposure into many small steps where the conditions are constant within each step but change between steps. It is then assumed that, if the steps are chosen correctly, the degradation produced in the step process will be approximately the same for that generated by the continuously changing environment. One factor which may help with this assumption is that the exposure conditions in outdoor aging change relatively slowly.

As a first attempt to test this idea, consider an experiment in the NIST SPHERE where the sample is first exposed for a given period of time using one set of environmental conditions and then exposed for an additional time period at a different set of conditions. For the model proposed above, the modulus ratio during the first exposure period should follow eq. (4) with parameters corresponding to the initial conditions: $R_{1\infty}$, and k_1 . If the degradation is assumed to be cumulative and eq. (4) is still followed, some algebra shows that during the second exposure period, R will follow this relationship

$$\mathbf{R} = (\mathbf{R}_1 - \mathbf{R}_{2\infty})e^{-k_2(t-t_1)} + \mathbf{R}_{2\infty}$$
(5)

where R_1 and t_1 are the values of R and t at the end of exposure period 1 while $R_{2\infty}$, and k_2 are the best fit parameters for the second exposure conditions. Moreover, this iterative procedure can be generalized to predict behavior when more than two step exposures are involved.

Experimental[†]

Of the 8 materials that exhibited degradation in the tests to date, the sealant with the most data is based on Kraton-D, a commercial styrene-butadiene-styrene triblock polymer. This material was selected for inclusion in this program because it displays the properties of a sealant, but contains double bonds that are sensitive to UV radiation. As a result, it shows environmental degradation in a relatively short time period which facilitates generating data more quickly. Consequently, results for these tests will be utilized to examine the model proposed above. Specimens were exposed to a controlled environment in the NIST SPHERE (Simulated Photodegradation via High Energy Radiant Exposure). This device is a specially constructed exposure chamber that allows complete control of environment: temperature, humidity, UV radiation, and mechanical strains. Details of the device are given elsewhere [7,8]. For the tests here the relative humidity was held constant at 50 %, and the UV irradiance was 50 W/m². During exposure, the sample is subjected to deformation via a triangle wave going from 0 % strain to a fixed maximum strain and then back to 0 % strain with a period of 28 minutes. The strain cycle is based on the fact that in applications the sealant is generally subjected to strains resulting from expansion and contraction of the structure resulting from temperature variations between day and night. The 28-minute period should be slow enough to minimize any frequency effects. The uncertainties in the experiments were ± 0.5 °C for temperature, ± 2 % for relative humidity, ± 1 % for the strain, and ± 5 W/m² for UV irradiance.

Specimens were characterized before, during, and after exposure by generating stress relaxation curves using ASTM C1735 with a pretreatment strain of 25 % followed by a test strain of 15 %. For the characterizations during exposure, the test was stopped, the sample cooled to room temperature, the characterization performed, and then the exposure test was continued. For each characterization, the results from four specimens were averaged and error bars representing 1 standard deviation were calculated. Figure 1 shows some typical curves for fresh and exposed specimens. The curves are parallel, and the plateau moduli for the analysis were taken by arbitrarily selecting the values at 100 s.



Figure 1: Stress relaxation curves for fresh samples (black) and specimens aged at 51 °C with 21 % maximum strain, for 2 d (pink), 4 d (blue), 8 d (green/black), 20 d (red), and 40 d (gray/blue). Points are average of 4 tests with error bars for 1 standard deviation.

Results and Conclusions

For the initial analysis of the exposure tests, sufficient data were available at 5 conditions: maximum strain of 21 % at temperatures of 31 °C, 41 °C, and 51 °C, and maximum strain of 11 % at 31 °C and 41 °C. Figure 2 shows a plot of the normalized modulus ratio vs exposure time as well as the predictions based on best fits with eq. (4). Again, each point represents the average of four tests with error bars of 1 standard deviation. It is clear that the model captures the trends in the data very well. Tests are now underway using 2 step exposure experiments where two different environments are combined in sequence. Initial results look promising.

[†] Certain commercial materials and equipment are identified in this paper in order to specify adequately the experimental procedure. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology nor does it imply necessarily that the items are the best available for the purpose.



Figure 2: Values for modulus ratio at 100 s for specimens exposed at indicated conditions as a function of exposure time. Points are an average of 4 specimens with error bars of 1 standard deviation. Solid curves are best fit results using eq. (4).

References

- R. Woolman, <u>Resealing of Buildings: A Guide to</u> <u>Good Practice</u>, 1995, Butterworth-Heinemann, Oxford.
- E. Grunau, <u>Service Life of Sealant in Building Con-</u> struction, 1976, Federal Minister for Regional Planning, Building and Urban Planning, Bonn, Germany.
- R. Chiba, H. Wakimoto, F. Kadono, H. Koji, M. Karimori, E. Hirano, T. Amaya, D. Sasatani, and K. Hosokawa, *Japan Sealant Industry Association*, 1992, pp.175-199.
- ASTM standard C1735, Standard Test Method for Measuring the Time Dependent Modulus of Sealants Using Stress Relaxation, 2011, ASTM International, West Conshohocken, PA.
- J. D. Ferry, <u>Viscoelastic Properties of Polymers</u>, 1980, John Wiley & Sons, New York, New York.
- 6. C. C. White, and D. L. Hunston, *Polymer Engineering* and Science, 2008, <u>48</u> (12), pp. 2317-2328.
- J.W. Chin, E. Byrd, N. Embree, J. Garver, B. Dickens, T. Finn and J. Martin, *Rev. Sci. Instrum.*, 2004, <u>75</u>, pp. 4951-4959.
- C. C. White, D. L. Hunston, K. T. Tan, J. Hettenhouser and J. D. Garver, *Rev. Sci. Instrum.* 2013, <u>84</u>, pp. 095113-1-095113-10.

INTERNATIONAL CONFERENCE ON ADVANCES IN CONSTRUCTION MATERIALS AND SYSTEMS, ICACMS 2017

BOND ENHANCEMENT OF REPAIR MORTAR VIA BIODEPOSITION

Didier Snoeck ⁽¹⁾, Jianyun Wang ⁽¹⁾, Dale P. Bentz ⁽²⁾ and Nele De Belie ⁽¹⁾

(1) Magnel Laboratory for Concrete Research, Faculty of Engineering and Architecture,

Ghent University, Technologiepark Zwijnaarde 904, B-9052 Ghent, Belgium

(2) Engineering Laboratory, National Institute of Standards and Technology,

100 Bureau Drive, Stop 8615, Gaithersburg, MD 20899, USA

Abstract ID Number (given by the organizers): 12

Keywords: Calcium carbonate, bond strength, slant shear, biodeposition.

Author contacts

Authors	E-Mail	Fax	Postal address
Didier Snoeck	didier.snoeck@ugent.be	/	Technologiepark Zwijnaarde 904, B- 9052 Ghent BELGIUM
Jianyun Wang	jianyun.wang@ugent.be	/	Technologiepark Zwijnaarde 904, B- 9052 Ghent BELGIUM
Dale P. Bentz	dale.bentz@nist.gov	/	100 Bureau Drive, Stop 8615, Gaithersburg, MD 20899 USA
Nele De Belie	nele.debelie@ugent.be	/	Technologiepark Zwijnaarde 904, B- 9052 Ghent BELGIUM

Corresponding author for the paper: Didier Snoeck

Presenter of the paper during the Conference: Nele De Belie

Total number of pages of the paper (the first excluded): 8

Instructions to authors submitting a final paper

Authors should submit a final paper corresponding to the accepted abstract <u>no later than 15 February 2017</u>. The organisers do not commit themselves to include in the Proceedings any paper received later than the abovementioned deadline. At least one of the authors must register and pay his/her registration fee during the advance period (before April 30th 2017) for their paper to be included in the final programme of the Conference. In case an author has more than one paper, the registration would be valid only for one paper and co-author/s should be registered for publication of multiple papers.

This document provides information and instructions for preparing a paper to be included in the Proceedings of the ICACMS 2017 Conference. Papers and all related correspondences should be written in English.

In the manuscript, the paper article should be preceded by one page containing the author(s) information (as in this page). The page should contain the paper title, abstract ID number (given by the organisers), keywords, e-mail and postal address of each author. The first name will indicate the principal/corresponding author and those that follow, the co-authors in the order of precedence.

BOND ENHANCEMENT OF REPAIR MORTAR VIA BIODEPOSITION

Didier Snoeck⁽¹⁾, Jianyun Wang⁽¹⁾, Dale P. Bentz⁽²⁾ and Nele De Belie⁽¹⁾

(1) Magnel Laboratory for Concrete Research, Faculty of Engineering and Architecture,

Ghent University, Technologiepark Zwijnaarde 904, B-9052 Ghent, Belgium

(2) Engineering Laboratory, National Institute of Standards and Technology,

100 Bureau Drive, Stop 8615, Gaithersburg, MD 20899, USA

Abstract

The bond between repair mortars and existing concrete substrates is critical for the longterm performance and durability of the repaired structure. The carbonation state of the substrate is one of the parameters that may affect this bond strength. The type of calcium carbonate polymorph (calcite, aragonite or vaterite) affects the nucleation and growth of cement hydration products. In this study, carbonation of the mortar substrate is promoted via the biodeposition of calcium carbonate by a ureolytic bacteria strain previously employed in bioconsolidation. X-ray diffraction and scanning electron microscopy were used to examine the interfaces between the repair material and the substrate, as well as the polymorph of the deposited calcium carbonate. It appeared that the approximately 50 μ m thick biodeposition film on the mortar surface mostly consisted of calcite and vaterite. The deposited crystals were full of bacterial imprints. Both the repair material and the substrate tended to show a good adherence to that layer. The bond, as assessed in this study by slant shear specimen testing, was improved by the presence of the biodeposition layer.

1. INTRODUCTION

Concrete is one of the most widely used construction materials on Earth. It is an ideal material to resist compressive forces, but when sufficient tensile forces are present, the restrained concrete may crack. And, without repair of the cracks, the durability can be critically compromised. One can decide to use a self-healing concrete during the design phase of construction [1-3], but repair of existing concrete structures will still often be needed. This manual repair should be made with care and precautions should be taken to assure that the repair is long-lasting, durable and efficient. If the bond between repair product and concrete substrate is not sufficient, delamination or spalling may occur. Therefore, one needs to make sure that the surface treatment of the substrate is properly executed.

One way of increasing and controlling the bond between the repair material and the concrete substrate could be the use of a biodeposition layer. One of the first patented applications on biodeposition was the protection of ornamental stone by a microbially deposited carbonate layer [4, 5]. This calcium carbonate precipitation can be biologically induced by bacteria. These bacteria synthesize the minerals in a unique form, characteristic of the specific species of bacteria. This biodeposition technique has also been applied on cementitious materials resulting in an increased resistance of mortar specimens towards chloride penetration, freeze/thawing and carbonation [4, 6-8]. It should merely be considered as a coating system as carbonate precipitation is mainly a surface-controlled phenomenon due

Page 1

to the limited penetration of bacteria into the porous cementitious matrix. Thin-section analysis revealed that the thickness of the bacterial layer was typically within the range of 10 μ m to 40 μ m; in which larger crystals up to 110 μ m could be found [4]. This layer can be a promising route to engineer the substrate surface for optimal bond strength characteristics.

In this paper, the bond strength was assessed by slant shear testing. Specimens with and without biodeposition layer were studied and the crystal formation composition and morphology were examined.

2. MATERIALS AND METHODS

2.1 Bacterial strain and cultivation condition

Bacillus sphaericus LMG 22257 (Belgian coordinated collection of microorganisms, Ghent) was used in this study. The bacteria were grown aseptically in the growth medium (400 mL per batch) which consisted of a blend of yeast extract (20 g/L) and urea (20 g/L). The culture was incubated at 28 °C on a shaker at 100 rpm [10.5 rad/s] for 24 h. Subsequently, the bacterial cells were harvested by centrifugation (7000 rpm [733.0 rad/s], 7 min) of the 24 h old grown culture and were resuspended in sterile saline solution (NaCl, 8.5 g/L). The concentration of the bacteria in the suspension was $1.5 \cdot 10^9$ cells/mL to $2 \cdot 10^9$ cells/mL. The obtained bacterial suspension was stored in a 4 °C refrigerator for further experimental use.

2.2 Mortar specimens

The standard used to make the mortar substrates was ASTM C882/C882M-13 on 'Bond strength of epoxy-resin systems used with concrete by slant shear'. Six Portland-cement mortar cylinders with a standard mixture composition according to EN 196-1 were cast (510 kg/m³ CEM I 52.5 N, 1530 kg/m³ silica sand 0/2, and 255 kg/m³ water). The specimens' diameter and height were 75 mm and 150 mm, respectively, and each had a diagonally cast bonding area at a 30° angle from the vertical, as per the ASTM standard.

The specimens were manually ground (bonded surface) by means of a sand paper until the desired roughness was reached. The International Concrete Repair Institute (ICRI) has a set of "roughness" surface profile chips (http://www.jdtechnical.com/Surface_Prep-JD.htm). An intermediate profile, similar to the CSP-5 chip, was targeted at an age of 28 d. All prepared surfaces were visually similar.

2.3 Biodeposition treatment

Mortar specimens (BAC) were partially immersed in a precipitation medium which consisted of urea (30 g/L), calcium nitrate (118 g/L) and yeast extract (5 g/L) for 24 h. The medium level was approximately 10 mm above the immersed surface (elliptical surface for applying repair material) of the mortar specimens. After that, the specimens were taken out from the precipitation medium and put upside down until surface dry. Subsequently, bacterial suspension was sprayed (approximately 0.5 mL/cm^2) all over each elliptical surface every 6 h for 4 times.

2.4 Repair material application and slant shear testing at 28 d

All mortar substrate specimens were soaked in tap water for 24 h. The face of the mortar section was put on an absorbent material for 10 min and was subsequently dried in air for 15 min. The repair material (Sika MonoTop-124 N¹) was mixed for 3 min. The prepared bonding surface was put horizontal and 2 mm of repair material was applied, followed by the application of a second mortar specimen. The entire cylinder (especially the sides) was(/were) covered with plastic foil and tape to exclude movement during hardening of the repair material. The specimens were kept horizontal for 48 h. Afterwards they were put in a moist room at 95 ± 5% RH and 20 ± 2°C until the repair product achieved an age of 28 d. Prior to testing, the loading surfaces of each cylinder were ground to produce a smooth testing surface. The composite specimen was loaded in compression and its strength was recorded. The bond strength was determined by dividing the load carried by the specimen at failure by the area of the bonded surface. The area of the bonded surface was reduced by that of any voids found in the bond layer on inspection after the test. Only voids larger than 3 mm in diameter were taken into account.

2.5 Characterization of the interface between mortar substrate and repair material

The mineral phases in the mortar-biodeposition layer-repair material interface were investigated by use of X-ray diffraction (XRD), Scanning Electron Microscopy (SEM) and thin section analysis. After the slant shear test, shards from the mortar surface, repair material surface and biodeposition layer were manually collected using a spatula.

A copper X-ray tube was used for the XRD analysis. The samples were manually trimmed into about 1 cm in diameter and 2 mm to 3 mm thick pieces to fit the sample holder. Scanning was performed from 10° to 70° two-theta with a step size of 0.039° .

Samples for SEM analysis were first subjected to a gold coating process. SEM analysis was performed on an instrument equipped with an EDX detector operating at an acceleration voltage of 20 kV.

To study the formed crystals near the interface between the mortar substrate and the repair material, thin sections (40 mm \times 25 mm \times 25 μ m) were prepared from the tested slant shear specimens, perpendicular to the interface and along the height of the cylindrical specimen (Figure 1). First, the specimens were cut to expose 40 mm \times 25 mm faces, which were then glued on a glass slide with a thickness of 2.9 mm. The combined sample was cut and polished until a height of the specimen and glass of 10.1 mm was reached. Next, the specimens were impregnated under vacuum with a fluorescent epoxy. The excess epoxy was polished away and an object glass was glued on the smooth surface. Finally, the glass slides were cut off and the remaining part was polished until a thin section with 25 μ m thickness was achieved. A cover glass was glued on top to protect the thin section. The thin sections were then analysed under normal and fluorescent light.

¹ Certain commercial products are identified in this paper to specify the materials used and the procedures employed. In no case does such identification imply endorsement or recommendation by Ghent University or the National Institute of Standards and Technology, nor does it indicate that the products are necessarily the best available for the purpose.



Figure 1: Slant shear setup (left) and thin section location (right)

3. **RESULTS**

3.1 Slant shear test

The results obtained when performing a compression test on the slant shear specimens, are shown in Table 1. The elliptical surface area was measured after the test. Possible defects at the edges and larger air voids were subtracted from the complete area. No distinct differences in elliptical surfaces were found on the studied samples series and the results thus could be compared. The bond strength results were characterized by an average and standard deviation of 12.6 MPa \pm 1.6 MPa and 14.3 MPa \pm 1.0 MPa for the REF and BAC specimens, respectively. The average results for the bacterially treated specimens are 13% higher compared to the un-treated specimens. The bacterial treatment thus seems to improve the overall bonding strength, but this preliminary result is only significantly different at a significance level of 0.18 (t-test). The BAC III specimen showed a lower force when performing the slant shear test for unknown reason. Additional future tests will elucidate this assumption and an improved treatment will be proposed.

	Force [kN]	Elliptical surface area [cm ²]	Bond strength [MPa]
REF I	99.15	9120.6	10.9
REF II	117.04	9131.6	12.8
REF III	126.78	9040.3	14.0
BAC I	135.7	9148.5	14.8
BAC II	136.21	9143.8	14.9
BAC III	119.99	9079.6	13.2

	Table 1: Force.	elliptical	surface area	and bond	strength	results
--	-----------------	------------	--------------	----------	----------	---------

Bentz, Dale; De Belie, Nele; Snoeck, Didier; Wang, Jianyun. "Bond Enhancement of Repair Mortar via Biodeposition." Paper presented at INTERNATIONAL CONFERENCE ON ADVANCES IN CONSTRUCTION MATERIALS AND SYSTEMS, Chennai, India. September 3, 2017 - September 8, 2017. Figure 2 shows the elliptical surface after performing the slant shear test. The surface of the reference samples is smooth and debonding of the repair material layer from the mortar surface occurred. Conversely, a clear whitish layer was observed on the bacterially treated specimens. This is from the calcium carbonate biodeposition due to the bacterial treatment. Partial debonding of the calcite layer deposited by the bacteria was seen by comparing the bonding surfaces after compression (Figure 2). This rougher surface is thus possibly primarily responsible for the increase in slant shear bond strength.



Figure 2: Bonding surfaces after performing the slant shear test

3.2 XRD analysis of the carbonates

The XRD spectra are shown in Figure 3 and the EDX spectra in Figure 4. The biodeposition film on the mortar surface mostly consisted of calcite and vaterite. The amount of both was quite significant, which is indicated by the sharp and strong dominant diffraction peaks in the spectrum (green lines). The compositions of the mortar substrate and repair material surfaces were quite similar, as they are both cementitious materials. They both contained calcite, vaterite and quartz. The detailed percentage of each mineral is unknown from this qualitative analysis. Yet it can still be seen that calcite was the main mineral, while the amount of vaterite was much less than that of calcite on the mortar substrate and repair material surfaces. This can be judged by the fact that the dominant diffraction peaks of vaterite were very weak in the spectra of the samples from mortar substrate and repair material surfaces (blue and red lines). No calcium hydroxide (CH) or calcium silicate hydrate gel (C-S-H) was found on the surfaces, suggesting that they were highly carbonated. Nor was aragonite indicated in any of the XRD spectra.

Vaterite is the major product in the pH-range between 8.5 and 10. Conversely, aragonite preferably forms at pH 11, while calcite is the dominant product when the pH is higher than 12 [9]. Vaterite is a metastable polymorph of calcium carbonate and is rare in natural environments [10]. It is unstable and transforms to calcite (or aragonite) at room temperature in an aqueous solution [11]. However, vaterite can be synthesized in chemical processes and often forms in the presence of microorganisms [12]. This gives an indication that bacteria and their secretion (mainly organic compounds) may facilitate the formation of vaterite.

The three kinds of surfaces had completely different morphologies. No obvious crystals were observed on the mortar surface. While on the repair material surface, particles of a size ranging from 2 μ m to 5 μ m were seen. Based on the EDX spectrum, these particles might be Ca-Mg-carbonates.



Figure 3: XRD spectra from mortar substrate, repair material and bio-deposition surfaces



Figure 4: EDX spectra from the mortar substrate (left) and the repair material (right) surfaces

The biodeposition film was full of bacterial imprints. These imprints could be seen as long elliptical shaped spots with a length of approximately 2 μ m. The EDX spectra indicated that the film was mainly composed of calcium carbonate. The film was not flat; instead, it was rough with a lot of pits. This could be due to roughening of the mortar surface with sand paper.
3.3 Thin-section analysis

Thin-section analysis is useful in viewing the interlayers between the mortar and the repair product in case of the reference samples and the interlayers between the mortar, calcium carbonate (biodeposition) layer and the repair material, respectively, for the bacterially treated specimens. In the case of the reference sample, the crack propagated through the interface of the mortar substrate and the repair material. This seems to be the weakest link when applying the repair material.

In the case of the bacterial treated specimens, it could be seen that debonding takes place in the repair material near the vicinity of the $CaCO_3$ layer or in the interface between the layer and the substrate. Both the bonding with the mortar substrate and the repair material seems to be sufficient to increase the bond strength.

The complete system in the bacterially treated specimens is shown in Figure 5. A clear whitish layer can be seen in between the mortar substrate and the repair material.



Figure 5: Formed CaCO₃ layer attributed to bacterial activity near the interface The figure width corresponds to $800 \,\mu m$

The whitish CaCO₃ layer in the bacterial specimens has an average thickness of $52 \,\mu\text{m} \pm 14 \,\mu\text{m}$ (n = 250). No corresponding white layer can be found in the reference specimens.

The formed $CaCO_3$ layer has a rough surface due to the formation of irregular crystals. Both the repair material and the mortar substrate tend to show a good adherence to that layer. This roughness could also partially lead to the increase in observed bond strength. Future tests will elucidate this assumption.

4. CONCLUSIONS

Based on this preliminary study of applying a biodeposition layer for bond enhancement of repair materials, the following conclusions can be drawn:

- A bacterial treatment has been shown to enhance the bonding of a repair product to a mortar substrate. Further research will focus on how to further optimize the bacterial layer to obtain a significantly higher bonding strength.
- The biodeposition film on the mortar surface mostly consisted of calcite and vaterite.
- Both the mortar substrate and the repair material exhibited a good bonding with the calcium carbonate crystals precipitated by the bacteria.

Page 7

Ongoing studies are focusing on further enhancing the bond strength of the composite system by further engineering of the biodeposition layer.

ACKNOWLEDGEMENTS

As Postdoctoral Research Assistants of the Research Foundation-Flanders (FWO-Vlaanderen), Didier Snoeck and Jianyun Wang want to thank the foundation for its financial support.

REFERENCES

- Snoeck, D. and N. De Belie, 'From straw in bricks to modern use of microfibres in cementitious composites for improved autogenous healing – a review', *Constr. Build. Mater.* 95 (2015) 774-787.
- [2] Van Tittelboom, K. and N. De Belie, 'Self-Healing in Cementitious Materials A Review', Materials 6 (6) (2013) 2182-2217.
- [3] Wang, J., D. Snoeck, S. Van Vlierberghe, W. Verstraete, and N. De Belie, 'Application of hydrogel encapsulated carbonate precipitating bacteria for approaching a realistic self-healing in concrete', *Constr. Build. Mater.* 68 (2014) 110-119.
- [4] De Muynck, W., N. De Belie, and W. Verstraete, 'Microbial carbonate precipitation in construction materials: A review', *Ecol. Eng.* 36 (2) (2010) 118-136.
- [5] Adolphe, J.P., J.F. Loubière, J. Paradas, and F. Soleilhavoup, 'Procédé de traitement biologique d'une surface artificielle', E. patent 90400G97.0 (1990).
- [6] De Muynck, W., K. Cox, N. De Belie, and W. Verstraete, 'Bacterial carbonate precipitation as an alternative surface treatment', *Constr. Build. Mater.* 22 (5) (2008) 875-885.
- [7] De Muynck, W., D. Debrouwer, N. De Belie, and W. Verstraete, 'Bacterial carbonate precipitation improves the durability of cementitious materials', *Cem. Con. Res.* 38 (7) (2008) 1005-1014.
- [8] De Muynck, W., S. Leuridan, D. Van Loo, K. Verbeken, V. Cnudde, N. De Belie, and W. Verstraete, 'Influence of pore structure on the effectiveness of a biogenic carbonate surface treatment for limestone conservation', *Appl. Environ. Microbiol.* **77** (19) (2011) 6808-6820.
- [9] Watanabe, J. and M. Akashi, 'Formation of various polymorphs of calcium carbonate on porous membrane by electrochemical approach', J. Cryst. Growth 311 (14) (2009) 3697-3701.
- [10] Lippmann, F., 'Sedimentary Carbonate Minerals', (Springer, UK, 1973).
- [11] Spanos, N. and P.G. Koutsoukos, 'The transformation of vaterite to calcite: effect of the conditions of the solutions in contact with the mineral phase', J. Cryst. Growth 191 (4) (1998) 783-790.
- [12] Giralt, S., R. Julia, and J. Klerkx, 'Microbial biscuits of vaterite in Lake Issyk-Kul (Republic of Kyrgyzstan)', J. Sediment. Res. 71 (3) (2001) 430-435.

Overview of an Ontology-based Approach for Kit-building Applications

Zeid Kootbally Department of Aerospace and Mechanical Engineering University of Southern California Los Angeles, CA, USA zeid.kootbally@nist.gov

Craig Schlenoff Intelligent Systems Division National Institute of Standards and Technology (NIST) Gaithersburg, MD, USA craig.schlenoff@nist.gov

Thomas R. Kramer Department of Mechanical Engineering Catholic University of America Washington, DC, USA thomas.kramer@nist.gov

Satyandra K. Gupta Department of Aerospace and Mechanical Engineering University of Southern California Los Angeles, CA, USA skgupta@usc.edu

Abstract—The Agility Performance of Robotic Systems (APRS) project at the National Institute of Standards and Technology (NIST) is using Web Ontology Language (OWL) ontologies for modeling in a robotic kitting workstation. The new technical idea for the APRS project is to develop the measurement science in the form of an integrated agility framework enabling manufacturers to assess and assure the agility performance of their robot systems. This framework includes robot agility performance metrics, information models, test methods, and protocols. This paper focuses on the information models and describes how they are used to introduce robot agility for the kitting domain. OWL class model files are generated automatically from XML schema model files. Files of OWL instances conforming to an OWL class model are generated automatically from XML instance files by automatically built translators.

Keywords-Web Ontology Language; kit building; robotics; agility; knowledge representation; information model;

I. INTRODUCTION

Robots will be a pervasive part of our lives in the coming decades. Whether it is behind the scenes in assembling products that you use in your everyday life, or helping you parallel park your car, robots are already playing a role in what we use and how we get around.

However robots are not good at everything. For the most part, robots perform best in highly structured environments, where objects are in well-known, predictable locations. Robots are also not known for "thinking on the fly" very well. They are best when they can be trained to perform a very specific activity, which requires a very specific set of motions, and that activity can be performed in the exact same way many hundreds or thousands times. Not surprisingly, robots have been adopted much more in high volume, repeatable operations such as car manufacturing than they have been in smaller job shop type operations where only a handful of similar products are being made at a given time.

Another way to describe this is that robots are not considered agile. But, in order for them to be useful to small manufacturers and to also allow larger manufacturers to offer more automated customization of high volume parts (think cars and cell phones), they need to be. The Agility Performance of Robotic Systems (APRS) project at the National Institute of Standards and Technology (NIST)¹ is addressing this challenge. The goal of this project is to enable agility in manufacturing robot systems and to develop the measurement science which will allow manufacturers to assess and assure the agility performance of their robot systems. Key areas of robot agility include: 1) the ability of a robot to be rapidly re-tasked without the need to shut down the robot for an extended period of time when a new operation needs to be performed, 2) the ability of a robot to recover from errors, so that when a part is dropped, for example, the robot can assess the situation and determine the best way to proceed to accomplish the goal, and 3) the ability to quickly swap in and out robots from different manufacturers so that a company is not tied to a single robot brand.

NIST is developing software, standards, and performance metrics to allow aspects of agility to happen. NIST is developing the Canonical Robot Command Language (CRCL) [1] which is a low-level messaging language for sending commands to, and receiving status from, a robot. CRCL is intended primarily to provide commands that are independent of the kinematics of the robot that executes the commands. This allows robots to be more easily swapped in and out since the robot commands are represented in a robot-agnostic format. NIST is also in the process of developing a set of performance metrics to measure robot agility, and is validating them at an upcoming Agile Robotics for Industrial

Gupta, Satyandra; Kootbally, Zeid; Kramer, Thomas; Schlenoff, Craig. "Overview of an Ontology-based Approach for Kit-Building Applications." Paper presented at First International Workshop on Semantic Robotics, San Diego, CA, United States. January 29, 2017 - February 1, 2017.

¹https://www.nist.gov/programs-projects/agilityperformance-roboticsystems

Automation Competition $(ARIAC)^2$ to be held in 2017 in conjunction with the Institute of Electrical and Electronics Engineers (IEEE).

One of the key aspects needed to enable robot agility is the ability for the robot to represent knowledge about the environment (and its own capabilities) in such a way that it can reason over it and take action based on what it has learned. NIST has chaired an IEEE Working Group which developed the IEEE 1872 Standard (Core Ontology for Robotics and Automation (CORA)) [2]. This standard defines a core ontology that allows for the representation of, reasoning about, and communication of knowledge in the robotics and automation (R&A) domain. This ontology includes generic concepts as well as their definitions, attributes, constraints, and relationships. These terms can be specialized to capture the detailed semantics for concepts in robotics sub-domains. The standard has chosen to use the Standard Upper Ontology Knowledge Interchange Format (SUO-KIF) [3] to represent concepts and their associated axioms.

In this paper, the authors describe how the concepts of three ontologies in Web Ontology Language (OWL) [4] format, consistent with CORA, were specialized to enable agility in industrial robotics and the infrastructure that was built to convert the concepts into various representations that could directly be applied to the robot control system. The effort described in this paper deals with kitting or kit building. In kitting, parts are delivered to the assembly station in kits that contain the exact parts necessary for the completion of one assembly object.

This paper is structured as follows: Section II describes the information models used within this project. The authors also briefly describe the tool used to translate XML (eXtensible Markup Language) Schema Definition Language (XSDL) and XML files to OWL files in Section III. Section IV details the design methodology that relies on OWL files to perform kitting in the APRS project. Section V gives conclusions and future work.

II. INFORMATION MODELS

The APRS project makes use of three ontologies that can be applied to the kitting domain. The authors used a set of closely related C++ software tools for manipulating XSDL [5], [6], [7] files and XML instance files and translating them into OWL class files and OWL instance files [8]. The authors use the translators to translate XML instance files to OWL instance files. The OWL class files are used as input to a tool that generates a database schema automatically. The APRS work in OWL generation was reported in [8]. Modest improvements have been implemented since then.

A. KittingWorkstation Model

The KittingWorkstation model describes the objects in the current kitting scenario. This file contains all of the basic information that was determined to be needed during the evaluation of the use cases and scenarios. The knowledge is represented in as compact of a form as possible with knowledge classes inheriting common attributes from parent classes. In Figure 1 and similar figures (which were generated by XMLSpy³), a dotted line around a box means the attribute is optional (may occur zero times), while a $..\infty$ underneath a box means it may occur more than once, with no upper limit on the number of occurrences. The main types (i.e., classes or datatypes) of the attributes of a kitting workstation are shown in Figure 1. More information on this model can be found in [9], [10].

B. Action Model

Planning for kitting relies on the Planning Domain Definition Language (PDDL) [11] domain file. PDDL is a community standard for the representation and exchange of planning domain models. In order to operate, the PDDL planners require a PDDL file-set that consists of two files that specify the domain and the problem. From these files, the planning system creates an additional static plan file. The authors explored the idea of automating the generation of PDDL domain and problem files by representing the components of a PDDL domain file [12] in an Action model. The Action model imports the KittingWorkstation model to include all the parameters defined in PDDL predicates and functions. Figure 2 depicts the information model for a PDDL domain file. A user only needs to create an XML file which consists of actions that a robot needs to perform to build a kit.

C. Robot Capability Model

The application of robotics in manufacturing assembly is hindered by their lack of agility, their large changeover times for new tasks and new products, and their limited reusability. One of the main causes for these hindrances is the lack of understanding of robot capabilities as they pertain to assembly tasks. In this context, a robot capability refers to the ability of a robot to perform a specific action on a specific object or set of objects. Measuring robot capabilities for a specific domain provide multiple advantages such as creating a process plan that assigns the best robot to each operation needed to accomplish the given task.

The Robot Capability model depicted in Figure 3 was produced based on the set of components identified from a thorough literature review [13]. The capability model

²https://www.nist.gov/el/intelligent-systems-division-73500/agilerobotics-industrial-automation

³Certain commercial/open source software and tools are identified in this paper in order to explain our research. Such identification does not imply recommendation or endorsement by the authors or NIST, nor does it imply that the software tools identified are necessarily the best available for the purpose.

Gupta, Satyandra; Kootbally, Zeid; Kramer, Thomas; Schlenoff, Craig. "Overview of an Ontology-based Approach for Kit-Building Applications." Paper presented at First International Workshop on Semantic Robotics, San Diego, CA, United States. January 29, 2017 - February 1, 2017.



Figure 1. Information model for kitting.

consists of pointers to elements defined in the Kitting-Workstation model as well as new elements. Mandatory elements include a pointer to the robot element for which the capability is defined (e.g., *robot-motoman*), the definition for the assembly action performed by the robot (e.g., *pick-up-part*), one or multiple references to the stock keeping unit for the object involved in the assembly action (e.g., *small-gear*), and the overall result for the current capability (e.g., can pick up this part 85 % of the time).

III. TRANSLATION DETAILS

To enable using OWL the authors have developed a set of automatic software tools for generating OWL class files from XML schemas and for generating XML to OWL instance file translators automatically. The authors use the translators to translate XML instance files to OWL instance files. A number of systems for converting XML files to OWL files have been developed. A survey of nine of these systems was made by Hacherouf *et al.* [14]. That survey does not include



Figure 2. Information model for PDDL.



Figure 3. Information model for robot capability.

the system developed by the authors, in this paper called the ISD OWL Generation System [8]. The ISD system is built in C⁺⁺, using YACC and Lex [15] for parsing. C⁺⁺ is generated from YACC by bison [16] [12] and generated from Lex by flex [17].

A. Obviating OWL Problems

A number of features of OWL [18], [19] and Protégé [20], a tool available for building OWL ontologies, make it impractical to build OWL models and instance files directly. The primary reason for this is that user errors in spelling the names of NamedIndividuals, properties, and classes are not recognized as errors.

OWL's open world assumption allows that anything might be true that is not explicitly ruled out (1) by OWL statements directly, or (2) by reasoning from statements that have been made. The open world assumption is appropriate in some contexts, however the kitting domain may be readily handled under a closed world assumption. Using an open world assumption introduces difficulty without providing any advantages.

Also, if the name of a class, property, or individual is used without being explicitly declared as such in the file (as happens when a name is misspelled), that class, property, or individual is implicitly declared. Protégé does provide some help with spelling by having an auto complete window to use when expressions are being constructed. A misspelled term will appear as one of the choices while the user types, if the first few letters are correct.

Another problem is that, while constructing an OWL file, it is easy to omit OWL statements one intends to make. Omitting any one statement or any set of statements after the header in OWL class and instance files will not be an OWL error and will not cause Protégé to flag any error or give any warning. The same would be true of many other OWL files.

Finally, Protégé does not check completely whether an OWL file conforms to the OWL specification. For example, if an OWL instance file imports an OWL class file and the prefix declared for both files is the empty prefix, no error will be signaled, even though the OWL specification says explicitly that this is not allowed.

Some of these issues can be detected, and research aimed at developing better OWL consistency checkers is ongoing [21]. One utility, Pellet, offers some support for advanced reasoning and debugging [22]. In our tests, the Pellet command line linter was able to detect spelling errors within OWL, but Pellet was unable to detect a missing statement. Further, Pellet seems to support OWL XML syntax, but was unable to parse functional style syntax. Limitations still remain.

The use of an undefined type in an XML schema file is an error, and readily available XML tools will detect and flag it. Similarly, a missing element in an instance file will be detected and flagged. If a definition of references to unique identifiers (IDREF) is made to a definition of unique identifiers (ID) that has not been used, that will be detected and flagged. If a portion of an XML schema file is omitted, in many cases, that will be detected when the file is read, and in most cases an error will be signaled if an instance file is read that conforms to the complete intended schema file. Thus, almost all spelling errors that will pass in OWL will fail in XML, and most errors of omission that will pass in OWL will fail in XML.

The translation tools do not make spelling errors or errors of omission. Hence, by using them on tested XML schemas and instance files, correct OWL files may be produced. In addition, it is easier to work with XML files since (1) they are structured while OWL files are not, and (2) XML files are about half as long as the equivalent OWL files.

IV. DESIGN METHODOLOGY

The authors have implemented the design methodology shown in Figure 4 to perform kitting with an industrial robot.

A manufacturing activity begins when the operator receives a request to fulfill an order. The request is submitted to a Task Scheduler that starts the kitting process. Information from the three ontologies described earlier is converted into a graph database with the Database Generator, a Java tool. The use of a graph database ensures real-time access to information on the environment, planning domain, and robot capabilities. The Database Generator "sees" an ontology as a graph. It has a top node (owl:Thing) and classes extending it. There are individuals that belong to classes and object properties connecting the individuals. Individuals can have data properties and annotations that can be represented as node properties and relationship properties or as relationship types.

Using the Action and Robot Capability ontologies, the Problem Generator [12] dynamically creates a PDDL problem file for the current domain. The Problem Generator parses the PDDL domain file to retrieve all the predicates and numeric-valued fluents (including robot capability values) and then uses a mapping file to retrieve their values in the graph database.

The current state of the world is updated by the Sensor Processing since objects in the environment could have been inadvertently moved since the previous time the database was updated. Once the problem file is autogenerated, both the problem and the domain files are used with a temporal planner to generate a plan file.

Next, an Executor application translates each PDDL action from the plan file into a series of Canonical Robot Command Language (CRCL) commands, and sends each to each Robot Controller. CRCL provides message content and syntax for a low-level robot command-status protocol. The Executor application continually tracks the execution status of each CRCL command, ensuring that the conditions and effects of the PDDL commands are met as the statements are processed. If failures are detected, e.g., dropped part, the Executor aborts the PDDL plan and calls for replanning. The current state of the world is maintained in a world model (WM) graph database which is updated by a sensor processing system, which continually estimates the state of the environment, synthesizes higher-level information about the state of the world by fusing lower-level sensor information, and updates the WM, which could be implemented in a distributed fashion.



Figure 4. Design methodology for kitting.

The design methodology described above has been used at NIST to perform kitting with two different robots (Figures 5 and 6). Kitting was performed with both robots to pick and place components of a gear box (Figure 7). The KittingWorkstation model consisted of elements describing the workstation (robot, gears, etc) while the Action model described the different pick-up and place actions that the robots can use to reach a goal state. The design methodology was not altered in the two scenarios, only the information in the OWL files was modified to match the goal state (e.g., building a kit with two parts vs. a kit with four parts).



Figure 5. Kuka Lightweight Robot.



Figure 6. Fanuc LR Mate 200iD.



Figure 7. Parts for a gearbox kit.

V. CONCLUSIONS AND FUTURE WORK

The Agility Performance of Robotic Systems (APRS) project aims at making industrial robots more agile to

tackle the challenges that small and medium manufacturers are facing. Implementing agility within the APRS project is performed with the kitting domain and starts with the definition of three information models which derive from the CORA model. This paper describes a KittingWorkstation model, an Action model, and a Robot Capability model. The KittingWorkstation model describes a kitting workstation including solid objects (e.g., parts) and data (e.g., poses). The Action model describes the structure components of a Planning Domain Definition Language (PDDL) domain file and is used to automate the generation of PDDL domain and problem files. The Robot Capability model expresses the capability of a robot at performing specific planning actions on a specific object or set of objects. Although significant efforts have been made to improve agility in manufacturing assembly, there is still a need to deal with action failures. It is the intention of the authors to develop a new model for failures. This model will encompass information on the different failures that a robot can encounter during kitting as well as the severity of the failures and a remediation plan for each type of failure.

ACKNOWLEDGMENT

Dr. Kramer acknowledges support for this work under grant 70NANB15H053 from the National Institute of Standards and Technology to the Catholic University of America.

Dr. Kootbally acknowledges support for this work under grant 70NANB15H249 from the National Institute of Standards and Technology to the University of Southern California.

REFERENCES

- [1] F. Proctor, S. Balakirsky, Z. Kootbally, T. Kramer, C. Schlenoff, and W. Shackleford, "The Canonical Robot Command Language (CRCL)," Industrial Robot: An International Journal, vol. 43, no. 5, pp. 495-502, 2016.
- [2] S. R. Fiorini, J. L. Carbonera, P. Gonçalves, V. A. Jorge, V. F. Rey, T. Haidegger, M. Abel, S. A. Redfield, S. Balakirsky, V. Ragavan, H. Li, C. Schlenoff, and E. Prestes, "Extensions to the Core Ontology for Robotics and Automation," Robotics and Computer-Integrated Manufacturing, vol. 33, pp. 3-11, Jun 2015, special Issue on Knowledge Driven Robotics and Manufacturing.
- [3] A. Pease, "Standard Upper Ontology Knowledge Interchange Format," 2009. [Online]. Available: http://suo.ieee.org/suokif.html
- [4] World Wide Web Consortium, "OWL 2 Web Ontology Language Structural Specification and Functional Syntax," 2009.
- [5] W3C, "XML Schema Part 0: Primer Second Edition," in http://www.w3.org/TR/xmlschema-0/, 2004.
- [6] -, "XML Schema Part 1: Structures Second Edition," in http://www.w3.org/TR/xmlschema-1/, 2004.

- [7] World Wide Web Consortium, "XML Schema Part 2: Datatypes, Second Edition," 2004. [Online]. Available: www.w3.org/TR/xmlschema-2
- [8] T. R. Kramer, B. H. Marks, C. I. Schlenoff, S. B. Balakirsky, Z. Kootbally, and A. Pietromartire, "Software Tools for XML to OWL Translation," National Institute of Standards and Technology, Gaithersburg, MD, USA, NIST IR 8068, Jul 2015.
- [9] S. Balakirsky, Z. Kootbally, C. Schlenoff, T. Kramer, and S. Gupta, "An Industrial Robotic Knowledge Representation for Kit Building Applications," in Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vilamoura, Portugal: IEEE, October 2012, pp. 1365-1370.
- [10] S. Balakirsky, Z. Kootbally, T. Kramer, R. Madhavan, C. Schlenoff, and M. Shneier, "Functional Requirements of a Model for Kitting Plans," in Proceedings of the Workshop on Performance Metrics for Intelligent Systems. ACM, 2012, pp. 29-36.
- [11] M. Fox and D. Long, "PDDL2.1: An Extension to PDDL for Expressing Temporal Planning Domains," Journal of Artificial Intelligence Research, vol. 20, pp. 431-433, 2003.
- [12] Z. Kootbally, C. Schlenoff, C. Lawler, T. Kramer, and S. Gupta, "Towards Robust Assembly with Knowledge Representation for the Planning Domain Definition Language (PDDL)," Robotics and Computer-Integrated Manufacturing, vol. 33, pp. 42-55, 2015, special Issue on Knowledge Driven Robotics and Manufacturing.
- [13] Z. Kootbally, "Industrial Robot Capability Models for Agile Manufacturing," Industrial Robot: An International Journal, vol. 43, no. 5, pp. 481-494, 2016.
- [14] M. Hacherouf, S. Bahloul, and C. Cruz, "Transforming XML Documents to OWL ontologies: A survey," Journal of Information Science, vol. 41, no. 2, pp. 242-259, 2015.
- [15] D. Brown, J. Levine, and T. Mason, Lex & Yacc. OReilly Media, October 1992.
- [16] C. Donnelly and R. Stallman, "Bison, The YACCcompatible Parser Generator," 2006. [Online]. Available: http://dinosaur.compilertools.net/bison
- [17] V. Paxson, W. Estes, and J. Millaway, "Flex, Version 2.5.31 A Fast Scanner Generator," 2003. [Online]. Available: http://www.gnu.org/software/flex
- [18] World Wide Web Consortium, "OWL 2 Web Ontology Language Primer (Second Edition) W3C Recommendation 11 December 2012," 2004. [Online]. Available: http://www.w3.org/TR/owl2-primer/
- "OWL 2 Web Ontology Language Structural [19] Specification and Functional Style Syntax (Second Edition) W3C Recommendation 11 December 2012," 2012. [Online]. Available: http://www.w3.org/TR/owl2-syntax
- [20] M. Horridge, "A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools, 1st ed," The University Of Manchester, Tech. Rep., March 2011.

Gupta, Satyandra; Kootbally, Zeid; Kramer, Thomas; Schlenoff, Craig. "Overview of an Ontology-based Approach for Kit-Building Applications." Paper presented at First International Workshop on Semantic Robotics, San Diego, CA, United States. January 29, 2017 - February 1, 2017.

- [21] B. Motik, I. Horrocks, and U. Sattler, "Adding Integrity Constraints to OWL," *Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions*, vol. 258, June 2007.
- [22] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz, "Pellet: A Practical OWL-DL Reasoner," Web Semantics: science, services and agents on the World Wide Web, vol. 5, no. 2, pp. 51–53, 2007.

Integrating Finite Element Analysis with Systems Engineering Models

Jerome Szarazi (Koneksys, United Kingdom);

> Axel Reichwein (Koneksys, United States);

Conrad Bock (National Institute of Standards and Technology, United States);

Abstract

In order to promote traceability, consistency, interoperability and better collaboration between systems engineering and Finite Element Analysis (FEA)-based simulation activities, we propose a tool-independent description of FEA models that integrates with the Systems Modeling Language (SysML), for future standardization. As technical systems become more complex, it is important to support traceability between systems engineering artifacts, such as requirements, and test cases, and corresponding FEA artifacts, such as FEA models, simulation conditions, and results.

While there is a standard for model-based systems engineering in the form of SysML, there is no standard description of FEA models. Existing FEA model descriptions are incomplete, tool-specific, informal or a combination of these. As a result, interoperability between FEA software applications is compromised, and communication between engineers is inefficient. A standard for the description of FEA models is difficult to develop, as the geometry, mathematics and physics of finite elements can vary greatly.

We propose a finite element mathematics specification based on recent works [1] and on the topological characteristics of finite elements that is formal, precise, and understandable to engineers. Mathematical expertise is still required to consistently set attributes of the specification but it can potentially capture new kinds of elements. The specification removes dependence on finite element names, which are sometimes inconsistent. We think that such a description is suitable for broad adoption among both FEA and systems engineers. We validate the new description of finite element mathematics by solving FEA problems using Python code we developed and demonstrate that these elements can be described in SysML.

Presented at the NAFEMS World Congress 2017

Stockholm, Sweden | 11-14 June 2017

Bock, Conrad; Reichwein, Axel; Szarazi, Jerome. "Integrating Finite Element Analysis with Systems Engineering Models." Paper presented at NAFEMS World Congress 2017, Stockholm, Sweden. June 11, 2017 - June 14, 2017.

1. Introduction and motivation

The motivation of this work is to better integrate FEA expert activities into the development lifecycle by integrating them with systems engineering modeling. We believe that it will benefit cross-disciplinary communication, traceability, model exchange, interoperability, model reusability, process knowledge and eventually process automation.

Efficient cross-disciplinary communication is essential for collaboration. System engineers have an overview of development activities by continually taking input from many stakeholders, making system-level decisions for improving the design, and communicating the results to cross-disciplinary teams for further design and validation. Based on system-level decisions, discipline-specific engineers create or reconfigure models for evaluating and validating new and modified designs. FEA models are used for validation across many engineering disciplines to simulate thermal, fluid, mechanical, electrical, magnetic problems or a combination of these. It is essential that FEA specialists, other engineering disciplines and system engineers share a common understanding of FEA models across disciplines to leverage a crossdisciplinary communication.

Sharing a common model representation across FEA tools will facilitate traceability from system requirements to FEA models. Product development is an iterative process where requirements change over time. The product specification may undergo many changes that are often decided at system level, but affect FEA models. To analyse the impact of a change, it is important that system requirements can be traced to FEA model artefacts. For example, a system cost reduction process may require a material replacement that should be translated into a material parameter change of the FEA model used to validate a safety requirement. By tracing links from requirements to models, system engineers can directly identify the impact of cost reduction on safety requirements and collaborate with FEA specialists to make better decisions. In addition, FEA engineers can better understand the context and value of their simulations in relation to customer needs.

Interoperability between applications is a major concern when selecting a tool, as it improves the data and information sharing within the same or with external organizations. It also avoids binding users to a single tool environment, making migration to another tool framework difficult. A standardized, tool independent FEA model representation will help to store and load a model from one application to the other. With standardized model interfaces, relationships and interdependencies between models of different abstraction can be defined. For example, this has been applied to integration of lumped parameter and systems engineering models [2].

Interoperability has two main levels, syntactic and semantic. The first facilitates the data exchange between tools by defining a common serialization format. The second helps to share context and define concept equivalences, enabling models to give the same simulation results across tools, and facilitate transformation between different kinds of tools. A tool independent FEA representation integrated with systems models will help to support syntactic and semantic interoperability between FEA tools, as well as model transformation between FEA tools and systems engineering tools.

Standards should help define unambiguous models that can be identified, understood and become reusable knowledge in engineering processes. For example, it is common that multiple simulations are required to validate a design. Validation procedures and other FEA tasks can be mapped into a workflow representing a chain of simulation actions performed on models, where model input and output can be connected. When the models used in workflows are interoperable, processes can be efficiently rearranged, process knowledge can be shared, and processes continuously improved.

This process knowledge can be exchanged with other disciplines. An example is the simulation of manufacturing processes using FEA. FEA is not only used for product design verification but also to simulate manufacturing processes. Typically, FEA calculates residual stress generated by forming, subtracting, or joining material in the assembly process. These simulations help to evaluate manufacturing process impact on design integrity by feeding-back residual stress results to the design. Another benefit is the improvement of simulation results and robustness by comparing simulation results to physical tests in a consistent manner. These comparisons can increase the quality of simulations.

Integrating FEA into system engineering requires an abstraction of FEA knowledge, leading to standards that will help overcome the heterogeneity of FEA model syntax and semantics. We start by evaluating existing standards in Section 2, then review FEA and some challenges in characterizing it abstractly in Sections 3 and 4, respectively. Section 5 proposes a new tool-independent specification for the mathematical portion of finite elements and provides a SysML model for it. Section 6 summarizes the paper and outlines future work.

2. Status of standardization

SysML is an open graphical modeling specification that extends a subset of the Unified Modeling Language (UML) [3,4]. UML has been successfully used for architecting software applications, while SysML provides simple and efficient constructs to model a large variety of systems engineering problems. It is tool and methodology independent and supports model and data interchange via the eXtensible Markup Language [5].

SysML defines diagrams that illustrate system component relationships under different views, as well as semantics for its notation. SysML provides diagrams for requirements, structure, behavior and parametric relationships. These diagrams support specification, analysis, design, verification of systems. SysML supports syntactic and semantic interoperability for system engineering models.

For FEA activities, International Standards Organization 10303 (STEP) Application Protocol 209 edition 2 (AP209ed2) is the reference standard for multidisciplinary analysis and design product data with respect to interoperability, legacy data archiving and information reusability [6]. It integrates simulation data management, computer aided design, computer aided engineering and product data management. One of the main benefits of using a STEP standard is that FEA uses Computer Aided Design (CAD) geometries as input and STEP AP242 is a successful standard for CAD 3D geometry, assembly and Product Manufacturing Information interoperability [7]. The use of these two application protocols could support a seamless data exchange between design and analysis. Considering that the transformation of CAD data to FEA information is a major bottleneck in analysis activities, supporting interaction between design and analysis in a bi-directional manner would improve productivity.

Unfortunately, when reviewing STEP AP209, we noticed it mainly captures Product Lifecycle Management (organisation, date, etc) and geometric information. For the simulation part, AP209 defines a FEA model with a collection of non-constrained string entities (analysis type, creating software, finite element name, material...) leading to an informal FEA description, as in Figure 1. STEP AP209 provides only syntactic interoperability, not semantic interoperability. It cannot ensure the same simulation results across FEA tools or support integration with other kinds of tools. A more formal FEA description for data exchange is needed.

#637538282= FEA_MODEL_3D('Identification',(#637538284),#637538291, 'NASTRAN BDF Converter v0.0.0',('NASTRAN'),'AnalysisModelType');

Figure 1: Example of a FEA model definition in a P21 file following AP209

Interoperability between software applications and the communication between engineers, including system engineers and FEA engineers, is therefore compromised. In order to integrate finite element analysis with system engineering we need to characterize FEA and define a tool-independent abstraction that can then be modelled in SysML.

3. Overview of FEA

Finite element analysis is a numerical method for finding approximate solutions to partial differential equation or energy minimizing problems by solving linear algebraic equations for steady state or ordinary differential equations for dynamic problems [8]. FEA models require derivatives of functions of space (geometry) and time, as opposed to lump parameter systems, which only involve derivatives of functions of time.

FEA is a very powerful numerical method with attractive properties:

- *Modularity:* A domain specific physics problem can be modelled algebraically for a simple polygonal entity using interpolation functions. The resulting algebraic model can be stored as a software module or function, called a library element.
- *Reusability:* A library element can be reused in analyses of multiple systems.
- *Scalability:* Library elements are mapped to mesh cells partitioning system geometry. A complex system composed of multiple finite elements can be assembled and described by a global stiffness matrix.

FEA is typically used to solve multidimensional physics problems describing a system under the action of body and surface loads and kinematically constrained. The algebraic element relating loads to unknowns is the stiffness matrix K in case of static problems, and the mass matrix M and damping matrix D in case of dynamic problems.

The FEA procedure is simple and reproducible. It can be described by a workflow consisting of three steps, the pre-processing phase (mesh generation, material assignment and boundary labelling), analysis definition phase (finite element selection, model parameter setting, solver definition and output settings), and post-processing phase (visualization and data export) [9].

FEA applications have software architecture that follows the workflow. At the file level, a completed analysis will be composed of pre-processing files (CAD and mesh files), an analysis definition file or input file, and results files. Modern FEA software provides graphical user interfaces to support workflow. This is very helpful considering that many parameter settings are geometrical, such as boundary conditions definition. Another useful feature is the bi-directional associative interface providing a seamless integration of CAD and FEA data.

Bock, Conrad; Reichwein, Axel; Szarazi, Jerome. "Integrating Finite Element Analysis with Systems Engineering Models." Paper presented at NAFEMS World Congress 2017, Stockholm, Sweden. June 11, 2017 - June 14, 2017.

4. Challenges in characterizing finite elements analysis

Finding a unified description of FEA is difficult in part because a large number of possible FEA analyses result from the many options involved:

- Modern software provides a vast collection of FEA library elements to simulate mechanical, thermal, electrostatics, magneto-statics, fluid and electromagnetism. Multiphysics problems, which are systems consisting of more than one component governed by their own physics principles, are also supported.
- System studies can be time independent and will be evaluated by running a steady state analysis. For time-dependent systems transient analysis will be performed.
- System stability can be assessed by performing a modal analysis to find resonance frequencies or modes. Buckling analysis helps to avoid failure modes for solid system components under compression.
- The analyses above require material model definitions, which can be linear or nonlinear. Material models can also involve coupling in another physics, as for example in thermo-mechanical problems. For solid material under some large deformation, geometric nonlinearities will be considered, resulting in a re-meshing operation at each simulation step.
- From a numerical perspective, discretizing an analysis by choosing one or more discretization variables can affect simulation performance.
- Depending on the problem abstraction, a problem can be described in 1D, 2D or 3D. For symmetric systems or systems under rotation, cylindric or polar coordinate system for example can be chosen to simplify the analysis.
- To run an analysis on the discrete geometric system model (mesh), the finite element type, which is the core numerical element of the analysis, needs to be selected.

The total combination of these choices would provide a rough evaluation of the number of possible FEA models. If we consider a line element for example, we could apply 4 different physics and support simulation in 3 different dimensions, then we would have 12 combinations and therefore 12 FEA library elements. As a result, modern FEA applications involve a vast number of choices and results that are difficult to track.

The FEA process generates many artefacts that are difficult to organize. Multiphysics problems can be solved by custom code or tool vendor solutions having their own artefacts. Software vendors define their own FEA model referencing schemas, some of which are nonexplanatory while others try to incorporate some of the model parameter information. Expert know-how is required to find equivalences between models. As AP 209 does not provide any schema to classify finite elements, tools use their proprietary names for their APIs and for referencing their FEA models.

5. New proposed FE mathematics specification

In order to ensure efficient communication between system engineering and FEA, we need a standard convention to describe finite elements. It should include a compact notation that contains all information necessary to solve an FEA problems in a platform independent manner, by automatically generating shape functions, element stiffness matrices based on a problem specification. When characterizing an FE simulation, we had previously concluded that many kinds of FEA library element models are possible. A classification schema for these models is needed.

In this paper, we only discuss specification of FE mathematics, which we define as the mathematical objects representing polynomials with one or many variables over a geometric domain. These are the core objects of FEA simulation. This section outlines the challenges in finding such a description and propose a compact notation and SysML model that is unambiguous and understandable by most engineers.

a. Previous work on FE mathematics classification

As we started our search for a finite element mathematics description, we soon realised that most descriptions are ambiguous. The same FE-mathematics can be referred to as many different names in the literature. For example, a linear line element is also described as a Lagrange line element. Many FE mathematics are named after their discoverer(s) but this is sometimes not even clear as to which name to use. A Lagrange triangle can for example also be called a Courant element. Furthermore, the use of mathematician names does not provide any information for non-specialists.

Describing FE mathematics with engineering names, such as beam or bar element, is also ambiguous. A beam, for example, can have 4 degrees of freedom or 2 degrees of freedom. In the first case, the representation of 2 transversal displacements and 2 rotation angles requires a cubic polynomial or Hermitte element. In the second case, it requires only a Lagrange element to interpolate the moment. To find a description of FE mathematics, it is important to remove name dependencies, separate the physics from the mathematics, and classify or characterize FE mathematics by their properties.

FE mathematics classification work seems to have been first undertaken by Ciarlet. Modern finite element literature refers also quite often to his definition. In [10] an FE mathematics is defined as a triplet (K, P, Σ) consisting of a geometric figure K, a set of basis functions or basis space P and degrees of freedom Σ (DOFs). Although the original definition is simple, it seems that Ciarlet's objective was to describe families of FE mathematics. To further categorize them, Ciarlet introduced the typing of the geometry K. By doing so, he could establish relationships between a typed geometry K and the degree of the basis functions that are valid for any dimension. Although this additional categorization provided great insight on mathematical properties, it made the description of FE mathematics more complex.

Ciarlet's work was extended in [11], leading to a periodic table of finite elements covering differential forms of all possible degrees for each function space families. The table helps to understand the relationship between FE mathematics families and the application scope of the captured FE mathematics. The limitation is it does not classify some common elements.

b. New proposed FE mathematics specification

The new FE mathematics description is based on the generic finite element definition of Ciarlet and the DOF type description as found in [1]. However, the triplet (geometry, DOF and basis space) is described using topology. Furthermore, the new FE mathematics description is coordinate-independent, in contrast to Ciarlet's more complex coordinate-dependant description. Also, the description of DOFs is uncoupled from the finite element geometry. In contrast, Ciarlet's description is based on a coupling between DOFs, finite element geometry and basis space, leading to a less compact and more complex classification of finite elements.

To address the problem, we propose a compact, simple and unambiguous description of the mathematics of finite elements. This description reflects the mathematical properties of finite elements and provides all necessary information for implementation. The proposed FE mathematics specification can describe any finite element as it can be applied to any discrete geometry and can specify any basis space. The description is composed of a triple as defined in [10], but the information entities (geometry, DOFs, basis space) can be defined independently of each other and are thereby reusable. Mathematical expertise is required to consistently set attributes of the three entities in a finite element, but the specification can potentially capture new kinds of elements. The specification removes dependence on the finite element names which are sometimes inconsistent.

The new proposed FE mathematics specification is defined as follows:

- Geometry K (line, polygon, polyhedron, prism...)
- DOF sets on geometry K:

$$\boldsymbol{\Sigma}(K) := \{ \boldsymbol{\Sigma}_{Cn}(K), \boldsymbol{\Sigma}_{Cn-1}(K), \dots, \boldsymbol{\Sigma}_{C0}(K) \}$$

where $\Sigma_{Cn}(K)$ is the set of DOFs applied to all n-faces of the geometry (0-faces C_0 : vertices, 1-faces C_1 : edges, 2-faces C_2 : faces, 3-face C_3 : volumes).

A DOF is defined by a type and a count which specifies the number of evaluated requirements on one n-face of K.

• Basis functions set or basis space **P**, defined by a string

Geometry type

The table below show a non-exhaustive list of geometries used in FEmathematics and their respective string encodings.

Geometry	Geometry name	Encoding string
••	Line	L
\bigtriangleup	Triangle	Т2
	Square	S
\bigcirc	Tetrahedron	T3
\bigcirc	Square pyramid	SP
Ð	Cube	С
\bigwedge	Triangular prism	ТР

Table 1: Listing of common geometric figures used in FEA and their corresponding encoding string

Degrees of Freedom type (DOF type)

The set of all possible DOFs Σ_f types is given by:

$$\Sigma_f = \{p(\boldsymbol{x}): \boldsymbol{P}\boldsymbol{E}; \ \nabla p(\boldsymbol{x}): \boldsymbol{F}\boldsymbol{D}; \ \frac{\partial^2 p(\boldsymbol{x})}{\partial^2 \boldsymbol{x}}: \boldsymbol{S}\boldsymbol{D}; \ \dots \}$$

where p(x) is a scalar or p(x) a vector-valued polynomial describing physical quantities. To facilitate reading, acronyms are used for literal operator definition as in [1], such as PE for point evaluation, see the table below.

DOF type	Definition [1]	Symbol	Mathematical expression of the linear form and shape function definition	
Point evaluation	Point evaluation of a function P at a vertex of the geometry	PE	$p(\mathbf{x})$	
First derivative	Point evaluation of all first derivatives of the function P at a vertex of the geometry	FD	$\nabla p(\mathbf{x})$	
Second derivative	Point evaluation of all second derivatives of the function P at a vertex of the geometry	SD	$\frac{\partial^2 p(\boldsymbol{x})}{\partial^2 \boldsymbol{x}}$	
Directional derivative	Evaluation of the directional derivative of the scalar function v in the direction n at the point x.	DD	$\nabla p(\mathbf{x}) \cdot \mathbf{n}$	
Tangential component	Evaluation of the vector- valued function v in the tangential direction along line of the geometry at the point x	VT	$\int_0^l p(x).tds$	
Normal component	Evaluation of the vector- valued function v in the normal direction along a face of the geometry at the point x	VN	$\int_0^l p(x) \cdot n ds$	
Interior moment	Interior moment degrees of freedom; that is, degrees of freedom defined by integration against a weight function over the interior of the domain K	IM	$\int_{K} \boldsymbol{p}. \boldsymbol{q} dS$	

Table 2: Definition of DOF types and corresponding encodings according to [1]

Presented at the NAFEMS World Congress 2017

Bock, Conrad; Reichwein, Axel; Szarazi, Jerome. "Integrating Finite Element Analysis with Systems Engineering Models." Paper presented at NAFEMS World Congress 2017, Stockholm, Sweden. June 11, 2017 - June 14, 2017.

Basis space P

The FE mathematics specification also defines basis space *P*. The geometry K, the DOFs $\Sigma(K)$ and the basis space *P* are interdependent attributes. In [10] and [11], FE mathematics is further categorized by defining a family of basis space P. Its definition is abstract and difficult for engineers to read.

To facilitate implementation of FE mathematics, the specification provides a mechanism to capture the set of basis functions P using lexicographic ordering of monomials, following many symbolic computing programs that generate monomials this way. The method consists of creating a dictionary of all possible monomials combinations or words. These generated words are then ordered and indexed. We chose graded lexicographic ordering, which first sorts by total degree, then lexicographically. As an example, the set $S_2 = \{1, x, y, xy, x^2, y^2, ...\}$ of generated monomials can be ordered by $grlex: ..., y^2 > xy > x^2 > y > x > 1$, following graded lexicographic ordering. In general, the representation of the monomials of the basis space P would only require using a subset of this ordered list. To represent this subset, we set 0 or 1 depending on whether a monomial word is used or not. A bit string to capture the monomials of the basis space P is then created following the graded lexicographic ordering. For example, the basis space $P_1 = \{1, x, y, xy\}$ used for a square can be represented as 11101. This bit string can be further compacted into the hexadecimal number 0x1D.

To complete the FE mathematics specification, we need to define the dimension. To create monomial words, we use predefined alphabet letters. This monomial alphabet is composed of a set of letters or symbols where one of the symbols represents a dimension. The set of generated monomials is therefore dimension dependant. In our examples, the letters x, y, z are used and represent space dimension. The set $S_1 = \{1, x, x^2, x^3, ...\}$ represents one dimensional monomials and S_2 would be the set of two-dimensional monomials. The space dimension is prefixed with ND where N is the space dimension. Depending on the field type, a function space can have one or many components. (scalar polynomial field, vector polynomial field, ...) The component index is specified by NC where N is the component index. The basis space $P_1 = \{1, x, y, xy\}$ would be encoded as in the table below.

Table 3: Example	of scalar	[,] polynomial	encoding
------------------	-----------	-------------------------	----------

Space dimension	Field component	Basis space code
2D	1C	0x1D

The encoding depicted in Table 3 (2D-1C-0x1D) can be translated into a symbolic math program capable of translating it to a human readable form. To illustrate we decode 2D-1C0x1FB. Again, the "2D" gives the dimension, which means the basis functions use two variables (x and y, see below). A set of monomials can be generated and then ordered following graded lexicographic ordering. The "1C" indicates that we have only one component therefore representing a scalar polynomial. The "0X1FB" is a hexadecimal number which can be translated into a bit word declaring what monomial to use in the ordered list. The resulting basis space is $P_2 = \{1, x, y, x^2, xy, y^2, x^2y, xy^2\}$, which is used for an FE mathematics of type "serendipity element" as described in [12].

monomials set S_2	<i>x</i> ⁵	xy^2	1	<i>x</i> ²	x^3y					
Graded lexicographic ordering	1	x	у	<i>x</i> ²	xy	y^2	<i>x</i> ³	x^2y	xy ²	<i>y</i> ³
0x1FB	1	1	1	1	1	1	0	1	1	0
Basis functions	1	x	у	<i>x</i> ²	xy	<i>y</i> ²	0	x^2y	xy ²	0

Table 4: Reading basis functions from basis space code 2D-1C0x1FB

The main benefit of this representation of basis space P is to provide all necessary information for a straightforward implementation, without requiring abstract mathematical knowledge. The basis space P_1 for example is referred to as a tensor space and is constructed by computing the tensor product of the basis space $B_1 = \{1, x\}$ and the basis space $B_2 = \{1, y\}$, which requires special expertise [13]. Using the representation above, the implementation of FE mathematics in custom code does not require knowledge about constructing spaces of basis functions.

Application examples

Figure 2 shows an example of a mathematical finite element used for beam calculation. This would read: " $\Sigma_0(K)$ defines a point evaluation and a first derivative requirement applied to each 0-face (vertex) of the line geometry *K*". This is a compact definition of a 1D-Hermitte element. Aggregation of physical information (transversal displacement, rotation) or geometric information (dimension, Cartesian...) can extend the specification.



Figure 2: Application example for a Hermitte element

Figure 3 shows another FE mathematics example used in thermal simulation. FE mathematics specification describes it as a triangle geometry K composed of two DOFs sets $\Sigma_0(K)$ and $\Sigma_1(K)$ where $\Sigma_0(K)$ defines point evaluation applied to each 0-face of K and $\Sigma_1(K)$ defines one point evaluation applied to each 1-face (edge) of K.



Figure 3: Application example for "a parabolic triangle"

The specification can potentially capture new elements. For example, the requirements previously defined for the triangle of Fig 3 could be reused and applied to any geometry, such as the hexagonal prism in Fig 4.



Figure 4: Application example for "a parabolic triangle"

Additional examples and their respective encodings are shown in the table below.

Table 4: Application examples of the FE specification

Figure	Common name	Compact description	Code
••	Linear element, Lagrange element, bar element	K = Line $\Sigma_{C_0(K)} = \{PE\}$ $P_K = 1D-1C0x3$	L COPE 1D-1C0x3
••	Quadratic element, Lagrange quadratic		L COPEC1PE1 1D-1C0x7
\bigtriangleup	Linear triangle, Lagrange linear triangle, courant element, T3	K = Triangle $\Sigma_{C_0(K)} = \{PE\}$ $P_K = 2D-1C0x7$	T2 C0PE 2D-1C0x7
\bigtriangleup	Quadratic triangle, T6	K = Triangle $\Sigma_{C_1(K)} = \{PE, 1\}$ $\Sigma_{C_0(K)} = \{PE\}$ $P_K = 2\text{D}-1\text{C0x3F}$	T2 C0PEC1PE1 2D-1C0x3F
	Hermitte triangle		T2 C0PEFDC2PE1 2D-1C0x3FF
	Linear square element, Lagrange square element	K = Square $\Sigma_{C_0(K)} = \{PE\}$ $P_K = 2D-1C0x1D$	S COPE 2D-1C0x1D
	Serendipity quadratic element		S COPEC1PE1 2D-1C0x1FB
	Argyris element		T2 C0PEFDSDC1DD1 2D-1C0x1FFFFF
	Raviart-Thomas	K = Square $\Sigma_{C_1(K)} = \{VN, 1\}$ $P_K = 2D-1C0x3-2C0x5$	S C1VN1 2D-1C0x3-2C0x5
\bigcirc	Linear Tetrahedron, 3- simplex linear Lagrange element	K = Tetrahedron $\Sigma_{C_0(K)} = \{PE\}$ $P_K = 3D-1C0xF$	T3 C0PE 3D-1C0xF

Presented at the NAFEMS World Congress 2017 Stockholm, Swed Bock, Conrad; Reichwein, Axel; Szarazi, Jerome. "Integrating Finite Element Analysis with Systems Engineering Models." Paper presented at NAFEMS World Congress 2017, Stockholm, Sweden. June 11, 2017 - June 14, 2017.

Representation in SysML

The new specification provides a compact and precise description of FE mathematics that can be modeled in a SysML block definition diagram. SysML block definition diagrams are based on UML class diagrams, where blocks are based on UML classes. A FE mathematics block composes a geometry block, one to many DOF set block(s) and a basis space block. Blocks are detailed by properties that specify types for their values. The model uses integers for space dimension and DOF count on a face. The basis space definition is of type string. The DOF type, the face type and the geometry are defined by enumeration literals, which are an ordered set of names. The enumeration for geometry, for naming of the faces and DOF type are shown in the diagram. Property values can be produced (derived) from other information, indicated by a forward slash ('/'). Such properties could be the number of faces composing the geometry or the total count of DOFs.



Figure 5: SysML block definition diagram of the FE mathematics specification

Bock, Conrad; Reichwein, Axel; Szarazi, Jerome. "Integrating Finite Element Analysis with Systems Engineering Models." Paper presented at NAFEMS World Congress 2017, Stockholm, Sweden. June 11, 2017 - June 14, 2017. The FE mathematics block definition diagram is an abstract model reflecting the general description of FE mathematics. An instance diagrams is derived from it and shows objects specified by property values.

The serendipity element of table 5 is captured in the instance diagram of figure 6 which specifies the geometry type as Square, basis space with the encoding "2D-1C0x1FB", the two DOFs set "C1PE1" and "C0PE" with their respectives DOFs type, DOF number and the facetype to which they are applied. FE mathematics objects can therefore be captured and documented in instance diagrams.



Figure 6: SysML diagram of an FE Mathematics instance

Validation and testing

The FE mathematics specification is used to calculate shape functions space, which form the basis of the dual space of linear forms represented by the DOFs. To find shape functions, the procedure transforms the basis space P into the shape function space, which is another basis, where one DOF, or by extension a discrete physics value (temperature, displacement, etc) located at a point or distributed on a point set of the geometry is represented by a shape function.

This procedure requires a definition for the reference coordinates of the geometry, which can be arbitrarily chosen. Depending on the coordinate choice, the shape functions will be different. The coordinate information is therefore implementation dependent and set by the FE code developer. This is one of the reasons why the FE mathematics specification has been defined in a coordinate independent manner. If shape function information is shared, then the reference coordinates of the geometry need to be communicated.

To validate the specification, we developed our own symbolic code using the Sympy library, a python based library for symbolic mathematics [14,15]. Another implementation of symbolic FE mathematics can be also found in [16]. Symbolic code promotes interoperability by supporting many serialization formats. In Sympy, a mathematical expression is stored as an object which can be "printed" into many formats such as a string, LaTex expression, MathML, some image formats such JPEG and help with the documentation process, as illustrated in the figures below. Other methods can transform the symbolic mathematical expression into code such C, Fortran, Javascript, Theano or python. We implemented our python code to calculate shape functions based upon the specification information. The figure below shows the information model for a linear triangle.



Figure 7: Computation of shape function's symbolic expression



Figure 8: Python object of a "parabolic triangle" for specific reference geometry coordinates

Shape functions are the basis of the discretization process and therefore an essential step in the FEA procedure. Physics variables are expressed with shape functions and DOFs. This expression is then implemented into the weak form, an equation, minimizing the approximation error or the energy of the physics problem, depending on the approach.

The resulting expression leads to an algebraic system relating DOFs to loads. In the case of steady state problems, stiffness matrices relate DOFs to loads. The figure below shows a stiffness matrix symbolically computed for a finite element describing a line element under the influence of axial loads in a 2D space. A Sympy module can transform symbolic expressions of stiffness matrices into lambda functions, by stating coordinates as inputs and stiffness matrix components as outputs. Lambda functions can calculate numerical values very fast. Using a lambda function of the stiffness matrix, a problem composed by many line elements can be assembled for a solver.

$\begin{bmatrix} \frac{(x_0-x_1)^2}{(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2)^{\frac{3}{2}}} & \frac{(x_0-x_1)(y_0-y_1)}{(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2)^{\frac{3}{2}}} \\ \frac{(x_0^2-x_1)(y_0-y_1)}{(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2)^{\frac{3}{2}}} & \frac{(y_0-y_1)^2}{(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2)^{\frac{3}{2}}} \\ -\frac{(x_0-x_1)(y_0-y_1)}{(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2)^{\frac{3}{2}}} & -\frac{(x_0-x_1)(y_0-y_1)}{(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2)^{\frac{3}{2}}} \\ -\frac{(x_0-x_1)(y_0-y_1)}{(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2)^{\frac{3}{2}}} & -\frac{(x_0-x_1)(y_0-y_1)}{(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2)^{\frac{3}{2}}} \end{bmatrix}$	$-\frac{(x_0-x_1)^2}{\left(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2\right)^{\frac{3}{2}}}-\frac{(x_0-x_1)(y_0-y_1)}{(x_0-x_1)(y_0-y_1)}}{\left(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2\right)^{\frac{3}{2}}}-\frac{(x_0-x_1)^2}{\left(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2\right)^{\frac{3}{2}}}}{\left(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2\right)^{\frac{3}{2}}}$	$-\frac{(x_0-x_1)(y_0-y_1)}{(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2)^{\frac{3}{2}}},\\ -\frac{(y_0-y_1)^2}{(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2)^{\frac{3}{2}}},\\ \frac{(x_0-x_1)(y_0-y_1)}{(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2)^{\frac{3}{2}}},\\ \frac{(y_0-y_1)^2}{(x_0^2-2x_0x_1+x_1^2+y_0^2-2y_0y_1+y_1^2)^{\frac{3}{2}}}.$
--	--	---

Figure 9: Symbolic stiffness matrix

6. Conclusion and future work

Model interoperability is essential to improving collaboration between FEA and other engineering disciplines. This is facilitated by systems engineering models acting as information hubs, which need to integrate FEA information. A successful integration of systems engineering models and FEA improves communication, traceability, model reusability and process knowledge in the development process. A key enabler for successful integration is standards. While there is a standard for systems engineering models (SysML), we found that standards for FEA are imprecise and ambiguous. A unified description of FEA is difficult in part because a large number of possible FEA analyses result from the many options involved.

To overcome this difficulty, we first decomposed FEA information by describing FE mathematics in a compact way that does not require abstract mathematical knowledge. The representation assigns degrees of freedom to topological entity types and provides a systematic method to capture basis functions. We tested the compact description on most FE mathematics and then created a SysML model to represent it. We tested this model by implementing symbolic code to generate shape functions. Some of these shape functions were also used to generated symbolic stiffness matrices.

We believe that the new FE mathematics representation can be consistently exchanged, providing all necessary information for implementation and FE mathematics model reusability. Implementation using symbolic computing further supports model reusability and the serialization of symbolic expressions into many formats promotes interoperability. As a result, we have an

unambiguous and compact specification of FE mathematics that is readable by most engineers and can potentially capture new elements.

The code is currently limited to using the string encoding of the FE specification, which will be addressed by a translator between SysML and the code, still under development. The specification of FE mathematics is a first step to consistently represent FEA problems. Extension of the new finite element specification to cover FEA physics is on-going work. The ultimate objective is to describe FEA models using SysML by extending the FE mathematics specification to capture relationships between systems engineering and FEA-related simulation information. This FEA information model should provide all information necessary to assemble a problem for a solver.

Acknowledgements

We would like to thank Peter Denno and Li Ma for input to this paper.

This work was performed under grant awards 70NANB14H251 and 70NANB16H174 from the U.S. National Institute of Standards and Technology.

Commercial equipment and materials might be identified to adequately specify certain procedures. In no case does such identification imply recommendation or endorsement by the U.S. National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References

- [1] Logg A. et Al. (2012). Automated solution of differential equation by the *finite element method*: Springer.
- [2] Dadfarnia, M., Bock, C., Barbau, R. (2016). An Improved Method of Physical Interaction and Signal Flow Modeling for Systems Engineering: Conference on Systems Engineering Research.
- [3] Object Management Group (September 2015). *OMG Systems Modeling LanguageTM*, version 1.4: http://www.omg.org/spec/SysML/1.4.
- [4] Object Management Group (March 2015). *OMG Unified Modeling LanguageTM*, version 2.5: http://www.omg.org/spec/UML/2.5.
- [5] W3C (September 2006) Extensible Markup Language (XML) 1.1 (Second Edition)

- [6] ISO 10303 STEP AP 209 ed2 (2013). Composite and metallic structural analysis and related design: ISO – International Organization for Standardization.
- [7] ISO 10303-242 STEP AP 242 "Managed model based 3D engineering: ISO

 International Organization for Standardization
- [8] Hulbert, G. (1992). *Time finite element methods for structural dynamics*: International journal for numerical methods in engineering. vol. 33, 307-331.
- [9] Bathe, K-J. (1996). Finite element procedures: Prentice hall.
- [10] Ciarlet, P. (2002). The finite element method for elliptic problems: SIAM.
- [11] Logg, A., Arnold D. (2014). Periodic table of finite elements: Siam News.
- [12] Arnold, D., Awanou, G. (2011). *The Serendipity Family of Finite Elements*: The journal for the Society for the foundations of Computational Mathematics.
- [13] McRac, A., Bercea, G-T., Mitchell, L., Ham, D., Cotter, C. (2014). Automated generation and symbolic manipulation of tensor product finite elements: SIAM Journal on Scientific Computing.
- [14] Van Rossum, G. (2007). *Python programming language*: http://www.python.org.
- [15] Meurer A, Smith CP, Paprocki M, Čertík O, Kirpichev SB, Rocklin M, Kumar A, Ivanov S, Moore JK, Singh S, Rathnayake T, Vig S, Granger BE, Muller RP, Bonazzi F, Gupta H, Vats S, Johansson F, Pedregosa F, Curry MJ, Terrel AR, Roučka Š, Saboo A, Fernando I, Kulal S, Cimrman R, Scopatz A. (2017). SymPy: symbolic computing in Python: PeerJ Computer Science.
- [16] Alnaes, M.S., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, C., Richardson, J., Ring, J., Rognes, M. E., Well, G.N. (2015): *The FEniCS Project Version 1.5*: Archive of Numerical Software. vol. 3.

Presented at the NAFEMS World Congress 2017

Emerging Flood Model Validation Frameworks for Street-Level Inundation Modeling with StormSense

Jon Derek Loftis College of William and Mary Gloucester Point, VA, USA jdloftis@vims.edu

Harry Wang College of William and Mary Gloucester Point, VA, USA wanghv@vims.edu

David Forrest College of William and Mary Gloucester Point, VA, USA drf@vims.edu

Sokwoo Rhee Gaithersburg, MD, USA sokwoo.rhee@nist.gov

NIST

ABSTRACT

Technological progress in flood monitoring and the proliferation of cost-efficient IoT-enabled water level sensors are enabling new streams of information for today's smart cities. StormSense is an inundation forecasting research initiative and an active participant in the GCTC seeking to enhance flood preparedness in the Hampton Roads region for flooding resulting from storm surge, rain, and tides and demonstrating replicability of the solution. Herein, we present street-level hydrodynamic modeling results at 5m resolution with conventional flood validation sources alongside new emergent techniques for validating model predictions during three prominent recent flooding events in Hampton Roads during Fall 2016: Hurricane Hermine, Tropical Storm Julia, and Hurricane Matthew. Emerging validation techniques include: (1) IoT-water level sensors, (2) crowd-sourced GPS maximum flood extent measurements, and (3) geospatial flooded area comparisons with drone-surveyed flood extents via ESRI's Drone2Map. Model uncertainty was validated against 5 newly-established tide gauges within the domain for an aggregate vertical root mean squared error of ± 8.19 cm between the sensor observations and model predictions. Also, geospatial uncertainty was assessed using mean horizontal distance difference as ± 4.97 m via 206 crowd-sourced GPS flood extents from the Sea Level Rise App.

CCS CONCEPTS

•Computer systems organization \rightarrow Embedded and cyberphysical systems; •Software and its engineering \rightarrow Design patterns;

KEYWORDS

Hurricane Matthew, Hydrodynamic Modeling, Internet of Things, Smart City, Global City Teams Challenge, Replicability, Citizen Science, Sea Level Rise, Drone2Map

SCOPE 2017, Pittsburgh, PA USA

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM. 978-1-4503-4989-5/17/04...\$15.00

DOI: http://dx.doi.org/10.1145/3063386.3063764

Cuong Nguyen NIST Gaithersburg, MD, USA cuong.nguyen@nist.gov

ACM Reference format:

Jon Derek Loftis, Harry Wang, David Forrest, Sokwoo Rhee, and Cuong Nguyen. 2017. Emerging Flood Model Validation Frameworks for Street-Level Inundation Modeling with StormSense. In Proceedings of The 2nd Workshop on Science of Smart City Operations and Platforms Engineering, Pittsburgh, PA USA, 21 April 2017 (SCOPE 2017), 6 pages. DOI: http://dx.doi.org/10.1145/3063386.3063764

1 INTRODUCTION

Cities are inherently complex systems subject to innumerable nonlinear influences on how to efficiently allocate their limited resources [17]. This is certainly true for handling emergency flooding conditions in the near-present and how best to prepare for the imminent flood-related disasters of the future. Analysis of the local sea level trend from one water level monitoring sensor at Sewells Point in the City of Norfolk depict a long-term increase in mean sea level of 4.59 mm/year ±0.23 mm/year since its establishment in 1928, projecting that rising sea levels will inevitably exacerbate flooding conditions from storm events in the future [3, 16].Cities, counties, and town governments, local institutions, and private contractors, provide many solutions, each of which must be evaluated in its own way. However, provision of these serviceable flooding solutions often impacts the availability of other services. Many existing smart cities solutions, such as those implemented in the Global City Teams Challenge (GCTC) action clusters, are designed to have a measurable impact on specific key performance indicators. Because many of today's smart city/community development efforts are isolated and customized projects, the National Institute of Standards and Technology (NIST) has launched the GCTC to encourage collaboration and the development of standards. The GCTC's long-term goal is to demonstrate a scalable and replicable model for incubating and deploying interoperable, adaptable, and configurable Internet of Things (IoT)/Cyber-Physical Systems technologies in smart cities/communities. This program aims to help communities benefit from working with others to improve efficiency and lower costs. NIST created the Replicable Smart City Technology (RSCT) cooperative agreement program to provide funding to enable awardee City/Community Partners to play a lead role in the team-based GCTC effort to pursue measurement science for replicable solutions [1]. The RSCT program was designed to support standards-based platform approaches to smart cities technologies that can provide measurable performance metrics.

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of the United States government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

SCOPE 2017, 21 April 2017, Pittsburgh, PA USA



Figure 1: Map of the Hampton Roads Region of Virginia with StormSense partner agreement cities outlined in gray superposed with radar-derived rainfall totals (in.) over 72 hours from Sept. 19-22, 2016, during Tropical Storm Julia. The region has an area of 1,365 sq. km. (527 sq. mi.) and can be modeled using lidar elevations and high-res. bathymetry at 5 m resolution.

The StormSense project brings together partnering municipal governments in Hampton Roads, Virginia, including: Newport News, the RSCT grant recipient, Norfolk, Virginia Beach, Hampton, Chesapeake, Portsmouth, Williamsburg, and York County along with the Virginia Institute of Marine Science (VIMS), with emphasis on replicating a flood forecasting and monitoring solution across the entire region (Figure 1). As an example, in one neighborhood in the City of Newport News that is subject to frequent flooding, typically a large number of emergency responders were required to assist in evacuating the complex [2, 10]. However, by remotely alerting residents that the water is rising quickly on the local stream, the past two flooding events have not required any emergency responders to assist them in evacuating, who were subsequently able to dedicate their emergency services elsewhere [18]. The goal of establishing a flood monitoring network can be cost-prohibitive, but in the long term, the anticipated benefits of improved quality of life for a region's denizens are sizable. The goal is to replicate this level of success throughout the cities of Hampton Roads by providing a greater density of water level sensors. As an added benefit, residents are taking responsibility for their assumed risk of living adjacent to floodplains, resulting in a marked spike in the number of residents who have opted for flood insurance, with 2,231 claims totaling \$25M in damage attributed to Hurricane Matthew [9]. Many of these properties are insured through the Federal Emergency Management Agency's (FEMA) National Flood Insurance Program (NFIP), but many properties outside of the surveyed floodplain do not have preferred risk policies.

A stakeholder workshop conducted on January 19, 2016 with representatives from Hampton Roads regional emergency management, storm water engineering and planning municipal staff, as well as academic and non-government organization partners uncovered a need for near-term, locally scaled, and 'realistic' scenarios to communicate risk [5]. Emergency managers are currently limited in their communications tools and know them to be inadequate [5, 6]. A better understanding of the decisions people are making

В REA OF MINIMAL FLOOD HAZAF

Figure 2: Prototype of Newport News Smart Cities Planning Dashboard with water level sensor installation sites relative to FEMA's flood zones and NFIP repetitive loss claims, and GIS flood-vulnerable properties identified via Hazus for a 100-year flood scenario in: A) Salters Creek, vulnerable to Coastal Flooding, and B) along Newmarket Creek, an inland regulatory floodway vulnerable to rainfall-induced flooding.

to adapt to flooding is needed. Differences are expected in both flood perception and behavior between urban and rural audiences. A pilot study conducted in 2015 examining information logistics for drivers on flooded roads in Norfolk found that decisions made about driving were strongly situational, based upon the importance, timing and location of the driving plans, but that a regional approach to communication was needed and lacking [6]. Time living in the area was an important factor in risk perception and that information comes from local knowledge, recognized sources of information, and sometimes a haphazard mix of both. Examining these issues in the context of flood communication and further elucidating the currently vague appropriate flood model parameters for accurate inundation prediction at 5 m scale in a broader context is needed, leading to the following flood research questions:

· How should bottom friction be appropriately parameterized for high-resolution street-level sub-grid inundation models?

I. Loftis et al.

StormSense

- · How should percolation/infiltration of rainwater through different density surfaces present in urban and rural environments be accurately accounted for in a high-resolution sub-grid model?
- · How should model results be deseminated to enhance flood preparedness, and what communication methods and messages influence flood risk decision-making and behaviors (including information-seeking and adaptive response)?

To attempt to address these questions, examples from a recent installment of water level sensors by the United States Geological Survey (USGS) and the City of Norfolk will be used as a proxy for a suite of ultrasonic water level sensors currently being installed in Hampton Roads via StormSense. The sensors will be fed into a geographic information system (GIS) flood risk dashboard (example depicted in Figure 2) to demonstrate the utility of a higher-density water level sensor network and how future flood forecasting efforts can be augmented for the betterment of citizen safety. The system can be used to study inundation extents and timing during flooding events in order to:

- (1) Build public awareness of inundation, and recurrent flooding through predictable geospatially locatable events,
- (2) Collect quality crowd-sourced inundation extent data to test and validate the numerical model via citizen science, and
- (3) Enhance both the scientific understanding of the physical systems involved and the public understanding of the science of inundation.

In pursuit of this, the StormSense model will be spatially validated during three 2016 flooding events: Hurricane Hermine, Tropical Storm Julia, and Hurricane Matthew, via the following emergent data sources in the subsequent sections: 1) IoT-water level sensors, 2) Crowd-sourced GPS maximum flood extent measurements collected using the crowd-sourced data-collection app, 'Sea Level Rise', and 3) Geospatial flooded area comparisons with drone-surveyed flood extents via Drone2Map software.

2 STUDY AREA AND MODEL INPUTS

Hampton Roads has been described as the second-largest population center at risk from sea level rise, with ;400,000 properties exposed to flood or storm surge inundation [19]. The region has a population of over 1.7 million people, living and traveling on roads exposed to both severe and increasing frequent chronic "nuisance" flooding [8]. Existing flood communication and messaging systems have not yet responded to the changing risk patterns brought by sea level rise and have not been able to meet the needs of diverse at-risk communications audiences. A better understanding of flood risk perception, information seeking behavior and decision-making can inform the development of new communications tools and flood risk messaging [20]. This is the perceived intersect between new IoT-technologies and emerging flood model validation methods. For each storm event, water levels driven via 36-hour tidewatch forecasts provided by VIMS at Sewells Point were used to drive surge and tides, alongside wind and pressure inputs used to drive the model atmospherically, similar to [14].

2.1 Groundwater Inputs

Recent advancements in hydrodynamic computation have enabled models to predict the mass and movement of flood waters to predict water velocities at increasingly finer scales. However, the current version of the sub-grid inundation model VIMS has developed does not fully incorporate a comprehensive groundwater model that slowly returns flood waters that infiltrate through the soil back to the nearest river. This is a valuable aspect of flooding relevant for city planning perspectives using sub-grid hydrodynamic modeling that has been successfully developed and employed throughout the Netherlands, Germany, and Italy [7]. There is an array of groundwater wells that exist in the Hampton Roads Region, bored and monitored by the USGS [4]. These temporally-varying values for hydraulic conductivity could provide some valuable input information for the hydrodynamic model via Richard's equation [13]. However, this does not currently account for the standard practice of near-surface groundwater displacement via pumping prior to anticipated flooding events conducted by cities with residents in the floodplains where a high water table regularly exacerbates even minor rainfall events [15]. Nevertheless, values observed near these sites prior to forecast simulations were used as the model's initial condition to estimate infiltration through previous surfaces, to counterbalance precipitation inputs, similar to [13].

In forecast approaches, groundwater influence is usually neglected, since typically storm surge is a short-term event, and groundwater recharge is more of a delayed and long-term process, however, it is becoming increasingly important to also consider in forecasting longer-term extratropical flooding events such as nor'easters where flooding and high winds can persist for 5 or more tidal cycles. Although, VIMS have been incorporating different forms of percolation of flood waters through different types of ground cover ranging from vegetated to impervious within the sub-grid model more recently [12, 13] there are still some potential applications of storm water that could be manually added to the existing sub-grid model version to account for surge flooding backups through storm water drainage without sufficient backflow prevention [13].

Precipitation Inputs 2.2

The inundation model could be used to guide decisions related to storm water management by using existing sensor-derived precipitation data in several cities. This could be expanded to include data observations from rain gauges that are currently operating on sewer and storm water pump stations in the localities, and from the Hampton Roads Sanitation District (HRSD), which combined currently amounts to ~ 130 sensors. With an iteratively interpolated series of precipitation measurements, further research could also be conducted with these sensors and the 10 proposed water level sensors to model localized microburst precipitation events. As evidenced in the series of interpolated rainfall images in Figure 1, a combination of these rainfall data with new observations from a high-density water level sensor network, would help better explain why one neighborhood experienced a drizzle and a neighboring community experienced a deluge during Hurricane Matthew. Furthermore, this could aid researchers to help model ways that

the city's systems could potentially be augmented for greater resilience to precipitation -induced flooding threats in the future. In the simulations presented herein, model results are calculated with temporally-varying precipitation inputs from the currently-private rain gauge data from HRSD.

WATER LEVEL SENSOR COMPARISON 3

StormSense is currently in the process of identifying and surveying sites for new IoT-bridge-mounted ultrasonic and microwave radar water level sensors in Newport News, Virginia Beach, and Norfolk, as outlined on the StormSense project's website at stormsense.com. These sensors will complement the previously installed array of 2 gauges operated by NOAA, 9 gauges recently installed in 2015-2016 via Hurricane Sandy relief funds operated by the USGS, and 1 gauge operated by VIMS in Hampton Roads. While these remote sensors are largely X-band radar sensors transmitting data through satellite signals, the IoT-sensors will enlist ultrasonic sensors and transmit data via cellular transmission protocols with the focus of creating a replicable cost-effective network of sensors. Some perceived utilities of water level sensors are viewed as follows:

- (1) Archiving of water level observations for flood reporting
- (2) Automated targeted advance flood alert messaging
- (3) Validation/inputs for hydrodynamic flood models

A collaboration between VIMS and the constituent partner Cities of: Newport News, Hampton, Norfolk, Virginia Beach, Portsmouth, Chesapeake, Williamsburg, and York County, in Hampton Roads, VA, will provide a prototype for strengthening emergency response times by providing spatial flood extent predictions in interactive map form at 5 m resolution. The plan for integrating the inundation model into a more permanent warning system involves planned connection with the new sensors to the citiesfi current Everbridge notification systems for alert messaging when the sensor observes flooding at user-specified elevations, and integration with model predictions for timely forecasted alerts once the sensors are tidallycalibrated. This approach demonstrates the benefits of replicating shared smart city solutions across multiple cities and communities that are facing similar flood challenges and it aligns with the goals of GCTC and RSCT programs. It is the hope that the recent installation of water level sensors provided by the efforts of the USGS can be used as an opportunity to demonstrate some of the benefits of added water level sensors while the alternative model validation approaches described herein may be replaced with more reputable and replicable monitoring methods soon.

A comparison of five existing water level sensors were used to temporally and vertically validate the hydrodynamic model's predictions. These sensors are located at: 1) National Oceanic and Atmospheric Administration (NOAA)'s Money Point sensor, the USGS's permanent sensors at 2) Rt. 17 near Portsmouth, and 3) Bailey Creek at Dock Landing Rd. in Chesapeake, a temporary USGS rapid-deployment gauge at 4) the Hague in Norfolk deployed only during Hurricane Hermine, and 5) water level sensor at Mayflower Crescent pump station maintained by the City of Norfolk. These 5 gauges resulted in an aggregate vertical root mean squared error (RMSE) of ±8.19 cm over the 36-hour Hurricane Hermine forecast simulation [14]. The four gauges present during Hurricane Matthew yielded a more favorable aggregate RMSE of ±4.69 cm. Tropical Storm Julia resulted in 12-14" of rainfall a week prior to Hurricane Matthew, ameliorating high water table drainage issues during the storm (Figure 1). Both storms produced minimal surge related coastal flooding and inundation impacts were far more profound inland, making coastal and estuarine water level sensors less practical for verification of inland inundation extents or depths.

4 **CROWDSOURCED GPS FLOOD EXTENTS**

Hurricane Hermine had a more significant storm surge measured by water level sensors in Hampton Roads and less rain, while the opposite was true for Hurricane Matthew. The relatively new citizen science 'Sea Level Rise' mobile app provided 206 points of geospatial data for use with validating predicted flood extents in Norfolk during Hurricane Hermine with a favorable Mean Horizontal Distance Difference (MHDD) of ±4.97 m (Figure 3A). Sites labelled B-D in Figure 3 represent the modeled maximum flooding extents calculated by the street-level hydrodynamic model in the flood-prone Larchmont neighborhood of Norfolk. Positioned on a peninsula bounded by the Elizabeth River to the west and the Lafayette River to the north and east, the area is no stranger to tidal 'nuisance' flooding. By measuring the horizontal distances from the GPS-reported points of maximum flooding extents from the 'Sea Level Rise App', to the edge of the model predicted maximum flooding extent contour line, an assessment of geospatial accuracy may be reached with minimal processing effort using the standard distance formula [17,18]. In Figure 3B at the houses along Richmond Crescent, the MHDD between the 65 GPS observation points and the model-predicted maximum flood extent contour line is ±6.72 m. Figure 2C depicts a slightly more favorable comparison with a MHDD of ±3.92 m along Cambridge and Carroll Crescent's 74 points. Finally, in Figure 3D, 48th Street near ODU and Hampton Blvd's 67 GPS observations were in agreement with model-predicted flooding extents with a MHDD of ± 4.48 m.

An apparent caveat of this geospatial MHDD approach is that it is only a relevant metric in areas with minimal surficial slope [17,18], like those that characterize Hampton Roads, VA. In areas with steeper slopes immediately adjacent to the shoreline, model over-prediction of several inches or even feet in the vertical may only manifest in minuscule increments of change on the horizontal scale. However, lack of these app data or crowdsourced data on ArcGIS online in the region during tropical Storm Julia or Hurricane Matthew (due to mass power outages) led to the use of emerging data validation methods from image analysis of drone videos via Drone2Map.

5 **DRONE2MAP FLOODED AREA** COMPARISONS

Useful information can be extracted from iterative image analysis of publicly uploaded drone footage to internet video repositories (e.g. YouTube, Vimeo, etc.) released under the umbrella of the shared creative commons license for research. Often, the rapid battery drain constraints of flight control and video streaming result in short drone flight times of ¡30 minutes. Thus, accompanying auxiliary data collection alongside flight-control information is often an afterthought. Through estimation of altitudes and retroactive construction of pre-programmed flight plans, a video may be parsed

Loftis, Jon 'Derek'; Nguyen, Cuong; Rhee, Sokwoo. "Emerging Flood Model Validation Frameworks for Street-Level Inundation Modeling with StormSense." Paper presented at Second International Workshop on Science of Smart City Operations and Platforms Engineering (SCOPE), Pittsburg, PA, United States. April 21, 2017 - April 21, 2017.

StormSense



Figure 3: Maximum flooding extents predicted for 2016 Hurricane Hermine on Sept. 2nd validated via crowdsourced GPS monitoring of maximum flooding extents on Sept 3rd in the Larchmont neighborhood of Norfolk, VA.



Figure 4: A) Single flooding image frame from drone video over drone video footage of Llewellyn Ave near Haven Creek Boat Ramp in Norfolk at 2:30pm on Oct. 10, 2016 after Hurricane Matthew. B) Laplace transform of pixel values for edge detection of water's edge. C) Resulting polygon (white) of water's edge from Drone2Map stitched edge-detected imagery representing extracted flood extents [14]. Source Video: https://youtu.be/PkvjnqDlTcQ

into a series of image frames that may be individually prescribed estimated GPS metadata to aid in flood model validation in places where there is an absence of validated reputable water level sensor measurements. Figure 4 depicts an estimated Drone2Map flooding extent captured via drone at 2:30pm on October 10th, 2016 on

Llewellyn Ave of a mostly oblique segment of video. Figure 4 depicts drone video footage at the intersection of Monticello Ave and Jefferson Ave in Norfolk from an overhead aerial view shows a more uniform depiction of flooding, although there are complications in each series of images form uniform edge detection methods. The oblique view suffers from line of sight complications with taller buildings, while the overhead viewpoint can be hampered by shadows. This can have potentially worse implications for detecting the dark transition between the water's edge and shadows cast by tall infrastructure becoming more apparent later in the day as the shadows elongate. When calculating the mean horizontal distance difference of the resulting edge-detected interface between wet and dry pixel values, the example is Figure 4 at the Haven Creek Boat Ramp in Norfolk at 3:00pm on Oct. 10, 2016, resulted in an AHDD of ±14.39 m from 137 points artificially constructed at 5 m regular intervals along the interface line.

6 **CONCLUSIONS**

The hydrodynamic model in Hampton Roads, VA, was effectively validated using 5 water level sensors within the model domain during Hurricane Hermine to yield a vertical RMSE of 8.19cm, as a primary time-honored model validation method that has been embraced by the hydrodynamic modeling community as a staple for determining the uncertainty of their predictions. Typically, the USGS provides a valuable service in the form of surveying high water marks after major flood events, but as none of these events were truly catastrophic flood events in Hampton Roads, VA, relative to the southern U.S. Eastern Seaboard, high water lines in the form of GPS maximum flood extent points from the citizen science App, 'Sea Level Rise' were used instead as a secondary form of model validation. Results from 3 sites in Norfolk yielded a MHDD of ±4.97 m during Hurricane Hermine. A tertiary flood validation approach involving the use of the newly-released Drone2Map software from ESRI was employed to successfully develop maximum

SCOPE 2017, 21 April 2017, Pittsburgh, PA USA

flooding extent polygons from aerial drone survey footage. The footage was parsed into frames at quarter-second intervals, and batch-processed through image analysis to run an edge detection algorithm using a Laplacian filter. This highlighted areas of the images where stark contrasts in wet/dry (dark/light) pixel values were used to highlight flooding extents at the time of the flight. The frames were georeferenced and mosaicked together in the software, and then compared with the analogous time-aware raster model output layer for a reasonably agreeable estimated AHDD of 14.39 m (±3 of 5 m resolution sub-grid cells). It is worth noting that in cases of heavy rainfall, this street-level sub-grid hydrodynamic modeling also performs the function of a hydrologic transport model to predict flow accumulation to aid in identification of areas that are most susceptible to flooding. This is useful for resilient building practices, as the model could also identify potential areas where development of green infrastructure could commence, with the understanding that a sub-grid model represents infrastructural features and many city lifelines better than most conventional hydrodynamic models [11]. In the future, smart city systems could evaluate the efficacy of candidate blueprint solutions to flood-related problems, and suggest how they could be addressed with a street-level inundation model (bold):

- Reduction of impervious surfaces is addressed by changes to spatially-varying soil infiltration values
- Land use changes is addressed by the model grid mesh modification to remove/add buildings/ infrastructure AND changes to spatially-varying soil infiltration values
- · Combination gray and green infrastructure opportunities are tested by changes to spatially-varying soil infiltration values in areas where modified green infrastructure lie
- · Increase in storm water "holding" management systems is modeled by Digital Elevation Model modification and adding sources/sinks for new holding reservoirs/ ponds

ACKNOWLEDGMENTS

The authors would like to thank the USGS and the City of Norfolk for their efforts in establishing and maintaining the recent suite of water level sensors in Hampton Roads used to help validate the results presented herein. Our thanks is extended to the Hampton Roads Sanitation District, and the public utilities, public works, and water works divisions in the Cities of Newport News, Norfolk, and Virginia Beach for making their privately-collected rainfall measurements at pump stations throughout Hampton Roads accessible for this research endeavor. Portions of this publication and research efforts are made possible through the encouragement and support of NIST via federal award #70NANB16H277.

Official contribution of the National Institute of Standards and Technology; not subject to copyright in the United States. Certain commercial products are identified in order to adequately specify the procedure; this does not imply endorsement or recommendation by NIST, nor does it imply that such products are necessarily the best available for the purpose.

REFERENCES

- [1] Federal Funding Opportuninty Announcement: Replicable Smart City Technologies Cooperative Agreement Program. https://www.nist.gov/document-360. (????). Accessed: 2017-03-02.
- Letter regarding flooding on Newmarket Creek and City Line Apartments from City of Newport News Fire Chief. https://wm1693.box.com/s/ qwsfc722peihdb70xsvfhzwxas348y0u. (????). Accessed: 2017-03-02.
- NOAA Tides and Currents: Sewells Point Sea level Trends https:// tidesandcurrents.noaa.gov/sltrends/sltrends_station.shtml?stnid=8638610. (????). Accessed: 2017-03-02.
- [4] USGS Groundwater Monitoring Sites: Tidewater Virginia. https:// groundwaterwatch.usgs.gov/StateMap.asp?sa=VA&sc=51. (????). Accessed: 2017-03-02
- 2016. Hampton Roads Planning District Commission Flooding Mitigation Stake-[5] holder Workshop. (Jan. 2016).
- [6] Howard Botts, Thomas Jeffrey, Wei Du, and Logan Suhr. 2015. Storm Surge Report. Technical Report. CoreLogic. http://www.corelogic.com/research/storm-surge/ corelogic-2015-storm-surge-report.pdf
- V. Casulli. 2015. A conservative semi-implicit method for coupled surfacesubsurface flows in regional scale. International Journal for Numerical Methods in Fluids 79, 4 (2015), 199-214.
- T. Ezer and L. P. Atkinson. 2014. Accelerated flooding along the US East Coast: on the impact of sea-level rise, tides, storms, the Gulf Stream, and the North Atlantic oscillations. Earth's Future 2, 8 (2014), 362-382.
- [9] FEMA. 2016. National Flood Insurance Program coverage the same as homeowner insurance or FEMA assistance. isn't http://www.vaemergency.gov/national-flood-insurance-program-(coverage-isnt-homeowner-insurance-fema-assistance/. (Dec. 2016)
- [10] J. Lawlor. 2012. City Line Apartments: flood prone and solutions in sight. Daily Press (2012). http://articles dailypress.com/2012-08-29/news/dp-nws-city-line-apts-20120829_1_ city-line-apartments-apartment-investment-and-management-government-ditch
- [11] J. D. Loftis. 2014. Development of a Large-Scale Storm Surge and High-Resolution Sub-Grid Inundation Model for Coastal Flooding Applications: A Case Study during Hurricane Sandy. Ph.D. Dissertation. College of William & Mary.
- J. D. Loftis, H. Wang, and R. DeYoung. 2013. Storm Surge and Inundation Modeling [12] in the Back River Watershed for NASA Langley Research Center. Technical Report NASA/TM-2013-218046, NASA,
- [13] J. D. Loftis, H. V. Wang, R. J. DeYoung, and W. B. Ball, 2016. Using Lidar Elevation Data to Develop a Topobathymetric Digital Elevation Model for Sub-Grid Inundation Modeling at Langley Research Center. Journal of Coastal Research Special Issue, 76 (2016), 134-148
- J. D. Loftis, H. V. Wang, and D. R. Forrest. 2016. Street-Level Inundation Modeling of Hurricanes Matthew and Hermine and Emerging Flood Monitoring Methods in Hampton Roads. Technical Report. VIMS Department of Physical Sciences. http://doi.org/10.21220/V5XW2G
- [15] J. D. Loftis, H. V. Wang, S. E. Hamilton, and D. R. Forrest. 2017. Combination of Lidar Elevations, Bathymetric Data, and Urban Infrastructure in a Sub-Grid Model for Predicting Inundation in New York City during Hurricane Sandy. Computers, Environment, and Urban Systems (2017). In Review
- [16] Molly Mitchell, Carl Hershner, Julie Herman, Dan Schatt, Pam Mason, and Emily Eggington. 2013. Recurrent Flooding Study for Tidewater Virginia. Technical Report. Virginia Institute of Marine Science. http://ccrm.vims.edu/recurrent_ flooding/Recurrent_Flooding_Study_web.pdf
- Sokwoo Rhee. 2016. Catalyzing the Internet of Things and Smart Cities: Global [17] City Teams Challenge. In 1st International Workshop on Science of Smart City Operations and Platforms Engineering (SCOPE). Global City Teams Challenge GCTC), 1
- [18] [7] H. Smith. 2016. After yet another City Line Apartments flood, FEMA steps in to help. Daily Press (2016). http://www.dailypress.com/news/newport-news/ lp-nws-fema-city-line-20161112-story.html
- [19] W. V. Sweet, J. Park, J. J. Marra, C. Zervas, and S. Gill. 2014. Sea level rise and nuisance flood frequency changes around the United States. Technical Report. NOAA Technical Report NOS COOPS 73.
- [20] T. Wahl, S. Jain, J. Bender, S. D. Meyers, and M. E. Luther. 2015. Increasing risk of compound flooding from storm surge and rainfall for major US cities. Nat. Clim. Change (2015). DOI: http://dx.doi.org/10.1038/nclimate2736

Proceedings of the ASME 2017 International Design Engineering Technical Conferences and **Computers and Information in Engineering Conference IDETC/CIE 2017** August 6-9, 2017, Cleveland, Ohio, USA

DETC2017-68054

DESIGN PATTERNS FOR VISUALIZATION-BASED TOOLS IN SUSTAINABLE PRODUCT DESIGN

Devarajan Ramanujan*

Department of Mechanical Engineering Massachusetts Institute of Technology Cambridge, MA

William Z. Bernstein Systems Integration Division National Institute of Standards and Technology Gaithersburg, MD

Karthik Ramani

School of Mechanical Engineering **Purdue University** West Lafayette, IN

ABSTRACT

Most design activities involve exploring and comparing existing designs. Thus, adopting an eco-conscious approach in the design exploration process can aid environmentally sustainable product design (SPD). One approach for supporting exploration in SPD is through tools based on information visualization (InfoVis). The use of InfoVis for SPD allows data-driven exploration of solutions that is rapid, direct, and supports investigation of questions that the designer may not have identified. Previous work has demonstrated the utility of InfoVis tools for different facets of the lifecycle, e.g. redesign, supply chain exploration, and life cycle assessment. These tools focus on projecting sustainability-related implications back to design. However, to fully realize their potential, future tools must synthesize data in a manner that helps designers view the effects of a design change on all downstream stages. Such tools will have to work across multiple data types, visual representations, and stakeholders. In this paper, we take the first steps towards addressing this challenge by formulating design patterns for visualization and interaction of product lifecycle data. These design patterns were synthesized by reviewing previous works that have successfully created visualization-based tools for SPD. The suggested design patterns can, (1) serve as a guide for creating integrated visualization-based tools for SPD, and (2) help create reusable visual components that aid in quick interface wireframing.

INTRODUCTION 1

Reducing the environmental (env.) impacts of products has become an important focus for industries [1]. Among the opportunities available for reducing the env. impact of a product, usually the design stage offers the most potential [2]. Integrating env. aspects of a product with its design creates the need for searching env. information, performing env. assessments, and outlining a suitable strategy [3]. In this paper, we focus on the search part of this process and look at visualization-based tools for exploration. The use of visualization-based tools in sustainable product design (SPD) is designer-driven and often involves exploring previous designs to

- compare env. impacts of design alternatives,
- support discovery of more benign design alternatives, and
- aid designers' understanding of correlations between design attributes and env. impacts for a single or a set of designs.

These activities are critical for supporting SPD, as quantitatively predicting a product's lifecycle env. impact based on its design

This work was authored in part by a U.S. Government employee in the scope of his/her employment. ASME disclaims all interest in the U.S. Government's contribution.

^{*}Address all correspondence devr@mit.edu

Bernstein, William; Ramani, Karthik; Ramanujan, Deverajan. "Design Patterns for Visualization-Based Tools in Sustainable Product Design." Paper presented at ASME 2017 International Design Engineering Technical Conference & Computers and information in Engineering Conference. IDETC/CIE 2017, Cleveland, OH, United States. August 6, 2017 - August 9, 2017.


FIGURE 1: ROLE OF INTEGRATED VISUALIZATION-BASED TOOLS FOR SPD. THE GOAL IS TO ALLOW DESIGNERS TO PER-FORM DATA-DRIVEN ANALYSES OF IMPLICATIONS IN DOWN-STREAM STAGES RESULTING FROM DESIGN CHANGES. PROD-UCT LIFECYCLE DATA FROM DOWNSTREAM STAGES AND RE-SULTS FROM ENVIRONMENTAL ASSESSMENT TOOLS (E.G. LCA) ARE PROJECTED BACK TO DESIGN THROUGH THESE TOOLS.

attributes is a prohibitively challenging task [4]. Our literature review revealed that current visualization-based tools are often developed to inform designers about the sustainability-related implications of specific facets of the lifecycle. While such tools project valuable information back to design, the effective use of such tools requires researchers to explore methods for holistic integration of data from all downstream stages [2]. To make use of lifecycle data effectively, future tools must synthesize data in a manner that helps designers view the effect of making a design change on all downstream stages. Such tools will work have to work across multiple data types, visual representations, and stakeholders. Figure 1 shows this integration layer that can allow designers to perform data-driven analyses of design changes in downstream stages. Lifecycle data from downstream stages and results from env. assessment tools, such as life cycle assessment (LCA), are projected back to design through the use of integrated visualization-based tools. These tools should support human sensemaking using both (1) data-driven methods that gather, process, and summarize data, and (2) user-driven methods that allow users to input their domain knowledge or datadriven insights during exploration. The need for a human-inthe-loop is critical as SPD involves tasks wherein, although the goal is known, designers rarely know the best approach for the problem, what questions to ask, and which among them are the right questions to consider [5]. Furthermore, ambiguities present in design representations, lack of information from downstream stages, and uncertainties in env. assessment all increase complexity in the SPD process [2].

In this paper, we take the first steps towards creating integrated visualization-based tools for SPD by formulating design patterns for visualization and interaction with regards to lifecycle data. For this, we adopt an inductive line of research [6] and derive these patterns by reviewing previous research that has successfully created visualization-based tools for SPD. Synthesizing a list of design patterns for visualization-focused tools in SPD is essential for reducing the barrier to their creation for both academic researchers and industry practitioners. Design patterns provide a tested heuristic for quick prototyping, reducing the cost of creating future tools. This is particularly significant in SPD as the growing complexity of production, consumption, and environmental systems necessitates flexible, usable, knowledge management tools that can integrate information across the lifecycle.

2 MOTIVATION FOR USING INFOVIS TOOLS IN SPD

Considering sustainability adds parameters and constraints to the design process, which increases complexity [7]. In SPD, the challenge of dealing with multiple inter-related parameters is compounded by the fact that methods and computer support tools for SPD are disconnected from those focused on design exploration. One primary reason for this disconnect is the mismatch in data representations used in these two contexts [8]. To illustrate, conducting an LCA for a part does not require knowledge about its form (shape), fit (tolerances), or function. However, these attributes are vital for assessing design intent. Reducing such gaps is essential for easing the disconnect between tools for design exploration and SPD.

Fully and semi-automated approaches based on techniques, such as neural networks and expert systems have been previously applied to SPD [9, 10, 11]. However, it is challenging to extend them towards exploration-focused tasks primarily due to its ill-defined nature. Previous research (see Section 4) has shown that InfoVis-the use of computer-supported, interactive, visual representations of abstract data to amplify cognition [12]—offers strong potential to support exploration tasks in SPD. Visualization-based tools combine the powerful pattern detection properties of the human visual system with the large data processing and manipulation capabilities of a computer system [13]. This allows visualization-based tools to support designers' insight generation processes and leverage their expertise in qualitative decision-making. An important aspect of InfoVis is keeping the human in the decision loop. This makes InfoVis-based techniques relevant towards tasks in SPD, which are largely open-ended. Another advantage of using InfoVis tools in SPD is the ability to create a common representation between domains (in our case environmental assessment and design exploration) by transforming data into graphical primitives [14].

NOMENCLATURE 3

Here, we define a *designer* as a person who is involved with the generation and development of ideas that leads to a new product [15]. On the other hand, we use the words, *researchers* and *practitioners*, to denote people in academia and industry, respectively, who are interested in creating new visualizationbased tools for designers. Design patterns refer to reusable implementation or solution strategies that can be customized to address challenges in a specific context. Originally published for object-oriented programming [16], design patterns have been successfully used in codifying solution strategies and best prac-

Paper presented at ASME 2017

tices in the context of information visualizations [17,18]. We use Shneiderman's task-by-type taxonomy [19] to categorize previous tools, and denote InfoVis tasks in this taxonomy by T#. The symbol *P*# is used for denoting the synthesized design patterns.

4 **RELATED WORK**

Creating InfoVis-based tools for SPD requires an understanding of visual representations for both design and sustainability-related data. These representations should be situated in an overall framework that allows designers to explore the design space. Most previous works on supporting design exploration in the context of env. sustainability usually focus on developing env. indices useful for the exploration process. There seems to be a research gap on methods and tools that support human-centered exploration for SPD. To understand barriers in this context, we review previous works that (1) develop computer-aided methods and tools for enabling SPD, (2) apply InfoVis techniques for aiding design exploration, and (3) promote env. sustainability through InfoVis.

4.1 Methods and tools for sustainable product design

Most design activities involve the reuse of existing design information to create new designs. Thus, supporting exploration of more benign alternatives in a design space (or repository) presents a significant opportunity for promoting SPD. Previous research in this area has looked at methods and tools for estimating env. performance of existing designs using neural networks [9, 10], simplified life cycle indices [20, 21], and functionbased relationships [22, 23]. A majority of these papers focus on estimating env. indicators relevant to the context of application. Even though such methods have been successful in facilitating reuse of design information, most of these methods cannot be extended towards interactive exploration of design spaces. There is a significant lack of research on understanding differences in representations between design data and sustainabilityrelated data from the standpoint of the designer. The disconnect in data representations translates to a reduction in the effectiveness of env. assessment tools for design. In their research project with industrial designers, Lofthouse [24] identified that eco-design tools, such as the LIDS wheel and the EcoReDesign programme, were incongruous to the design process. Similarly, Rio et al. [8] highlight challenges related to data interoperability between env. assessment and product design. In response to these challenges, researchers have proposed expert systems [11] that provide sustainability-related recommendations in early-phase design. Other approaches (see Section 4.3) propose the use of InfoVis as a means to commonize representations in design and sustainability assessment.

4.2 Visualization-based tools for design exploration

Within engineering design, visualization finds application in both scientific visualization (SciVis) and InfoVis. Card et al. [12] differentiate the two methods in that SciVis is typically applied to scientific/physically based data (e.g. engineering stress and fluid velocity) while InfoVis is directed towards abstract, nonphysically based data (e.g. parameter spaces and product/supply chain structures). Design spaces usually contain non-physical data without a prescriptive mapping to visual representations. This makes the study of InfoVis for design exploration a research discipline in its own right. The use of InfoVis in design exploration is predominantly driven by the need to (1) characterize and navigate multi-dimensional design spaces [25], (2) understand parameter trade-offs for design optimization [26, 27], and (3) generate insights, patterns, and trends for design decisionmaking [28, 29]. Most InfoVis applications in design exploration relate to situations wherein the designer is exploring multidimensional datasets with limited knowledge about the relationships in the data. Unlike traditional data exploration, visual exploration using InfoVis requires users to interact with abstract data representations. A range of standard and custom visualization methods have been utilized in previous work to support exploration processes. Interested readers are directed to works by Chi [30] and Keim [31] that present taxonomies for commonly used data visualization techniques. InfoVis-based tools in SPD often rely on such visualization techniques to relate metrics from env. assessment to design parameters.

4.3 Visualization-based tools for sustainability

Previous efforts integrating visualization and env. sustainability has been directed towards, (1) eco-visualization: the use of visualization to promote sustainable behavior in endusers, and (2) visualization for aiding sustainable product design: the use of InfoVis to support designers practicing SPD. Eco-visualizations have been explored to provide eco-feedback on energy consumption, carbon loads, water usage, and waste recycling [32, 33]. They have also been used for, promoting conservation-related dialogs among community members in the context of paper printing [34], and for developing persuasive media that promote eco-conscious behavior [35]. Along similar lines, Lilley [36] present a list of attributes for behavior changing devices. These attributes can be adapted for developing ecovisualizations that promote sustainable behaviors in end-users.

Creating information visualizations for the designer presents a different set of challenges when compared to designing ecovisualizations for the end-user. The primary reason is that information visualizations for designers focus on supporting insight generation through interactive exploration of design variables and performance/output parameters. On the other hand, visualization for end-users often tend to focus on providing feedback only on performance/output parameters (i.e. energy efficiency and transportation impacts). Contextualizing visualizations to the design process adds additional parameters (and dimensions) making insight generation quite challenging. Studying mechanical designers, Bernstein et al. [7] discovered that vi-

Paper presented at ASME 2017

sual representations that interface LCA with design are necessary for promoting SPD. By translating design and env. sustainability variables into graphical primitives, visualization-based methods enable concurrent presentation of data and support designers' insight into the generation processes.

Espinosa et al. [37] present a technique for generating data visualizations useful for env. LCA. The authors develop VisEIO-LCA, a tool that can visualize product-related env. data. Otto et al. [38] use a glyph-based information representation and visualization approach to represent lifecycle information. The authors generate multiple visualizations for analyzing LCA results of a desktop computer. Ramanujan et al. [39] propose a framework, and a multi-dimensional visualization tool that allows exploration of 3D model repositories. The tool helps the designer explore part repositories driven by similarities in design attributes (i.e. material, manufacturing, and function) while simultaneously considering the env. implications of a chosen design. A mutually co-ordinated visualization tool for simultaneous exploring supply chain and design data in the context of eco-conscious redesign is described by Bernstein et al. [40].

Although there are several commercial tools that support SPD [41], only a handful of them use interactive data visualizations that allow users to explore the decision space. A notable exception is SourceMap, a tool for visualizing env. indicators for a product's supply chain [42]. SourceMap represents supply chain data as node-link diagrams that are overlayed on geographical maps. Apart from the works discussed above, other researchers [43, 44, 45, 46, 47, 48] have also looked at visualization techniques for product design and env. assessment.

5 METHODOLOGY

To identify visualization-based tools in SPD, we conducted a review of previous literature. We identified previous works through Google Scholar, Scopus, Science Direct, and the ASME Digital Collection using combinations of the following SPD and visualization related keywords: {*sustainable design, sustainable* product design} and {information visualization, visual analyt*ics*, *visual exploration*}. We filtered the resulting papers based on, (1) relevance to sustainable product design, (2) presence of a visualization-based tool in the paper. Additionally, if we found multiple papers by the authors that discuss the same concepts, we opted to include the most comprehensive version of those paper in the review. Table 1 lists papers that were selected from the search process. We included QuestVis [49] as the discussions herein are valuable for visualizing env. indicators.

We used Shneiderman's task-by-type taxonomy (TTT) [19] to infer design patterns from the tools that were reviewed. The TTT consists of 7 InfoVis task-types (overview, zoom, filter, details on demand, relate, history, & extract) for creating visual analysis tools. We used the TTT over other taxonomies present in InfoVis literature since (1) it forms one of the fundamental

IABLE 1: LIST O	F VISUALIZATION-BASED TOOLS IN SPD RE	Ξ-
VIEWED TO IDENTIJ	Y DESIGN PATTERNS. FOR EACH TOOL, REI	
EVANT TASKS-TYPE	S FROM THE TTT ARE ALSO SHOWN.	

Name	Jame Supported Activities		
shapeSIFT [39]	Exploration of 3D part repositories	overview, filter	
VisEIO-LCA [37]	EIO-LCA results visualization	zoom, details on demand	
OM-Glyphs [38]	LCA & LCIA results visualization	extract	
QuestVis [49]	Policy making & Indicator visualization	zoom, filter, history, extract	
ViSER [40]	Concurrent exploration of supply chain and product architecture	details on demand, relate, extract	
SourceMap [50]	Supply chain visualization and analytics	overview, zoom, details on demand	
Uchil & Chakrabarti [45]	Interfacing LCA results with product design	filter, relate	
i-Tree [46]	Transformation of impactful LCA flows to eco-potentials	zoom, filter, relate	

basis for more recent works on InfoVis taxonomies, and (2) a task-based classification allows us to extract design patterns that can map to high-level features that can be implemented in an interface. The latter is significant as low-level implementations used by researchers and practitioners are strongly influenced by the context for which their tools are built. Therefore, providing guidance for low-level implementations (such as the kind of visual representation or software to use) might be overly constraining. Table 1 shows the corresponding TTT principles that are embodied in the tools that we reviewed.

6 MAPPING THE TASK-BY-TYPE TAXONOMY TO VISUALIZATION-BASED TOOLS IN SPD

For each task type in the TTT, (1) we provide examples of their usage based on works in Table 1, and (2) infer design patterns for InfoVis-based tools in SPD relevant to this task-type.

6.1 Overview (T1): Obtain overview of design space

Overviews are intended to provide a macro-level perspective about decision spaces. Good information visualizations, according to Shneiderman, follow the mantra: Overview first, zoom and filter, then details on demand [19]. In the context of SPD, overviews can be used for (1) showing variations in env. indicators or design parameters over a collection of designs, (2) presenting a global perspective of a system across specific dimensions (i.e. geographical overlays, calendar charts), and (3) visu-



FIGURE 2: (A) THE SHAPESIFT PROTOTYPE [39] PROVIDES AN OVERVIEW OF PART ALTERNATIVES USING A SQUARIFIED LAYOUT VISUALIZATION. HERE, THE AREA OF EACH CELL IS SCALED INVERSELY TO THE ENVIRONMENTAL INDICATOR OF THE CORRESPONDING PART. (B) AN OVERVIEW OF A SUPPLY CHAIN FOR A PRODUCT IS SHOWN USING A GEOGRAPHICAL MAP IN SOURCEMAP [50].

alizing the overall structure of the design/decision space.

6.1.1 Illustrative examples in SPD

shapeSIFT: The shapeSIFT tool (Fig. 2(a)), uses a squarified layout for visualizing a set of similar shaped parts obtained from sketch-based retrieval. Here, the env. indicator (cumulative energy demand per unit volume) is mapped to cell size. Relevant design parameters like material, manufacturing, and function are encoded using secondary variables, e.g. color.

Sourcemap: An overview of the geographic location of each supply chain node, is presented on a map overlay in Sourcemap (see Fig. 2(b)). In the context of SPD, such overlays help in understanding env. indicators such as carbon footprint resulting from transporting goods/services across geographies. Sourcemap allows designers to easily identify hotspots for improvement and realize alternative supply chains that are both economically viable as well as more environmentally benign [42].

6.1.2 Suggested design patterns

Visual overviews are useful in providing a lay of the land so that designers can gauge properties for the entire or a significant portion of the design space. Based on the design of shapeSIFT and SourceMap, we infer the following design patterns for SPD.

- P1: Indicator overviews: Visual presentation of an overview for env. indicators allows designers to make quantitative and/or qualitative comparisons across design alternatives. Interactions such as sorting and filtering of overviews benefit designers in quickly characterizing the exploration space [51].
- P2: Eco-prominence: Env. indicators can be represented using the most prominent visual variable on that visualization. When possible, visual variables for env. indicators should allow pre-attentive processing of information [52].
- P3: Eco-persistence: Persistent visualizations of env. indicators help designers understand the env. performance of design



FIGURE 3: (A) MULTILEVEL INDICATOR BROWSING SUP-PORTED IN THE QUESTVIS INTERFACE [49]. THREE AGGRE-GATION LEVELS OF OUTPUT INDICATORS ARE ACCESSIBLE TO USERS. COLOR ENCODES THE MAGNITUDE OF CHANGE IN AN IN-DICATOR. (B) ILLUSTRATION OF A GEOGRAPHICAL OVERVIEW IN VISEIO-LCA [37]. USERS CAN ZOOM INTO IMPACTED REGIONS (IN RED) AND GATHER DETAILS.

spaces. In visualizations that overlay env. and other design data, visual variables coding env. indicators should remain persistent through view transformations.

6.2 Zoom (T2): Zoom in on items of interest

Zooming is an interaction technique for narrowing the search or decision space by changing the magnification level of the visualization. Methods for zooming include (1) geometric zoom, where object sizes are visually magnified, and (2) semantic zoom, where the size and/or context of the object changes in relation to the information being presented [53]. In the context of SPD, zooming can be useful in controlling the level-of-detail for aggregated data such as env. indicators [49], and for increasing magnification to provide additional detail [37].

6.2.1 Illustrative examples in SPD

OuestVis: This interface enables multilevel indicator browsing using expandable color-encoded boxes (Fig. 3(a)). Three levels of magnification are accessible, including (1) a top level showing aggregate values for each of the 13 categories, (2) an intermediate level showing the most important indicators in a specific category, and (3) the lowest level showing all 294 output indicators across 13 categories. QuestVis preserves transparency of the aggregation process by allowing users to zoom into details. Another advantage of multilevel indicator browsing is that the overview visualization uses very little screen space, and allows users to easily compare indicators across different scenarios.

VisEIO-LCA, SourceMap: VisEIO-LCA and Sourcemap use geographical overlays. In VisEIO-LCA, a geographic map visualizes locations affected by env. impacts (Fig. 3(b)). Users can magnify the displayed map to access more detailed information about specific regions. Similarly, SourceMap uses pan+zoom interaction for navigating multiscale geospatial datasets.

i-Tree: The i-Tree uses a multi-level approach for products and processes to analyze complex inventories. The highest level



FIGURE 4: (A) SHAPESIFT [39] ALLOWS USERS TO FILTER RESULTS BASED ON SIMILARITIES IN MATERIAL, MANUFAC-TURING, AND FUNCTION. ALTERNATIVES THAT ARE FILTERED OUT ARE GRAYED IN THE SQUARIFIED LAYOUT VISUALIZATION. (B) OUESTVIS [49] ALLOWS USERS TO FILTER OUT SCENARIO OVERVIEWS FOR OUTPUT INDICATORS BASED ON THE MAGNI-TUDE OF CHANGE OF THAT INDICATOR. HERE, SCENARIOS IN BLUE REPRESENT GREATEST INCREASE OF THE CORRESPOND-ING INDICATOR. IN THIS CASE, FILTERED SCENARIOS ARE EN-TIRELY REMOVED FROM THE VIEW.

provides a view of the entire life cycle. Here, flows and processes representing sub-assemblies are shown. For gathering more detail, designers can traverse down levels and shrink the scope of the life cycle to a single manufacturing process or operation. Therefore, designers can zoom in on areas of interest in the lifecycle while retaining an overall view of the entire product.

6.2.2 Suggested design patterns

The discussed tools primarily use zooming as a navigation aid for drilling down on design data and env. indicators from overviews. We infer the following design patterns for this task.

- P4: Intent-based aggregation: Zooming can be used as an interaction strategy for promoting transparency by facilitating drilling down on aggregated env. indicators and designrelated metrics. This is particularly useful for SPD as env. indicators are often constructed by aggregating across impact categories using subjective weighting factors [54]. Aggregation also enables designers with varying levels of expertise in env. assessment to use a common visual platform for exploring the data in various levels of detail. Interested readers are directed to the paper by Elmqvist & Fekete [55] that reviews InfoVis-based guidelines for aggregation.
- P5: Multiscale design exploration: Visual tools for SPD should potentially support multiscale exploration across both (1) system hierarchies and (2) lifecycle stages. Tools should also help designers maintain multiscale (across hierarchies or stages), as well as multifocus (across a set of selected alternatives) awareness in regards to env. performance. Such multiscale exploration tools have also been used for exploring the dynamics of env. systems [56].

6.3 Filter (T3): Filter out uninteresting items

Exploring large collections of designs presents the challenge of downselecting a set of feasible alternatives under imposed constraints. Filtering serves as an interactive technique to locate and focus on specific alternatives or regions of interest [57]. Filtering can provide dynamic visual feedback about interesting items through emphasis. This can be achieved by either (1) removing uninteresting items, or by (2) making them less prominent by changing a visual variable. Filtering can also facilitate collaborative, multi-user pruning of the exploration space.

6.3.1 Illustrative examples in SPD

shapeSIFT: The shapeSIFT interface encodes design metadata (such as material, manufacturing, and functions) using a coloring scheme (Fig. 4(a)). The underlying framework [39] provides a method for measuring similarities along multiple dimensions using known taxonomic structures for the metadata. The result of each similarity computation is a scalar measure of similarity between two parts along each dimension. This allows users to set numerical similarity thresholds (based on a reference part) and downselect parts that are retrieved by a sketch query. Parts that are *filtered* out are visually de-emphasized from the squarified layout view by graying out their background. Since the overall dimensionality of the data is preserved, shapeSIFT can facilitate collaborative pruning wherein, each domain expert (i.e. material, manufacturing) can individually set similarity thresholds.

OuestVis: In QuestVis, an overview visualization of the scenario space is presented to the user as a scatter plot. The magnitude of change of a particular output indicator (such as water use. and industrial energy use) is encoded using color (Fig. 4(b)). In this case, a filtering operation based on the magnitude of change in the indicator allows users to selectively remove scenarios from the viewport. To preserve context, QuestVis displays each scenario (point) in the same spatial location.

Uchil & Chakrabarti: The developed interface uses filtering for setting the system boundary of the results displayed in the env. assessment. For example, if a designer selects material extraction and manufacturing, then the LCA results update to show the impacts resulting from these stages.

i-Tree: The i-Tree emphasizes flow-specific ecoimprovements when a user focuses on a particular material or energy flow. Visual filtering is achieved by retaining color on the focused items and fading the rest of the product life cycle.

6.3.2 Suggested design patterns

The discussed tools use filtering as a means for reducing the visual complexity of overview visualizations by de-emphasizing uninteresting alternatives or dimensions. Query-based interactions are also used to widening or refocusing the search space. By providing dynamic visual feedback during filtering, such tools also allow designers to explore relationships between resulting env. indicators and design parameters.

Bernstein, William; Ramani, Karthik; Ramanujan, Deverajan. "Design Patterns for Visualization-Based Tools in Sustainable Product Design." Paper presented at ASME 2017 International Design Engineering Technical Conference & Computers and information in Engineering Conference. IDETC/CIE 2017, Cleveland, OH, United States. August 6, 2017 - August 9, 2017.



FIGURE 5: (A) VISER [40] EMPHASIZES INWARD AND OUTWARD NODE LINKS ON MOUSE HOVER. SELECTING A NODE GENERATES A TOOLTIP WITH DETAILED INFORMATION. (B) SOURCEMAP [50] DETAILS THE CARBON FOOTPRINT, AND SUPPLIER INFORMA-TION ON NODE SELECTION.

- P6: Emphasis on design similarities: Filtering interactions should potentially emphasize similarities in design attributes. This can be useful for SPD tasks wherein designers explore similar alternatives that are more benign. Emphasis can be achieved by (1) removing items from the scene while retaining a visual variable (i.e. spatial location, color), or (2) retain items on scene and de-emphasize filtered out items by changing visual variables, e.g. color, spatial location, or size [51].
- P7: Collaborative pruning: Visual interfaces for SPD should potentially facilitate filtering large design spaces with respect to sustainability as well as design-related metrics. Facilitating collaborative pruning of the design space among all involved stakeholders in SPD allows them to realize globally feasible outcomes, especially when information is scarce [58].

6.4 Details on demand (T4): Select an item of a group and get details when needed

When exploring a large decision space the number of data points in overviews create a high level of visual complexity. Although it is beneficial to preserve context while exploring the data, limitations in screen sizes and resolution makes it impractical to visualize details about all data points. Details on demand serves as an interaction technique that provides additional information about single/multiple selected objects while preserving the overview. Examples for this interaction technique include, information tooltips presented on mouse hover, and a separate detail viewport that provides information on selected objects.

6.4.1 Illustrative examples in SPD

ViSER, SourceMap: ViSER implements details on demand via information tooltips. When a user selects a node from the supply chain tree, details such as, ID label, node number, and env. impacts are displayed in a tooltip (Fig. 5(a)). Furthermore, when a user hovers over a node, the ID label is shown as a tooltip and edges directly connected to that node (both inward and outward) are emphasized using color. Similarly, SourceMap presents details about the supplier and the env. footprint of a supply stage as



FIGURE 6: VISER [40] IMPLEMENTS TWO MUTUALLY COOR-DINATED PANES REPRESENTING A SUPPLY CHAIN TREE AND A PRODUCT ARCHITECTURE GRAPH. SELECTING A PART FROM THE ARCHITECTURE PANE, UPDATES THE SUPPLY CHAIN TO SHOW PATHWAYS FOR THE CORRESPONDING PART.

a tooltip (Fig. 5(b)). SourceMap also presents example calculations for carbon footprint on the tooltip, promoting transparency in env. assessment.

VisEIO-LCA: A detail window implemented in VisEIO-LCA allows users to explore detailed information pertaining to different visualizations. For example, while exploring a visualization of the economic matrix and toxic release inventory (TRI) vector data, the detailed view presents a magnified view of a userselected region. Similarly, when the user selects a point from the scatter plot visualization, the name of the sector and the exact value for the specific point is displayed in the detail window.

6.4.2 Suggested design patterns

Details on demand serves as a useful interaction technique for presenting additional data dimensions not captured in the existing visualization. These details can be presented as raw metadata or embedded visualizations. This allows presentation of details on overview visualizations while preserving the context of the overall design space. We infer the following design pattern.

P8: Interactive detailing of hidden dimensions: User interactions that select or focus visual elements should dynamically present details on non-visualized dimensions for multidimensional data such as env. indicators or design metadata. Results can be presented through mechanisms such as dynamic labels, tooltips, and visual emphasis.

6.5 Relate (T5): View relationships among items

A lifecycle of a typical product consists of multiple interacting products and lifecycle stages. This creates the need for multiple representations for lifecycle data. SPD requires designers to comprehend several such representations to fully evaluate design alternatives. Providing visual cues for relating data across these representations simplifies navigation and decision-making processes. For example, relating designs by similarity measures allows spatially representing similar designs based on a reference design. This allows designers to easily identify sub-spaces that contain designs with similar characteristics.

6.5.1 Illustrative examples in SPD

ViSER: This interface consists of two mutually coordinated views that relate a supply chain of a product system to its physical architecture. Here, selecting a part from the product architecture graph (Fig. 6) emphasizes the corresponding pathway on the supply chain graph. This linking allows designers to relate supply chain related metrics, e.g. stage time and cost, to productrelated metrics, e.g. modularity. Also, resulting env. impacts are simultaneously overlayed on both views. This allows designers to ask questions like "what is the most impactful supply chain stage for the most impactful and most modular part?"

Uchil & Chakrabarti: The interface developed by the authors used a multi-view visualization to help designers identify unforeseen relationships between design parameters and LCA results. Dynamic sliders present in the interface help designers explore the effect of a change in system boundary or functional unit on the resulting env. impact.

i-Tree: The i-Tree uses TRIZ (Theory of Inventive Problem Solving) to relate LCA results to design. To do so, the authors develop a graphical ontology that guides designers in mapping the product life cycle, identifying and emphasizing critical flows, and tracking eco-improvement. i-Tree relates eco-improvement and env. impact by visualizing criticalities- a measure of the available improvement based on the efficiency of the product or process-using a Sankey chart.

6.5.2 Suggested design patterns

Relating and visually presenting dependencies in env. indicators and lifecycle data can help designers contextualize env. assessment to design-related lifecycle variables. We infer the following design patterns based on our review.

- P9: Co-ordination of lifecycle views: Lifecycle data should be presented using multiple mutually coordinated views. This allows designers to explore multiple representations of lifecycle data and evaluate sources of env. footprint. In these views, relations between data across the visualized lifecycle stages should be interactively emphasized. Interested readers are directed to the work by Wang et al. [59] that discusses when and how to design multiple-view systems.
- P10: Linking indicators through the lifecycle: Env. indicators should be visualized in the context of a specific lifecycle stage (such as the manufacturing or use phase) and linked to show its contribution to the overall env. footprint of the design or the system.

6.6 History (T6): Keep a history of actions to support undo, replay, and progression of refinement

SPD is an open-ended process and, most often, there is little guidance for the designer about the direction of search. Furthermore, as one of the goals in SPD is to facilitate learning, it is important for visualization-based tools in SPD to reduce the cost



FIGURE 7: QUESTVIS [49] SHOWS HISTORY OF USER EXPLO-RATIONS ON THE SCENARIO OVERVIEW WITH A TRAIL. EACH CIRCLE REPRESENTS A FUTURE SCENARIO AND THE CIRCLE'S COLOR ENCODES CHANGE IN A SPECIFIC OUTPUT INDICATOR.

of exploration. This can be achieved by allowing designers to (1) visualize the consequence(s) of their decisions through the exploration process, and (2) retrace their steps and branch out in a different direction [60]. In collaborative SPD tasks, visualizing exploration pathways across the group can provide individual designers with an awareness of one another's decisions.

6.6.1 Illustrative examples in SPD

QuestVis: Scenario space overviews in the QuestVis interface display a trail marking recently selected scenarios (Fig. 7). This helps users create associations between input choices to the QUEST model and the output indicators for future scenarios.

6.6.2 Suggested design patterns

We found that only QuestVis allows for saving the entire history of the exploration process. While some other tools have the ability to save current state, the need to save the entire history trail is critical with the increasing complexity of the exploration process and the need for collaborative decision-making. We infer the following design patterns based on our review.

- P11: Eco-location: Exploration paths should be visually presented to designers so that they can serve as navigation aids. Such visualizations should overlay consequences (env. indicators) due to input choices for facilitating designers' learning from previous decision scenarios.
- P12: Shareable exploration trails: In collaborative tools exploration paths of each designer should be visualized in a shareable manner. This facilitates group-awareness of the exploration processes [61]. Saved exploration states across shared pathways should serve as anchor points for branching out neighboring regions in the design/decision space.

6.7 Extract (T7): Allow extraction of sub-collections and of the guery parameters

Extraction is an interaction task that allows users to (1) create and export sub-collections containing items of interest, and (2) save and export settings, for interactions controls implemented by an InfoVis tool [19]. The overall goal is to reduce the tediousness associated with reporting or reusing previous knowledge obtained through a set of interaction procedures.



FIGURE 8: SUB-MATRIX OF OM-GLYPHS (RIGHT) EXTRACTED FROM A LARGER CO₂ GLYPH MATRIX OF THE ENTIRE INVEN-TORY (LEFT) [38]. THIS EXTRACTION STEP ENABLES USERS TO VISUALIZE GREATER DETAIL FOR A SELECTED SET OF GLYPHS. IN EACH GLYPH, THE RADIUS OF THE CENTRAL SPHERE REPRE-SENTS THE TOTAL SUM OF CO2 EMISSIONS FOR A SPECIFIC COM-PONENT IN AN ASSEMBLY. THE PROTRUSIONS FROM THE SPHERE ENCODE CONTRIBUTIONS FROM EACH LIFE CYCLE STAGE.

Providing a means for extraction is an important capability for visualization-based tools in SPD as reporting is an integral part of comparative env. assessment and design exploration.

6.7.1 Illustrative examples in SPD

OM-Glyphs: Otto et al. [38] present a matrix-style visualization of OM-glyphs that encode carbon dioxide emissions for an entire inventory of an electronic product (Fig. 8). In this view, the large number of glyphs makes it difficult for comparing glyphs with one another, and detecting data imperfections (based on the sphericity of the base sphere). To overcome such a limitation, users can extract a sub-collection of OM-glyphs in the form of a sub-matrix. As a smaller number of OM-glyphs are shown in this view, each glyph can be shown in greater detail.

QuestVis: The QuestVis interface allows users to save their value judgments (positive, neutral, and negative) for output indicators in the QUEST model. This set of values can be reloaded onto the interface for future exploration tasks.

6.7.2 Suggested design patterns

Extracting a set of alternatives from a larger design space can help designers develop functional units for contextualizing env. indicators to design. For example, extracting rotation-based transmission parts from a repository can allow designers to understand the env. impact of transmitting a unit torque. Furthermore, creating affordances for extracting and exporting queries, interesting results, and exploration methods can help ease the barriers to collaboration, leading to the following design pattern.

P13: Exploration snippets: Designers should be able to create saveable snippets of both data and metadata pertaining to the exploration process. Adoption of any designed visualizationbased tools for SPD can be furthered by facilitating reuse, reporting, and exchange, of such snippets.

7 DISCUSSION & LIMITATIONS

While the set of design patterns here serve as a review of best-practices based on previous tools, the implementation of

TABLE 2: SUGGESTED DESIGN PATTERNS FOR VISUALIZATION-BASED TOOLS FOR SPD (P#) MAPPED TO TASKS IN THE TASK-BY-TYPE TAXONOMY (T#)

		Indicator overviews	P1
T1	Overview	Eco-prominence	P2
		Eco-persistence	P3
T2 Zoom	Intent-based aggregation	P4	
	20011	Multi-scale design exploration	P5
T3 Filter	Filter	Emphasis on design similarities	P6
	Tiller	Collaborative pruning	P7
T4	Details on demand	Interactive detailing of hidden dimensions	
T5 Relate	Palata	Co-ordination of lifecycle views	P9
	Relate	Linking eco-indicators through the lifecycle	P10
T6 History	History	Eco-location	P11
	mstory	Shareable exploration trails	P12
T7	Extract	Exploration snippets	P13

these patterns into future tools requires design judgement. In other words, the suggested design patterns should be viewed as templates on which future tools can be modeled. We recognize that the use of our design patterns do not guarantee the success of the developed tool. Developing a checklist for creating a successful interface is prohibitively hard, not unlike the idea of creating a checklist for designing successful products. In both cases, success depends on human creativity, insight, and skill. Even so, we believe that creating a list of design patterns or templates is worthwhile as it provides practitioners with a starting point for rapid prototyping and evaluation of different implementations. This is particularly critical as the cost of creating information interfaces and evaluating them is often significantly high [62].

Furthermore, we recognize that the number of existing Info-Vis tools for SPD is relatively small. Hence, our design patterns do not represent a full landscape of guidelines. In the future, we intend to re-visit our list and possibly offer a more comprehensive classification. To this end, we envision that a well-structured set of design patterns would open new research opportunities for studying both advantages and challenges in using various types of visual representations and interactions in future tools.

8 CONCLUSIONS

In this paper we have presented a set of design patterns for visualization-based tools for SPD by analyzing previous works in this domain using the task-by-type taxonomy. A summary of mappings between identified design patterns (P#) to the task-bytype taxonomy (T#) is presented in Table 2. The primary contribution of our paper is a set of design patterns useful for creating future visualization-based tools for SPD. We believe that our suggested design patterns will (1) serve as a guide for cre-

ating integrated visualization-based tools for SPD and (2) contribute towards reusable visual components to aid in quick interface wireframing. One could imagine that with communitywide agreement on design patterns such as the 13 suggested here, research groups could begin developing reusable visualization components as plug-and-play features to prototype interfaces.

ACKNOWLEDGMENTS

We thank Dr. Devadatta Kulkarni, Dr. Senthil Chandrasegaran, and Prof. Niklas Elmqvist. This research work is partially supported by NSF (EEC 0935074, CMMI 110619) and the TATA Design Innovation Grant. The contents of this paper do not necessarily reflect the views or opinions of these agencies.

DISCLAIMER

No approval or endorsement of any commercial product by NIST is intended or implied. Certain commercial equipment, instruments or materials are identified in this report to facilitate better understanding. Such identification does not imply recommendations or endorsement by NIST nor does it imply the materials identified are necessarily the best available for the purpose.

REFERENCES

- [1] Boons, F., 2002. "Greening products: a framework for product chain management". J Clean Prod, 10(5), pp. 495-505.
- [2] Ramani, K., Ramanujan, D., Bernstein, W. Z., Zhao, F., Sutherland, J., Handwerker, C., Choi, J.-K., Kim, H., and Thurston, D., 2010. "Integrated sustainable life cycle design: A review". J Mech Design, 132(9), p. 091004.
- [3] Brezet, H., and van Hemel, C., 1997. Eco-design A promising approach to sustainable production and consumption. UNEP, United Nations Publications.
- [4] Skerlos, S. J., 2015. "Promoting effectiveness in sustainable design". Procedia CIRP, 29, pp. 13-18.
- [5] Waage, S. A., 2007. "Re-considering product design: a practical road-map for integration of sustainability issues". J Clean Prod, 15(7), pp. 638-649.
- [6] Babbie, E., 2015. The practice of social research. Cengage.
- [7] Bernstein, W. Z., Ramanujan, D., Devanathan, S., Zhao, F., Sutherland, J., and Ramani, K., 2010. "Function impact matrix for sustainable concept generation: A designers perspective". In Proceedings of the ASME 2010 IDETC/CIE, pp. 377-383.
- [8] Rio, M., Reyes, T., and Roucoules, L., 2013. "Toward proactive (eco) design process: modeling information transformations among designers activities". J Clean Prod, 39, pp. 105-116.
- [9] Sousa, I., Wallace, D., and Eisenhard, J. L., 2000. "Approximate life-cycle assessment of product concepts using learning systems". *J Ind Ecol*, **4**(4), pp. 61–81.
- [10] Park, J.-H., and Seo, K.-K., 2006. "A knowledge-based approximate life cycle assessment system for evaluating environmental impacts of product design alternatives in a collaborative design environment". Adv Eng Inform, 20(2), pp. 147-154.
- [11] Hernandez, N. V., Kremer, G. O., Schmidt, L. C., and Herrera, P. A., 2012. "Development of an expert system to aid engineers in

the selection of design for environment methods and tools". Expert Systems with Applications, 39(10), pp. 9543-9553.

- [12] Card, S., Mackinlay, J., and Shneiderman, B., 1999. Readings in information visualization: using vision to think. Morgan Kauf.
- [13] Munzner, T., 2014. Visualization Analysis and Design. CRC Press.
- [14] Senay, H., and Ignatius, E., 1994. "A knowledge-based system for visualization design". IEEE Computer Graphics and Applications, 14(6), pp. 36-47.
- [15] Morris, R., 2009. The fundamentals of product design. AVA.
- [16] Gamma, E., Helm, R., Johnson, R., and Vlissides, J., 1994. Design patterns: elements of reusable object-oriented software. Pearson.
- Chen, H., 2004. "Toward design patterns for dynamic analytical [17] data visualization". In Proc. SPIE, 5295, Visualization and Data Analysis 2004, pp. 75-86.
- [18] Heer, J., and Agrawala, M., 2006. "Software design patterns for information visualization". IEEE Trans Vis Comput Graph, 12(5), pp. 853-860.
- [19] Shneiderman, B., 1996. "The eyes have it: A task by data type taxonomy for information visualizations". In Proceedings of IEEE 1996 Symposium on Visual Languages, pp. 336-343.
- [20] Huang, H., Liu, Z., Zhang, L., and Sutherland, J., 2009. "Materials selection for environmentally conscious design via a proposed life cycle environmental performance index". The International Journal of Advanced Manufacturing Technology, 44, pp. 1073–1082.
- [21] Kaebemick, H., Sun, M., and Kara, S., 2003. "Simplified lifecycle assessment for the early design stages of industrial products". CIRP Annals-Manufacturing Technology, 52(1), pp. 25-28.
- [22] Devanathan, S., Ramanujan, D., Bernstein, W., Zhao, F., and Ramani, K., 2010. "Integration of sustainability into early design through the function impact matrix". J Mech Design, 132(8).
- [23] Haapala, K., Poppa, K., Stone, R., and Tumer, I., 2011. "Automating environmental impact assessment during the conceptual phase of product design". In Proceedings of the AAAI 2011 Spring Symposium: Artificial Intelligence and Sustainable Design, pp. 21-23.
- [24] Lofthouse, V., 2006. "Ecodesign tools for designers: defining the requirements". J Clean Prod, 14(15), pp. 1386-1395.
- [25] Basole, R. C., Qamar, A., Park, H., Paredis, C. J., and McGinnis, L. F., 2015. "Visual analytics for early-phase complex engineered system design support". IEEE Computer Graphics and Applications, 35(2), pp. 41-51.
- [26] Winer, E., and Bloebaum, C., 2002. "Development of visual design steering as an aid in large-scale multidisciplinary design optimization. part i: method development". Structural and Multidisciplinary Optimization, 23(6), pp. 412-424.
- [27] Simpson, T. W., Carlsen, D. E., Congdon, C. D., Stump, G., and Yukish, M. A., 2008. "Trade space exploration of a wing design problem using visual steering and multi-dimensional data visualization". In Proc. of the 4th AIAA Multidisciplinary Design Optimization Specialist Conference, pp. 7-10.
- [28] Stump, G. M., Yukish, M., Simpson, T. W., and Harris, E. N. "Design space visualization and its application to a design by shopping paradigm". In Proc. of the ASME 2003 IDETC/CIE, pp. 795-804.
- [29] Zhang, X., Simpson, T., Frecker, M., and Lesieutre, G., 2012. "Supporting knowledge exploration and discovery in multidimensional data with interactive multiscale visualisation". J Eng Design, 23(1), pp. 23-47.

- [30] Chi, E. H., 2000. "A taxonomy of visualization techniques using the data state reference model". In Proc. of the 2000 IEEE Symposium on InfoVis, pp. 69-75.
- [31] Keim, D., et al., 2002. "Information visualization and visual data mining". IEEE Trans Vis Comput Graph, 8(1), pp. 1-8.
- [32] Holmes, T. G., 2007. "Eco-visualization: combining art and technology to reduce energy consumption". In Proc. of the 6th ACM SIGCHI Conference on Creativity & Cognition, pp. 153-162.
- [33] Froehlich, J., Findlater, L., and Landay, J., 2010. "The design of eco-feedback technology". In Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 1999-2008.
- [34] Pousman, Z., Rouzati, H., and Stasko, J., 2008. "Imprint, a community visualization of printer data: designing for open-ended engagement on sustainability". In Proc. of the 2008 ACM Conference on Computer Supported Cooperative Work, pp. 13-16.
- [35] Kim, T., Hong, H., and Magerko, B., 2010. "Designing for persuasion: toward ambient eco-visualization for awareness". In Persuasive technology. Springer, pp. 106-116.
- [36] Lilley, D., 2009. "Design for sustainable behaviour: strategies and perceptions". Design Studies, 30(6), pp. 704-720.
- [37] Espinosa, O. J., Hendrickson, C., and Garrett, J. "Domain analysis: a technique to design a user-centered visualization framework". In Proc. of the 1999 IEEE Symp. on InfoVis, pp. 44-52.
- [38] Otto, H. E., Mueller, K. G., and Kimura, F., 2004. "Efficient information visualization in LCA: Application and practice". The International Journal of Life Cycle Assessment, 9(1), pp. 2–12.
- [39] Ramanujan, D., Bernstein, W. Z., Benjamin, W., Ramani, K., Elmqvist, N., Kulkarni, D., and Tew, J., 2015. "A framework for visualization-driven eco-conscious design exploration". J Comput Inf Sci Eng, 15(4), p. 041010.
- [40] Bernstein, W. Z., Ramanujan, D., Kulkarni, D. M., Tew, J., Elmqvist, N., Zhao, F., and Ramani, K., 2015. "Mutually coordinated visualization of product and supply chain metadata for sustainable design". J Mech Design, 137(12), p. 121101.
- [41] Ramanujan, D., 2015. "Data representation methods for environmentally conscious product design". PhD thesis, Purdue U.
- [42] Bonanni, L., 2011. "Sourcemap: eco-design, sustainable supply chains, and radical transparency.". ACM Crossroads, 17(4), pp. 22-26.
- [43] Gutiérrez, E., Lozano, S., and Adenso-Díaz, B., 2010. "Dimensionality reduction and visualization of the environmental impacts of domestic appliances". J Ind Ecol, 14(6), pp. 878-889.
- [44] Morbidoni, A., Favi, C., and Germani, M., 2011. "Cad-integrated lca tool: Comparison with dedicated lca software and guidelines for the improvement". Glocalized solutions for sustainability in manufacturing, pp. 569-574.
- [45] Uchil, P., and Chakrabarti, A., 2015. "An interface between life cycle assessment and design". In ICoRD15-Research into Design Across Boundaries Volume 2. Springer, pp. 251-259.
- [46] Russo, D., Serafini, M., and Rizzi, C., 2016. "Triz based computer aided lca for ecodesign". Computer-Aided Design and Applications, 13(6), pp. 816-826.
- [47] Mutel, C., Cardellini, G., Froemelt, A., Heeren, N., Jaggi, A., Marcus, M., de Saxceé, M., Seong, L. K., and Steubing, B., 2015. Brightway2 LCA framework. Available at http://brightwaylca.org. Accessed 20-01-2016.

- [48] Kuczenski, B., and Beraha, S., 2015. "Antelope: A web service for publishing life cycle assssment models and results". In Proceedings of the 2015 ISSST.
- [49] Munzner, T., Barsky, A., and Williams, M., 2011. "Reflections on questvis: A visualization system for an environmental sustainability model.". Scientific Visualization: Interactions, Features, Metaphors, 2(240-259), p. 36.
- [50] Sourcemap, 2015. Starbucks supply chain visualization on sourcemap. Available at http://free.sourcemap.com/view/6039. Accessed 06-02-2015.
- [51] Stasko, J., 2013. CS 7450-information visualization. Lecture notes, www.cc.gatech.edu/~stasko/7450/13/Notes/interaction.pdf.
- [52] Healey, C. G., Booth, K. S., and Enns, J. T., 1996. "High-speed visual estimation using preattentive processing". ACM Transactions on Computer-Human Interaction (TOCHI), 3(2), pp. 107-135.
- [53] Perlin, K., and Fox, D., 1993. "Pad: an alternative approach to the computer interface". In Proceedings of the 20th annual conference on computer graphics and interactive techniques, ACM, pp. 57-64.
- [54] Owens, J. W., 1996. "LCA impact assessment categories". The International Journal of Life Cycle Assessment, 1(3), pp. 151–158.
- [55] Elmqvist, N., and Fekete, J.-D., 2010. "Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines". IEEE Trans Vis Comput Graph, 16(3), pp. 439-454.
- [56] Sips, M., Köthur, P., Unger, A., Hege, H.-C., and Dransch, D., 2012. "A visual analytics approach to multiscale exploration of environmental time series". IEEE Trans Vis Comput Graph, 18(12), pp. 2899-2907.
- [57] Ahlberg, C., and Shneiderman, B., 1994. "Visual information seeking: Tight coupling of dynamic query filters with starfield displays". In Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 313-317.
- [58] Pirolli, P., and Card, S., 1995. "Information foraging in information access environments". In Proc. of the 1995 ACM SIGCHI Conference on Human Factors in Computing Systems, ACM Press/Addison-Wesley Publishing Co., pp. 51-58.
- [59] Wang Baldonado, M. Q., Woodruff, A., and Kuchinsky, A. "Guidelines for using multiple views in information visualization". In Proc. of the 2000 ACM Conference on AVI, pp. 110-119.
- [60] Dunne, C., Henry Riche, N., Lee, B., Metoyer, R., and Robertson, G., 2012. "GraphTrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history". In Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp. 1663-1672.
- [61] Park, K., Renambot, L., Leigh, J., and Johnson, A., 2003. "The impact of display-rich environments for enhancing task parallelism and group awareness in advanced collaborative environments". In Workshop on Advanced Collaboration Environments.
- [62] Jones, W., Spool, J., Grudin, J., Bellotti, V., and Czerwinski, M., 2007. "Get real!: what's wrong with hci prototyping and how can we fix it?". In CHI'07 Extended Abstracts on Human Factors in Computing Systems, ACM, pp. 1913–1916.

The Science Behind Community Resilience Assessment: An Overview of the Center for Risk-Based Community **Resilience Planning**

Bruce R. Ellingwood^a, John W.van de Lindt^a, and Therese P. McAllister^b

^aDepartment of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523 USA ^bEngineering Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899 USA

Abstract: Community resilience depends on the performance of the built environment and on supporting social, economic and public institutions that are essential for response and recovery of the community following a hazard event. The social needs of a community are not reflected in codes, standards, and other regulatory documents currently used to design individual facilities. A new approach is necessary, one which is interdisciplinary in nature and reflects the complex inter-dependencies among the physical, social, and economic systems on which a healthy and vibrant community depends. The Center for Risk-Based Community Resilience Planning, headquartered at Colorado State University in Fort Collins, Colorado, and eight partner universities, was established by the National Institute of Standards and Technology to advance the measurement science for understanding the factors that make a community resilient, to assess the likely impact of natural hazards on communities, and to develop risk-informed decision strategies that optimize planning for and recovery from disasters. This presentation summarizes the approach taken by the Center's research teams during its first two years to advance the science underlying community resilience assessment and risk-informed planning and recovery strategies.

Keywords: Buildings, Civil infrastructure, Decision algorithms, Natural hazards, Risk.

1 Introduction

Community resilience depends on the performance of the built environment and on supporting social, economic, and public institutions that, individually and collectively, are essential for immediate response and long-term recovery within the community following a hazard event. The performance of the built environment, which is a key factor in community resilience, is largely determined by codes and standards, which are applicable to individual facilities and have the primary objective of preserving life safety under severe events. Current design standards, such as ASCE Standard 7 [1], generally do not address facility performance in the period

Ellingwood, Bruce; McAllister, Therese; van de Lindt, John. "The Science Behind Community Resilience Assessment: An Overview of the Center for Risk-Based Community Resilience Planning." Paper presented at 12th International Conference on Structural Safety & Reliability, Vienna, Austria. August 6, 2017 - August 10, 2017.

of recovery following an event. Moreover, the design of interdependent transportation systems, utilities, and communication systems currently is based on different performance criteria. Accordingly, there is no assurance that all systems required for community resilience will perform at a consistent level during and following an extreme hazard event. Furthermore, the resilience goals of a community are based on social needs and objectives that are specific to its character - its prior experience with natural hazards, the vulnerability of the population, economic and financial drivers and resources, and local building regulations and construction practices, all of which are factors that fall outside the purview of infrastructure design. Finally, much of the needed science only exists in rudimentary form at present. This includes sciencebased measurement tools to address deficiencies in current practices and to evaluate performance and resilience at the community scale, fully integrated supporting databases, and riskinformed decision frameworks to support optimal life-cycle technical and social policies aimed at enhancing community resilience [2, 3].

Natural and man-made disasters in the United States are responsible for over \$55 billion in average annual costs in terms of injuries and lives lost, disruption of commerce and economic networks, property damaged or destroyed, the cost of mobilizing emergency response personnel and equipment, and recovery of essential services [4,5]. The losses due to such events are increasing more rapidly than the growth in population or the Gross Domestic Product. The potential exists for even larger losses in the future, given that population and infrastructure development in hazard-prone areas of the United States [as of 2010, 39 % of the U.S. population lived in a county that touches the coast (www.noaa.gov)] are increasing dramatically and that climate change may affect both the frequency and severity of the extreme events from natural hazards [6]. A new approach is needed, one that reflects the complex inter-dependencies among the physical, social and economic systems on which community well-being depends. The vast majority of research on community resilience in the past decade has focused on the impact of severe earthquakes on the physical infrastructure in communities [7, 8, 9, 10] or on the impact of hazard events on community social institutions [e.g., 11, 12]. Little attention has been paid to other natural hazards, including those that might be impacted by climate change [6, 13]. Resilience assessment has become a national imperative [5, 14], not only in the United States but in Europe and the countries in the Asia-Pacific Rim.

The Center for Risk-Based Community Resilience Planning, headquartered at Colorado State University in Fort Collins, Colorado, was established by the National Institute of Standards and Technology (NIST) in 2015. The Center's overarching goals are to establish the measurement science for understanding the factors that make a community resilient, to assess the likely impact of natural hazards on communities, and to develop risk-informed decision tools and strategies that optimize planning for and recovery from hazard events and are consistent with local community financial constraints, values and preferences. In this paper, we introduce this Minisymposium on Enhancing Urban Resilience Under National Hazards by presenting an overview of the Center's research activities during its first two years, including multiple hazards and their cascading effects on infrastructure, the role of supporting economic networks and social systems on community resilience and post-hazard event recovery, the impact of aging infrastructure, and the identification and articulation of performance metrics and requirements. We include several community resilience testbeds, which have been designed to allow Center research teams to initiate, test, and modify essential community resilience assessment models

Ellingwood, Bruce; McAllister, Therese; van de Lindt, John. "The Science Behind Community Resilience Assessment: An Overview of the Center for Risk-Based Community Resilience Planning." Paper presented at 12th International Conference on Structural Safety & Reliability, Vienna, Austria. August 6, 2017 - August 10, 2017.

and algorithms early in the program and to facilitate the essential interdisciplinary collaborations and approaches to community resilience assessment that will be necessary for the Center's ultimate success. We conclude with a summary of significant challenges for community resilience research and Center research activities designed to address many of them.

2 **Summary of Center Research Tasks**

The Center is engaged in three major research thrusts to accomplish its programmatic goals: (1) developing the interdependent networked community resilience modeling environment (IN-CORE) to quantitatively assess alternative community resilience strategies; (2) instituting a standardized data ontology, robust architecture, and management tools to support IN-CORE; and (3) performing a comprehensive set of testbeds and disaster hindcasts to validate this advanced computational modeling environment and to lead to practical risk-informed decision tools for community resilience assessment and risk mitigation.

Thrust (1) Development of the IN-CORE modeling environment: Initial efforts within the Center during its first two years have focused on Thrust 1, on which all else depends. Thrust 1 consists of eight tasks, which are running concurrently, rather than sequentially: (1) Hazards modeling; (2) Damage, loss, and recovery of physical infrastructure systems modeling; (3) System modeling of community resilience; (4) Development of the IN-CORE modeling environment; (5) Model validation by hindcasting; (6) Development of performance metrics for community resilience; (7) Resolution and scalability studies; and (8) Development of optimization strategies and decision methods for community resilience.

Natural hazards considered include earthquakes and tsunamis, tornadoes, hurricane winds, coastal storm surge and inundation, riverine flooding, and wildfires. In some cases, the individual hazard formulations are reasonably well-understood, especially for earthquakes where the National Earthquake Hazard Reduction Program has produced advanced knowledge of earthquake hazards and civil infrastructure response over the past four decades. In other cases (riverine flooding, storm surge, tornado, and wildfires), improved models of the hazard and its impact on the built environment are required for quantitative community resilience assessment. Similarly, the state of the art regarding the performance of individual constructed facilities (e.g., buildings, bridges, buried piping, electrical transmission and distribution systems) and the integrity of individual infrastructure systems (electrical, gas, and water distribution systems) during hazard events is reasonably mature. Such systems are interconnected, however, and their functioning is dependent on the availability and functioning of other connected systems [15].

The distinctive features of each hazard (e.g., advance warning time, area affected, type and severity of damage, populations displaced) have caused hazard mitigation methodologies to be strongly hazard-dependent. Multiple hazards, and differences in community response to them, or synergies that might be achieved in policies to mitigate risk or enhance community resilience under multiple hazards, have received limited attention [16]. Furthermore, the performance of physical infrastructure systems during and following a hazard event may be positively correlated depending on the spatial scale of the hazard and the interconnected nature of their successful (or unsuccessful) operations within the community. These positive spatio-temporal correlations for physical systems must be considered in resilience assessment and combined

Ellingwood, Bruce; McAllister, Therese; van de Lindt, John. "The Science Behind Community Resilience Assessment: An Overview of the Center for Risk-Based Community Resilience Planning." Paper presented at 12th International Conference on Structural Safety & Reliability, Vienna, Austria. August 6, 2017 - August 10, 2017.

with social and economic models which affect both the ability to prepare for and recover from an event. Finally, the numerous sources of uncertainties associated with the life cycle performance of infrastructure systems mandate a risk-informed decision-making approach to assess facility and community risks and to identify cost-effective strategies to enhance community resilience [17].

Accordingly, the first two years of the Center have seen the compilation of existing stochastic models for performance of the built environment in the form of hazards and multi-dimensional fragilities; the development of new stochastic models of infrastructure performance where none previously existed; consideration of interdependencies including physical, social, and economic interactions; development of computable general equilibrium models to assess the economic impacts of a specific hazard on a community in terms of job loss, declines in tax base, reductions in household income, and other measures of community well-being; inclusion of the effects of household demographics; and validation of the fundamental resilience assessment algorithms for specific communities and hazard scenarios with default stochastic models (summarized in Section 3).

Thrust (2) Development of standardized data ontologies, robust architecture, and management tools to support IN-CORE: IN-CORE is being designed to be an open-source, integrated multi-scale, multi-hazard computational modeling environment, providing extension points where software developers can add new modules to simulate response, interactions, and recovery of major infrastructure, social, or economic systems. Since community resilience assessment involves multiple domains of science, engineering, and socio-economics, defining a common language within those communities is essential. Databases in international resiliencerelated research are being developed by different users, using different methods and have different purposes. It is essential that these databases interface seamlessly in the IN-CORE environment. This task focuses on collecting and analyzing user requirements, metrics, and measurements within the user communities and stakeholders, at the same time that the algorithms being developed in Thrust (1) are coded into the computational modeling environment.

Thrust (3) Validation of the modeling environment from testbeds and hindcasts to develop decision tools: While thrust (3) must await the completion of IN-CORE Version 1, which is anticipated in year 3 of the Center, hindcasts and field studies have progressed in year 2, as described in Section 3.

3 **Community Resilience Studies**

The Center has several community resilience studies in progress to provide focus to Center engineers in developing the physical infrastructure models – hazards, fragilities, and infrastructure interdependencies - needed for community resilience assessment. These studies are intended to allow the Center's interdisciplinary research teams to initiate, test, and modify resilience assessment models and algorithms prior to the time IN-CORE becomes fully operational; to stress these assessment models in a controlled manner; to examine varying degrees of dependency between physical, social and economic infrastructure systems; to allow issues of scalability in modeling to be addressed; to inform the subsequent development of more refined community resilience assessment methods; and to facilitate interdisciplinary collaborations and

Ellingwood, Bruce; McAllister, Therese; van de Lindt, John. "The Science Behind Community Resilience Assessment: An Overview of the Center for Risk-Based Community Resilience Planning." Paper presented at 12th International Conference on Structural Safety & Reliability, Vienna, Austria. August 6, 2017 - August 10, 2017.

approaches to community resilience assessment that will be essential for the Center to achieve its goals. Each of these studies, summarized below, has a somewhat different objective. We refer to some of these studies as *testbeds*; these are intended to study the impact of various scenario hazards on communities that have not yet been exposed to such hazards, and are used for purposes of algorithm and database development, to resolve issues of modeling interdependencies, scaling, and resolution, and to capture the interfaces between physical/social/economic systems. Other studies are referred to as *hindcasts*; in contrast to the testbeds, the hindcasts represent an attempt to reproduce what actually happened during an extreme hazard event, using the physics-based models of physics-based infrastructure behavior and integrated social and economic databases developed in other Center research tasks. The hindcasts requires careful treatment of data from past events to ensure they do not bias the development of the models. We view both testbeds and hindcasts as essential ingredients of validating the algorithms and databases in IN-CORE.

The Centerville Virtual Community Testbed (Earthquake and Tornado) 3.1

The Centerville Testbed involves a virtual community of about 50,000 in the Central United States. It is an average community in most respects (population demographics were obtained from the American Community Survey, published by the US Census Bureau), although there are pockets of low-to-moderate income residents who may be especially vulnerable to a natural hazard event. Its diversified economy includes commercial/retail, professional services, education/healthcare, industry and public sectors. The physical infrastructure includes a variety of residential, commercial and industrial buildings, bridges and transportation facilities, and utility networks, each of which represents a distinct spatial infrastructure topology [18, 19, 20, 21, 22]. Physical and socioeconomic systems are modeled in this testbed as interdependent; the performance of physical infrastructure under stipulated scenario hazards provides the initial conditions to the social and economic models needed to obtain a broad perspective of the impact of such hazards on community well-being. A recent issue of Sustainable and Resilient Infrastructure [23] is devoted to the Centerville Testbed. Topics covered include a description of Centerville, its building inventory, transportation network, coupled water and electrical power network, and its community economic and social demographics; a building inventory fragility analysis to support multiscale community resilience assessment; a multi-objective optimization approach for allocating retrofit resources to minimize population dislocation and economic losses; performance of electrical power networks during scenario tornadoes; performance of interdependent electrical and water systems under scenario earthquakes; and the use of computable general equilibrium (CGE) economic models for assessing impact of natural hazard events on a typical community.

Seaside, Oregon Testbed (Earthquake and Tsunamis) 3.2

In contrast to Centerville, Seaside, OR is a real community, a popular oceanfront resort of approximately 6,500 permanent inhabitants, situated on the Pacific coast about 80 miles NW of Portland, OR. The Oregon/Washington coast is approximately 120 km directly east of the Cascadia Subduction Zone, which is capable of generating Magnitude 9 earthquakes and producing a devastating tsunami of as much as 30 m in height that would reach the Oregon coast within 10 to 15 minutes. While seismic fragility analysis is relatively mature, the vast majority of fragilities for tsunami loading that have been developed are based on post-event observations, which are dependent on the site at which the tsunami made landfall and in which velocities are

Ellingwood, Bruce; McAllister, Therese; van de Lindt, John. "The Science Behind Community Resilience Assessment: An Overview of the Center for Risk-Based Community Resilience Planning." Paper presented at 12th International Conference on Structural Safety & Reliability, Vienna, Austria. August 6, 2017 - August 10, 2017.

unknown. Accordingly, one of the Center research tasks developed a methodology to generate physics-based tsunami fragility functions using tsunami flow depth, flow velocity, and momentum flux, in various combinations, as intensity measures [24]. There has been an increase in interest in tsunami effects following the Great East Japan Earthquake and Tsunami of 2011, several papers on tsunami effects have been published recently, and the new ASCE Standard 7-16 contains tsunami provisions for the first time.

3.3 Shelby County, TN Testbed (Earthquake, Riverine Flooding, and Climate Change)

Shelby County, TN is located in the southwest corner of Tennessee, and its western border is formed by the Mississippi River. Shelby County is a metropolitan area covering 2,033 km², with a population of approximately 940,000 inhabitants, and includes the City of Memphis, with a population of 660,000 inhabitants. Shelby County was selected for several reasons. First, it provides a realistic test of the Center's integrated engineering/social/economic models in IN-CORE for a reasonably large metropolitan area with a diverse economy and demographics. Second, a study of urban resilience under earthquakes from the New Madrid Seismic Zone in SE Missouri (about 50-60 km NW of Memphis and capable of generating Magnitude 8 earthquakes) was conducted a number of years ago by the Mid-America Earthquake Center (MAEC) (using less advanced models) [25] and can be used to benchmark the IN-CORE analysis. Third, a significant part of Shelby County lies within the Mississippi Embayment, where soil liquefaction is a significant problem and existing fragilities will have to be modified to model damage to the built environment. Fourth, the City of Memphis depends on the infrastructure provided by surrounding Shelby County, and the testbed will provide an opportunity to investigate how the resilience of a community depends on the resilience of the surrounding urban area. Fifth, unlike California, Shelby County is ill-prepared for a Magnitude 8 earthquake; older buildings and other infrastructure are not earthquake-resistant, e.g. unreinforced masonry. And finally, a large part of Shelby County falls within the Wolf River Basin drainage; one of the Center teams has initiated a study of the impact of climate change on riverine flooding in the Wolf River Basin for the remainder of the 21st Century [26], which will enable the Center to perform an analysis of competing risks and risk mitigation strategies for a major urban area.

3.4 Galveston and Bolivar Peninsula, TX (Hurricane and Storm Surge)

Galveston lies on a barrier island approximately 75 km southeast of Houston, TX on the Gulf of Mexico. It is a port city, with a population of approximately 50,000, and its economy depends on the viability of the port. The dominant hazards for Galveston are hurricane and coastal storm surge. The most recent hurricane was Hurricane Ike, which occurred in September, 2008, with wind speeds of 230 km/hr, \$US 37.5 billion in damages, and 195 fatalities. The Galveston study is attempting to reproduce the impact of Hurricane Ike on the Galveston-Bolivar Peninsula using the physics-based hazard and infrastructure response models developed during the first two years of the Center's research program. It has elements of both a hindcast (the event has already happened) and a testbed (the coupled hurricane-storm surge models will attempt to reproduce the impact of Ike on the community).

Ellingwood, Bruce; McAllister, Therese; van de Lindt, John. "The Science Behind Community Resilience Assessment: An Overview of the Center for Risk-Based Community Resilience Planning." Paper presented at 12th International Conference on Structural Safety & Reliability, Vienna, Austria. August 6, 2017 - August 10, 2017.

3.5 Joplin, MO Tornado Hindcast

Joplin, MO is a community of approximately 50,000 inhabitants, within a surrounding Metropolitan Statistical Area of about 175,000 inhabitants. In May, 2011, it was struck by a catastrophic EF5-rated multiple vortex tornado, with wind speeds exceeding 320 km/hr. The tornado killed 158 people and caused nearly \$US 3 billion in damage. It also ranks as the costliest single tornado in U.S. history. The Joplin, MO tornado prompted an investigation by the National Institute of Standards and Technology [27], which provided extensive documentation of the impact of the tornado on the community. A Center Team has recently initiated a hindcast of the Joplin tornado which utilizes many of the tools developed during the first two years of the Center including the tornado wind field model [28] and physical models of infrastructure [e.g. 29]. As part of the hindcast, researchers are focusing on a series of increasingly complex analysis starting with single sector damage validations, e.g. building damage, electrical power network disruption and damage, and then moving to coupled sector validations. A CGE model for the community is also under development, which will allow a yearly time stepping analysis to examine the economic recovery of Joplin using the functionality and recovery models that are being developed within the Center. The final validation step for the hindcast will be to model at least four physical infrastructure sectors, demographics and related population dislocation, and economic recovery from 2011 to the present day for direct comparison with what actually occurred.

Community Resilience Metrics 4

Community leaders and planners need to know how resilient their communities are in order to develop effective risk-informed public policy for enhancing community resilience. Community resilience metrics fall into two general categories: current and desired (future) resilience conditions. The current condition includes the resilience of existing systems before a hazard event, or physical damage, loss of functionality, and social and economic impacts after a hazard event. The future condition provides a goal for planning and recovery purposes. Resilience metrics need to address both categories for community systems, such as physical systems, economic vitality, and social well-being. Design codes and standards traditionally address the performance of individual facilities - buildings, bridges and roadways, utility systems - without considering how the interactions between these systems affect the welfare of the community as a whole. Functionality and recovery are not included in codes and standards; nor are economic vitality – attracting and retaining businesses and employment opportunities, revenue sources, household income, domestic economic output - or social well-being - safety and financial security, availability of food, water and shelter, population dislocation, access to health services. All are essential ingredients of community resilience.

The testbeds and hindcasts summarized in the previous section included direct damage to physical infrastructure – buildings, bridges and other transportation infrastructure, and water/wastewater and electrical power networks – as well as damages which present challenges to economic vitality and social well-being within the community. These different physical, social, and economic systems are integrated to provide quantitative measures of community resilience, including post-hazard measures, such as economic losses due to direct damage, loss of income and employment as a result of damage to residential, commercial/retail buildings, declines in the community tax base and population dislocation [30, 31, 32, 33]. The ability to assess such

Ellingwood, Bruce; McAllister, Therese; van de Lindt, John. "The Science Behind Community Resilience Assessment: An Overview of the Center for Risk-Based Community Resilience Planning." Paper presented at 12th International Conference on Structural Safety & Reliability, Vienna, Austria. August 6, 2017 - August 10, 2017.

metrics quantitatively and to determine how changes in policy might affect them is essential, both in planning for a hazard event as well as in planning for post-event recovery.

5 Concluding Remarks

Community social needs and objectives (including post-hazard recovery) are not reflected in codes, standards, and other regulatory documents applied to design of individual facilities, necessitating an approach that reflects the complex inter-dependencies among the physical, social, and economic systems on which a resilient community depends. Modeling the resilience of communities and cities to natural hazards depends on many disciplines, including engineering, social sciences, and information sciences.

A review of the literature [e.g., 2, 3, 34] has identified a number of critical challenges confronting the development of a resilient built environment, among them: (1) Quantitative metrics and tools for assessing community resilience are required to improve resilience in the built and modified natural environment and should be tailored to the specific needs of each community; (2) Community resilience plans and guidance are needed to help communities plan for hazardspecific performance, and for restoring community infrastructure systems in a cost effective and timely manner; (3) Existing building and infrastructure systems must be considered in community resilience planning as well as new construction, recognizing that existing buildings (a) may not meet current codes and standards; (b) often cannot be modified economically to meet modern design and construction practices; and (c) may have deteriorated due to structural aging or a lack of maintenance; and (4) Codes and standards with consistent performance goals for all buildings and infrastructure systems are a key component for achieving a resilient community.

The Center's current work is addressing these challenges by refining existing community infrastructure models, modeling dependent or cascading multiple hazards, developing stochastic models of the recovery processes, considering algorithms for propagating deep uncertainties in hazards and community responses, including the impact of climate change, and developing decision support algorithms and practical decision tools.

Acknowledgement

The research herein was funded by the Center for Risk-Based Community Resilience Planning, a Center of Excellence funded through a cooperative agreement between the U.S. National Institute of Science and Technology and Colorado State University (NIST Financial Assistance Award Number: 70NANB15H044).

References

- [1] ASCE Standard 7-10. Building code requirements for minimum design loads in buildings and other structures. American Society of Civil Engineers, Reston, VA, 2010.
- [2] McAllister, T.P. Developing guidelines and standards for disaster resilience of the built environment: a research needs assessment. NIST Technical Note 1795, National Institute of Standards and Technology, Gaithersburg, MD, 2013.

Ellingwood, Bruce; McAllister, Therese; van de Lindt, John. "The Science Behind Community Resilience Assessment: An Overview of the Center for Risk-Based Community Resilience Planning." Paper presented at 12th International Conference on Structural Safety & Reliability, Vienna, Austria. August 6, 2017 - August 10, 2017.

- McAllister, T.P. Research needs for developing a risk-informed methodology for [3] community resilience. J. Struct. Engrg. ASCE 142(8):C4015008, 2016.
- [4] Gilbert, S. Disaster resilience: a guide to the literature. NIST Special Publication 1117, National Institute of Standards and Technology, Gaithersburg, MD, 2010.
- [5] National Academies. *Disaster resilience: a national imperative – summary*. Committee on Science, Engineering and Public Policy, The National Academies, 2012 (www.nap.edu/html/13457/13457 summary.pdf)
- [6] Adapting infrastructure and civil engineering practice to a changing climate. American Society of Civil Engineers, Reston, VA, 2015.
- [7] Bruneau, M. S., Chang, S., Eguchi, R., Lee, G., O'Rourke, T., Reinhorn, A.M. Shinozuka, M., Tierney, K., Wallace and Wintervelt, D.V. A framework to quantitatively assess and enhance the seismic resilience of communities. Earthquake Spectra 19:733-752, 2003.
- Chang, S.E. and Shinozuka, M. Measuring improvements in the disaster resilience of [8] communities. Earthquake Spectra 20:739-755, 2004.
- [9] Cimellaro, G. P., A.M. Reinhorn, and M. Bruneau. Framework for Analytical Quantification of Disaster Resilience. Eng. Struct., 32(11); 3639-3649, 2011.
- [10] Mieler, M., Stojadinovic, B., Budnitz, R., Comerio, M., and Mahin, S. A framework for linking community-resilience goals to specific performance targets for the built environment. Earthquake Spectra, 31(2015): 1267-1283, 2015.
- [11] Cox, R. and Perry. K-M. Like a Fish Out of Water: Reconsidering Disaster Recovery and the Role of Place and Social Capital in Community Disaster Resilience. American Journal of Community Psychology 48:395-411, 2011.
- [12] Van Zandt, S., Peacock, W., Henry, D., Grover, H., Highfield, W., & Brody, S. Mapping social vulnerability to enhance housing and neighborhood resilience. Housing Policy Debate, 22(1), 29-55, 2012.
- [13] Lee, J.Y. and Ellingwood, B.R. A decision model for intergenerational life-cycle risk assessment of civil infrastructure exposed to hurricanes under climate change. Rel. Engrg. & System Safety 159:100-107, 2016 (DOI:0.1016/j.ress.2016.10.022).
- [14] Presidential Policy Directive 21 (PPD 21). Critical infrastructure security and resistance. The White House, 2013 (www.whitehouse.gov.)
- [15] Rinaldi, S.M., Perenboom, J.P. and Kelly, T.K. Identifying, understanding and analyzing critical infrastructure interdependencies. IEEE Control Systems 21(6):11-25, 2001.
- [16] Critical Assessment of Lifeline System Performance: Understanding Societal Needs in Disaster Recovery. *Report GCR 16-917-39*, prepared by the Applied Technology Council, Redwood City, CA. National Institute of Standards and Technology, Gaithersburg, MD, 2016.
- [17] Ellingwood, B.R. Assessment and mitigation of risk from competing low-probability, high-consequence hazards. Australian J. of Struct. Engrg. 9(1):1-8, 2009.
- [18] Ouyang, M., and Dueñas-Osorio, L. Time-dependent resilience assessment and improvement of urban infrastructure systems. Chaos: An Interdisciplinary Journal of Nonlinear Science 22(3): 033122, 2012.
- [19] Bocchini, P., Frangopol, D. M., Ummenhofer, T., and Zinke, T., Resilience and sustainability of civil infrastructure: toward a unified approach., J. Infrastruct. Syst. 20: 04014004, 2014,
- [20] Zhang, W. and Wang, N., Resilience-Based Risk Mitigation for Road Networks. Structural Safety 62: 57-65, 2016.

Ellingwood, Bruce; McAllister, Therese; van de Lindt, John. "The Science Behind Community Resilience Assessment: An Overview of the Center for Risk-Based Community Resilience Planning." Paper presented at 12th International Conference on Structural Safety & Reliability, Vienna, Austria. August 6, 2017 - August 10, 2017.

- [21] Guidotti, R., Chmielewski, H., Unnikrishnan, V., Gardoni, P., Therese McAllister and van de Lindt, J.W. Modeling the Resilience of Critical Infrastructure: The Role of Network Dependencies. Sustainable and Resilient Infrastructure 1(3):153, 2016.
- [22] Lin, P., Wang, N. and Ellingwood, B.R. A risk de-aggregation framework that relates community resilience goals to building performance objectives. Sustainable and Resilient Infrastructure, 1(1):1-13, 2016
- [23] Ellingwood, B.R., van de Lindt, J.W. and McAllister, T.P. Developing measurement science for community resilience assessment. Sustainable and Resilient Infrastructure 1(3):93, 2016.
- [24] Attary, N., van de Lindt, J.W., Unnikrishnan, V.U., Barbosa, A.R. and Cox, D.T. Methodology for development of physics-based tsunami fragilities. J. Struct. Engrg. ASCE (http://dx.doi.org/10.1061/(ASCE)ST.1943-541X.0001715), 2016.
- [25] Steelman, J. and Hajjar, J.F. Capstone scenario applications of consequence-based risk management for the Memphis Testbed. Report of the Mid-America Earthquake Center, University of Illinois at Urbana-Champaign, 2008.
- [26] Xue, X., Wang, N., Ellingwood, B.R. and Zhang, K. The impact of climate change on riverine flooding at a community scale. Proc., ICOSSAR 2017, Vienna, Austria, 2017.
- [27] Kuligowski, Erica D. Lombardo, Franklin T. Phan, Long T. Levitan, Marc L. Jorgensen, David P. Technical Investigation of the May 22, 2011, Tornado in Joplin, Missouri. National Institute of Standards and Technology (NIST). NCSTAR 3. Gaithersburg, MD, 2014. (http://dx.doi.org/10.6028/NIST.NCSTAR.3)
- [28] C. Standohar-Alfano and J.W. van de Lindt. An Empirically-Based Probabilistic Tornado Hazard Analysis of the U.S. using 1973-2011 Data., ASCE Natural Hazards Review, 16 (1), 04014013, 2015. (10.1061/(ASCE)NH.1527-6996.0000138)
- [29] Unnikrishnan, V.U. and J.W. van de Lindt. Probabilistic Framework for Performance Assessment of Electrical Power Networks to Tornadoes. Journal of Sustainable and Resilient Infrastructure, 1:3-4, 137-152, 2016.
- [30] Cutter, S.L. Burton, C.G., and Emrich, C.T. Disaster resilience indicator for benchmarking baseline conditions. J. Homeland Security and Emergency Management, 7(1): Article 51, 2010.
- [31] Xiao,Y., and Van Zandt, S. Building community resiliency: Spatial links between household and business post-disaster return. Urban Studies, 49(11), 2523-2542, 2012.
- [32] Rose, A., "Resilience and sustainability in the face of disasters," Environ. Soc. Trans 1:960-100, 2011.
- [33] Francis, R., and Behailu, B. A metric and frameworks for resilience analysis of engineered and infrastructure systems. Reliability Engineering & System Safety 121, 90-103, 2014.
- [34] Lounis, Z., and McAllister, T. P. Risk-based decision making for sustainable and resilient infrastructure systems. J. Struct. Engrg, ASCE 142: F4016005, 2016.

Ellingwood, Bruce; McAllister, Therese; van de Lindt, John. "The Science Behind Community Resilience Assessment: An Overview of the Center for Risk-Based Community Resilience Planning." Paper presented at 12th International Conference on Structural Safety & Reliability, Vienna, Austria. August 6, 2017 - August 10, 2017.

IDETC201768457

A Collaborative Data Management System for Additive Manufacturing

Yan Lu, Paul Witherell, and Alkan Donmez

National Institute of Standards and Technology, Gaithersburg, Marvland 20899 Email: [yan.lu, paul.witherell, alkan.donmez]@nist.gov

ABSTRACT

As additive manufacturing (AM) continues to mature as a production technology, the limiting factors that have hindered its adoption in the past still exist, for example, process repeatability and material availability issues. Overcoming many of these production hurdles requires a further understanding of geometryprocess-structure-property relationships for additively manufactured parts. In smaller sample sizes, empirical approaches that seek to harness data have proven to be effective in identifying material process-structure-property relationships. This paper presents a collaborative AM data management system developed at the National Institute of Standards and Technology (NIST). This data management system is built with NoSQL (Not Only Structured Query Language) database technology and provides a Representational State Transfer (REST) interface for application integration. In addition, a web interface is provided for data curating, exploring, and downloading. An AM data schema is provided by NIST for an alpha release, as well as a set of data generated from an interlaboratory study of additively manufactured nickel alloy (IN625) parts. For data exploration, the data management system provides a mechanism for customized web graphic user interfaces configurable through a visualization ontology. As a collaboration platform, the data management system is set to evolve through sharing of both the AM schema and AM development data among the stakeholders in the AM community. As data sets continue to accumulate, it becomes possible to establish new correlations between processes, materials, and parts. The functionality of the data management system is demonstrated through the curation and querying of the curated AM datasets.

1 INTRODUCTION

Additive manufacturing (AM) is a process capable of building up complex shapes layer upon layer directly from a threedimensional digital representation. As a production technology, AM minimizes the time lag between design and realization of a part while enabling novel design opportunities. However, since both the shape and the material properties of parts are formed during the AM process, manufacturers and end users have difficulty in predicting whether parts will meet specifications (shape, surface and material properties) and ensuring repeatability, consistency, and reliability across different AM machines and locations. Qualification of additively manufactured parts is one of the most serious hurdles to the broad adoption of additive manufacturing [1]. A key step in overcoming these hurdles is to characterize AM materials and geometry--process-structure-property relationships.

To develop such relationships, sufficient information about feedstock materials, processes, machines, and final part performance tests must be collected and analyzed accordingly. Such an effort requires a well-populated database containing the necessary pedigreed information. This task is difficult and extremely expensive for any single company, especially given the limitation of currently available standard methods and protocols for AM feedstock material characterization and AM part property testing [2].

Several data system development efforts have been reported for AM. For example, the Senvol Database [3] provides researchers and manufacturers with open access to industrial AM machines and materials information. A search on the Senvol Database finds properties of most industrial additive manufacturing machines and materials, and information on their compatibility. Granta, a material information management technology provider, has collaborated in several European Framework Seven projects in the field of material information management. The results from those projects are not currently available for public access, and the underlying data structure ('schema') is proprietary to Granta users. The Materials Selection an Analysis Tool (MSAT) database led by Department of Defense (DOD) was built based on GRANTA:MI [4], which captured some data from NASA and it is available to approved users[5]. None of the existing AM databases are publicly accessible, and none provide a sufficient amount of reliable material property data for much of the necessary AM analytics. These two hurdles are significant barriers to adoption of this emerging transformative technology as a viable production alternative to traditional manufacturing methods, especially for small manufacturers.

To obtain the multitude of data sets rich enough for AM geometry-material-process-structure-property relationship identification, the full processing history of thousands of samples

Donmez, M; Lu, Yan; Witherell, Paul. "A Collaborative Data Management System for Additive Manufacturing." Paper presented at 37th Computers and Information in Engineering Conference (CIE), Cleveland, OH, United States. August 6, 2017 - August

9. 2017.

and parts must be monitored and captured. This creates a significant challenge for any individual organization. For a community, however, the matter becomes more realizable. The AM community expects that, with appropriate data driven modeling methods, heterogeneous data sets with a multitude of geometries, material types, and processes can be mined to develop empirical correlations between material properties, microstructure, part geometry and process parameters. Such advancements promise to speed up AM process, material, part qualification and to lead to new AM material discoveries.

This paper discusses our effort to develop a collaborative AM data management system. This data management system is built on the latest NoSQL database technology, providing data storage in clouds structured by an AM lifecycle data schema. The collaborative data management system provides a REST interface for application integration as well as a user-friendly web interface for data curating, exploring, and downloading. Regular users can explore the database through customized web GUIs. Authorized users can contribute to the database population by populating an agreed upon AM schema. Initially, an AM data schema is provided by NIST [6], while iterative improvements are planned through collaborative community development. The first section of this paper presents the overall collaborative AM data management system architecture. Then the second section introduces the initial data model designed for AM lifecycle data management. In the third section, data

curation methods are introduced, and a mechanism to develop customizable data exploration views is shown. The final section summarizes the paper, our current development status, and discusses future work.

2 ARCHITECTURE

The top requirements on the collaborative AM data management system are data curation and sharing by multiple types of stakeholders. The open database system will be freely accessible to the AM community including vendors, manufacturers, and researchers. Consistent information should be provided to all stakeholders. Changes made on the database should be validated and relevant stakeholders are notified of any changes. Active and prior schemas will be available to the AM community, accompanied with clear definitions and descriptions of the data attributes. Users are encouraged to participate in improvement and optimization of the common AM schema. To enable these, a friendly GUI is needed with an effective navigation and data browsing scheme. A standard Application Programming Interface (API) will be made available for integration with other applications. For example, loading data into data analytics software such as R and Matlab will help the AM community to develop process-structure-property relationships for AM materials. Finally, it is desired that the system, to the greatest extent possible, be developed on proven technology.



Figure 1: A collaborative data management system architecture

Based on the requirements, several conceptual designs were investigated, including those using commercially available software. However, to avoid the use of proprietary software technologies and proprietary data models, we turn to NIST's own Material Data Curation System (MDCS), a product of the Material Genome Initiative [7][10]. The core of MDCS, the MDCS Server, was identified as a foundational system on which to build our data management system. The NIST MDCS server provides a means for wrapping MongoDB, an open source NoSQL database, and capturing, sharing, and transforming data into a structured Extensible Markup Language (XML)-based format. The server allows users to curate data with user-selected templates encoded in XML. These templates provide data entry forms that can be saved in a MongoDB database to be queried, explored and exported as XML files.

Figure 1 shows the final design of our open access and open structure database system architecture. The system can be summarized as an aggregation of seven main components in five layers. Proceeding from the bottom of the figure up the components are as follows:

- 1. Layer 1 a MongoDB database: The document-based database stores AM material data and provides scalability for the accumulation of data sets. It can be deployed in a private or public cloud.
- Layer 2 NIST MDCS Server: The NIST MDCS server 2. wraps MongoDB by providing XML-based templates for data population. It provides basic functions for material data curation, exploration, query, and template composition. In addition, it provides a REST API for application integration.
- Laver 3 NIST AM Database Schema: The schema is 3. developed to model the data sets in the AM data management system. The schema template is uploaded to the MDCS server for data population. It can be modified in any XML tool, and active and prior schemas can also be viewed and downloaded by end users.
- 4. Layer 3 Data Navigation Configuration: To enable a user friendly and customizable GUI, an ontology based data exploration mechanism is employed for effective web-based data navigation and browsing by different end users. The navigation configuration file is a Resource Description Framework (RDF)/XML file. The entity annotations defined in the RDF/XML are used to describe specific views and create links to the database items that will automatically be queried. The navigation schema can be automatically updated when the data schema changes. Access to development of this function is limited to the developer role.
- 5. Layer 4 A NIST-hosted web portal: A web portal is developed for AM data curating, exploring, and downloading. AM material data, AM machine data, build data, and test data can be explored through a customizable navigation schema. In addition, an AM-specific data query capability is provided for searching data records. New data can be curated and added to the database. (This feature is limited to those with appropriate administrative rights.)

- 6. Layer 4 Data Analytics Application Integration: Using the REST application interfaces, data analytics capabilities provided by Matlab, R, and Python applications can be integrated into the collaborative data management system. Data analysis functions provided by the software can be fully utilized to analyze material properties and mine process-structure-property relationships.
- 7 Layer 5 - AM Build, Material, and Test Data: AM material data, AM machine data, and the complete process history of AM builds can be curated into the data management system through the web user interface or through software tools (e.g. customized Python applications). AM feedstock test data, AM simulation data, and data generated from AM testbeds can be captured into the data system as well. Fusion of these data sets will facilitate AM material characterization, process characterization, and part qualification.

The next several sections elaborate on the main functional components of the NIST collaborative data management system.

3 AM DATABASE SCHEMA

The collaborative data management system is designed for collaborations among AM stakeholders including manufacturers, material vendors, machine builders and the AM research community. The system should capture, store, and manage the data through the entire AM lifecycle and value chains, including part/specimen design, feedstock material data, AM machine data, process parameters, process monitoring and control data, part/material test data, and other pedigreed information. Each of these factors potentially contribute to qualifying the process, asbuilt parts, materials, and identifying process-structure-property relationships. For example, to highlight possible challenges, process parameters, the feedstock material handling, and even the design of the component can all cause differing thermal properties from layer to layer, leading to excessive internal stresses and build failures. Hence, in addition to feedstock material data, the full processing history of manufactured parts must be captured, stored, and linked in attempt to relate materials, designs, and process parameters with process outcomes. The individual AM machine calibration and maintenance history, as well as their geographic locations, can also contribute to variability of part properties.

The data structure, or schema, provides the backbone of the collaborative data management system. The relationships created by the structure are essential to supporting meaningful data curation and data retrieval. As the collaborative data management system matures through iterations, it is the data structure that will need to evolve to support the development of meaningful relationships that can be used to query and analyze the database.

As part design information and test data are linked to individual builds, we structure the data sets into four types of main entities,

Donmez, M; Lu, Yan; Witherell, Paul. "A Collaborative Data Management System for Additive Manufacturing." Paper presented at 37th Computers and Information in Engineering Conference (CIE), Cleveland, OH, United States. August 6, 2017 - August

9. 2017.

as shown in Figure 2. The first entity type "amMaterial" captures vendor material information and vendor provided material properties. The second entity type "amMaterialTest" captures feedstock material properties, e.g., flowability, spreadability etc., provided by the AM community. The third type of entity "amMachine" captures AM machine information. The data captured by these three types are independent of specific builds and can be provided by material vendors, material researchers

and machine owners. The fourth type of database entity "amBuild" captures the information related to a specific AM build, including part and specimen design, pre-process, inprocess and post process information, as well as test information. Figure 3 shows the data structures of an AM build record in an XML tree structure. These structures are representative of the initial data structure of the database.



Figure 3: AM Build XML schema.

AM Data Curation

Curation of pedigreed data about feedstock material, build process and built-material testing and results is done through a XML-based template. To ensure data integrity, the current instantiation of the curation system only allows data entry by users with appropriate credentials. Figure 4 shows the current design of the template used to instantiate the amMaterial entity in the web portal. An alternative entry mechanism is to curate

"A Collaborative Data Management System for Additive Manufacturing." Paper presented at 37th Computers and Information in Engineering Conference (CIE), Cleveland, OH, United States. August 6, 2017 - August

9, 2017.

Donmez, M; Lu, Yan; Witherell, Paul.

and edit data by using a POST request as provided by the MDCS REST API. The documents of the REST API are accessible through the web portal as well.



Figure 4: AM vendor provided material data curation.

4 DATA NAVIGATION CONFIGURATION

The functionality and ease-of-use of the data management system is heavily dependent on the abilities and appearance of the Navigation Pane. Custom navigation can help improve efficiency by grouping, highlighting, or hiding objects. The flexibility of the Navigation Pane allows us to continuously improve usability by making it easier for users to continue to identify the database objects they require as the schema evolves. As shown in Figure 5, The NIST collaborative AM data management system navigation Pane and Views can be customized through an ontology based configuration tool. Here, Protégé, a free, open-source platform is used to construct the collaborative data management system data navigation schemas. Figure 5 shows the current design of the data exploration view and the corresponding ontology configurations.

AM Material Database # Home C2 Curate Q Explore @ Vev			
Database	1	CarponterMicroMolt625 yml	
> AMBuilds			
> AMMachine		0	
Coronico		General Information	
✓ Metal		Vendor	Carpenter
> AluminumAlloy		Material ID	3213103-0003
> CobaltAlloy			
> CopperAlloy		Powder Properties	
✓ NickelAlloy			
✓ IN230		Powder Tap Density (g/cc)	5.1
✓ IN625		D10	 Value: 13.4
Micro-Melt@ 625AM		D50	 Value: 28.4
EOS NickelAlloy IN625		000	- Miles (2.0
✓ IN713		000	 value, 43.0
✓ IN718			
✓ IN738		Chemical composition in Percentage	9
✓ Ni-HX		Conformance	AMS 5000F
✓ Ni-K5		Carbon	Min: None
✓ Nitinol60			 Max: 0.1
PreciousMetalAlloy		Manganese	Min: None
> Steel			 Max: 0.5
Polymer		Sulfur	Min: None
AMTests			 Max: 0.015
> Chemistry		Chromium	 Min: 20
✓ Dimensional			 Max: 23
✓ HCF		Molybdenum	• Min: 8
B HCF_01			 Matc 10
HCF_10		Al	Min: None Min: 0.4
HCF_22	-		 max. 0.4

(a) Data Exploration view



Figure 5: Customizable AM data navigation

As shown in Figure 5(b), the class hierarchy defined in the ontology dictates the data navigation pane, as shown in Figure 5(a). The annotations of the leaf nodes as highlighted in the class hierarchy of the ontology define the document types to be searched for, the instances to be shown, and the information of the instances to be displayed. The annotation 'filter' constrains the search path. The annotation 'project' defines the instance to be listed and displayed in the left side navigation pane, and the annotation 'projection_view' defines the information to be displayed for all the instances.

Donmez, M; Lu, Yan; Witherell, Paul.

"A Collaborative Data Management System for Additive Manufacturing." Paper presented at 37th Computers and Information in Engineering Conference (CIE), Cleveland, OH, United States. August 6, 2017 - August The ontology is modeled in the RDF/XML format and uploaded to our data management system to configure the data exploration view. The navigation schema can be designed based on enduser's requests.

5 RESULTS AND FUTURE WORK

The NIST collaborative data management system has successfully gone through a series of functional tests and security tests both on a local server and in a cloud. Currently an instance of such a data management system is deployed on a NIST public server

The first batch of data was captured and curated in this server, including a set of round robin test data from a NIST MSAM (Measurement Science for Additive Manufacturing Program) [8] sponsored project. The project investigated four nickel alloy 625 builds produced on three different Laser Powder Bed Fusion machines in argon or nitrogen environments. Except for the machine serial number and build environment, all other conditions of the four builds were specified, including: build geometry, powder lot, process parameter sets, and heat treatment. Each build produced 13 coupons for tensile, high-cycle fatigue (HCF), low cycle fatigue (LCF) and metallurgy tests. The samples cut from mechanical test coupons include both XY build orientation and Z direction orientation. Figure 6 shows the coupon drawing.



Figure 6: NIST-EWI round robin test coupon drawings

All the build coupons, except those two for chemical analysis, received a stress relief at 1038°C, followed by a hot isostatic pressure thermal treatment at 1121°C. Test samples were cut from the post-processed coupons and mechanical tests were conducted by a third party lab. There are total 40 tensile test reports, 24 HCF reports, 24 LCF reports, two chemical analysis results, 4 hardness test reports, and 8 grain size analysis results available.

Large volumes of structured and unstructured data were generated during the processes, and in various formats, such as CAD models, STL files, excel data sheets and pdf documents. In addition, AM machine data was collected from the machine vendor. All of the stated data was captured in our data management system through manual curation. The first build data was manually entered though the web portal. The other three build data sets were edited on top of the first build data set offline and curated using the REST API. In total, the first set of data instantiation included one AM material document for Carpenter micro-melt 625, three AM machine documents capturing the information about three different EOS M 270 machines, and four AM build documents. The data is instantiated and can be explored based on current navigation schema, as shown in Figure 5 (a).

At its initial release, our system aims to provide a mechanism that will allow AM users to view and download the NIST schema as well as view the first batch of data sets. Its setup allows the AM community to provide feedback on the suitability of the data types, definitions, content, and structure. As the collaborative data management system is meant to support changes in needs and priorities, the tool will easily be able to update schema variations while preserving the previous schema in its archives. While our initial success was based on the demonstration of remote access and data retrieval, we have also demonstrated that the direct integration of Matlab with the AM data management system through the RESTful API simplified any repeatability analysis of the NIST Round Robin data.

After initial Alpha and Beta releases, the next phase of our open data effort will incorporate vetted AM data submitted by the AM community. We look to expand the open data management into a comprehensive collaboration platform, and provide an additional analytical framework for data management with data analytics tools and simulation engines. This effort aligns with NIST's goal for open data to expand public access to NIST research results while maintaining the quality and integrity of NIST data resources. Progress and updates on the NIST-hosted AMMD data management system can be found at https://ammd.nist.gov. The locally-downloadable version of the AMMD data management system software can be found at https://github.com/usnistgov/SIAM-Project.

ACKNOWLEDGEMENT

We thank Mr. Phillippe Dessauw and Mr. Alden Dima for the support on the system development, Dr. Zachary Trautt for his help on the introductions of MDCS server and our two summer interns, Nathan Brockett and Tawsif Rafid, for their contributions to data population.

DISCLAIMER

Certain commercial equipment, or materials, suppliers, or software are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

9. 2017.

Donmez, M; Lu, Yan; Witherell, Paul. "A Collaborative Data Management System for Additive Manufacturing." Paper presented at 37th Computers and Information in Engineering Conference (CIE), Cleveland, OH, United States. August 6, 2017 - August

REFERENCES

- [1]. Wing, I., Gorham, R., and Sniderman, B., 3D opportunity quality assurance and parts qualification, for http://dupress.com/articles/3d-printing-quality-assurancein-manufacturing/
- [2]. Characterization of Additive Manufacturing Materials, http://www.nist.gov/el/isd/sbm/camm.cfm
- [3]. Senvol Database, http://senvol.com/database/
- [4]. GRANTA MI: Additive Manufacturing Package, https://www.grantadesign.com/products/mi/am.htm
- [5]. Material Tool, Selection and Analysis http://maptis.nasa.gov/FeaturesMSAT.aspx
- [6]. Lu, Y., Choi, S., Witherell, P., 2015. "Towards an integrated data schema for additive manufacturing conceptual modeling". ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Boston, MA.
- [7]. Materials Data Curation System, https://mgi.nist.gov/materials-data-curation-system
- [8]. Measurement Science for Additive Manufacturing, http://www.nist.gov/el/isd/sbm/msam.cfm

National Science and Technology Council, Materials Genome Initative for Global Competitiveness, Office of Science and Technology Policy, Washington DC; 2011

- [9]. Kelly, S., "Development and Measurement Analysis of Design Data for Laser Powder Bed Fusion Additive Manufacturing of Nickel Alloy 625 Final Technical Report", https://ewi.org/eto/wpcontent/uploads/2015/04/70NANB12H264_Final_Tech_R eport_EWI_53776GTH_Distribution_Vol_1.pdf
- [10]. Dima, A., Bhaskarla, S., Becker, C. et al. "Informatics Infrastructure for the Materials Genome Initiative", JOM (2016) 68: 2053. doi:10.1007/s11837-016-2000-4

Toward Semi-Autonomous Information Extraction for Unstructured Maintenance Data in Root Cause Analysis

Michael Sharp, Thurston Sexton, and Michael Brundage

National Institute of Standards and Technology Systems Integration Division, 100 Bureau Dr, Gaithersburg, MD 20899, USA {michael.sharp,thurston.sexton,michael.brundage}@nist.gov https://www.nist.gov/el/systems-integration-division-73400

Abstract. To facilitate root cause analysis in the manufacturing industry, maintenance technicians often fill out "maintenance tickets" to track issues and corresponding corrective actions. A database of these maintenance-logs can provide problem descriptions, causes, and treatments for the facility at large. However, when similar issues occur, different technicians rarely describe the same problem in an identical manner. This leads to description inconsistencies within the database, which makes it difficult to categorize issues or learn from similar cause-effect relationships. If such relationships could be identified, there is the potential to discover more insight into system performance. One way to address this opportunity is via the application of natural language processing (NLP) techniques to tag similar ticket descriptions, allowing for more formalized statistical learning of patterns in the maintenance data as a special type of short-text data. This paper showcases a proof-of-concept pipeline for merging multiple machine learning (ML) and NLP techniques to cluster and tag maintenance data, as part of a broader research thrust to extract insight from largely unstructured natural-language maintenance logs. The accuracy of the proposed method is tested on real data from a small manufacturer.

Introduction 1

Multiple industries often use root cause analysis (RCA) techniques to diagnose the underlying cause(s) of problems (Shrouti et al., 2013). Within the manufacturing industry, there are a variety of RCA techniques that are utilized: Six Sigma, including DMAIC (Define, Measure, Analyze, Improve, Control) and DFSS (Design-For-Six-Sigma) (BOUTI and KADI, 1994), Failure Mode and Effect Analysis (FMEA) (Liu et al., 2013), and fishbone diagrams, also known as Ishikawa diagrams (Juran and Godfrey, 1999) are just a few. While there are many techniques, instances of RCA are often problem-specific studies, where results are not readily available for wide retrieval in future studies. A framework was developed previously in Brundage et al. (2017) to help alleviate this issue,

 $\mathbf{2}$ Information Extraction of Maintenance Logs for Diagnostic Analysis

providing mechanisms for accessing previous problems to aid in diagnosing the root cause. However, the developed framework relies on readily structured descriptions of causes-effects-treatments; generally such patterns are derived from raw information tracked via a Computerized Maintenance Management System (CMMS). Such clearly structured information is rarely found in practice, as technicians often inconsistently record informal prose rather than clearly filling in discrete fields for causes, effects, and treatments. Such inconsistencies make it difficult to perform diagnosis procedures. This paper begins to address that issue by providing NLP and ML techniques to prepare, clean, and tag the data for use in the diagnosis framework. It is aimed at cases where a CMMS may not be properly implemented, or in a well managed CMMS to help capture in-explicit information from any free form descriptions or comments within the system.

2 Motivation

Using NLP techniques in a maintenance data-set, unlike the more popular applications of NLP that have huge amounts of casual and/or complete-sentence phrases (such as Yelp[®] reviews or a Twitter[®] feed), requires treatment of documents generally smaller in nature, which at times have entries with fragmented sentence structure or are written in domain-specific shorthand. In addition, to the authors' knowledge, no definitive corpus or thesaurus of maintenance log terms and terminology currently exists that spans all companies in an industry setting. Nor, in many cases, would one be appropriate, as each company even each work site — will often develop their own short hand vocabulary and "tribal knowledge" set that could be meaningless to anyone not immersed in that environment. This work seeks to provide a method for working within these environments in order to characterize and categorize often dissimilar entries.

The authors envision a mature information extraction tool as autonomously tagging and structuring extracted information from the short, often fragmented entries that are common characteristics of many industrial maintenance logs. Achieving such vision would require both additional data sets, as well as comparative analysis with a wider breadth of existing NLP and ML techniques. The following is presented as a preliminary proof of concept (PoC), articulating the basic road map shown in Fig 1. The result of this PoC expresses the viability of implementing computer-augmented maintenance history analysis.

Methodology 3

Solutions to classification problems are generalized based on their method of training: supervised, semi-supervised, or unsupervised. This paper uses a supervised classification, meaning that pre-defined ground truth labels must be assigned to a training set by an "expert", and a trained model can then be used to predict labels for previously unseen entries. The data set used in this work consists of 779 hand-labeled entries from a manufacturing company's actual maintenance log over the period of several months.

3



Fig. 1. An overview of the information extraction pipeline for maintenance log text, as used in this work.

3.1 Process Overview

Using text-mining methods in general, and specifically NLP as implemented here, requires several components:

- 1. Data collection and labeling
- 2. Text preprocessing and language embedding (vectorization)
- 3. Classification model training and validation

This process is summarized briefly in Fig. 1. The data set of 779 entries was manually labeled from short form problem descriptions to train a classifier (Sec. 3.2). To represent the logs numerically, vectorization was done using a version of Topic modeling closely related to Latent Semantic Analysis (Sec. 3.3). Finally, a Support Vector Machine (SVM) was trained to label documents based on the provided training set (Sec. 3.4).

3.2 Data Collection and Labeling

Each of the 779 entries was tagged with ground-truth "effect" labels via an "expert human classifier". During this "data labeling" step, the expert was tasked with tersely describing the observed "effect" of a problem recorded in the original log. For example:

Log Text Conveyor belt between machine and wash conveyor worn Effect Label Conveyor Belt Failure

A total of 352 unique target "effect" labels were given by the expert, with the most common label occurring 39 times. A significant number of entries, 270 entries, have a unique label that is not repeated within the data set. To aid in classification quality for the more "important" effects, those with less than 4 occurrences, are re-labeled as "miscellaneous" or "uncharacterized". Placing these "uncharacterized" labels in the training set helps to solidify the model's recognition of high-frequency effects, and avoid prediction in areas of high uncertainty. In effect, one could interpret an uncharacterized label as any entry that is "not significantly represented" within the model¹. Using a vector of zeros

 $^{^{1}}$ This may be viewed as equivalent to a form of outlier detection (or noise filtering), where the labels used frequently by the expert are considered "useful" for classification of future problems.

4 Information Extraction of Maintenance Logs for Diagnostic Analysis

to represent these "not significantly represented" labels improves classification accuracy of identified labels by around 50%, while also providing well-defined identification of labels either unlike any in the training set, or those that the model is highly uncertain about. Any new or uncertain labels identified by the model could then be delivered to a user for better expert identification, and subsequent inclusion into a retrained model.

Text Preprocessing and Vectorization 3.3

To train a classifier on the labeled data, a numeric representation of the textbased maintenance logs (the documents) is needed. This semantic language embedding is done in two steps: 1) preprocess the text to cleanse it of non-useful artifacts, and 2) vectorize the processed text into manageable form by using a topic model.

Raw-text Preprocessing Cleaning each query (i.e. each issue as described in a string of words) by removing extraneous punctuation and inconsequential words (a.k.a. stop words) (Leskovec et al., 2014) is common practice in NLP techniques. After cleaning, the string is parsed into words and phrases up to N words long and placed in sparse word frequency matrix via the Bag of Words (BoW) technique. During the construction of this matrix, common pluralizations of words are combined (i.e. treated as a single entry) and tokens that have a very low occurrence rate within the corpus (less than 3 instances) are removed. In this context, a token is a word, or group of ordered words (phrase) that appear in the corpus. This dimensionality reduction aids both in convergence and processing time for the language embedding and classification algorithms.

Language Embedding To create a numeric feature space, useful for computerized classification, this work loosely follows the process used in Latent Semantic Analysis (Dumais, 2004). The entire set of text (the *corpus*) is represented as a term-document Matrix, a frequency-based vectorization of word occurrence (BoW, as referred earlier).

To reduce the dimensionality of this corpus word-frequency feature space, a reduced-rank Principal Component Decomposition was performed, and the top n-largest principal components were retained so that the rank-n approximation had 90% variance retention. One interpretation of each of the n the principal component vectors is as a weighted combination of words/phrases forming a topic in the corpus. Thus the name, "Topic Model".

Finally, due to very low term-overlap between maintenance topics within this corpus, the document vectors were weighted by the most common token found elsewhere in the corpus. This was done with an ordinary-least-squares (OLS) mapping, essentially a prediction on how a phrase would be most commonly talked about in the rest of the corpus. The authors hypothesize that this might take advantage of the natural structure of maintenance-like data, and preliminary results suggest accurate performance under this weighting scheme.

3.4 Classification: SVM Model

The last phase of the presented method is to train a supervised classifier on the expert diagnostic tags provided for each training entry. Here, a Support Vector Machine (SVM) classifier was selected for this task, which has been previously shown to have excellent ability in text classification (Joachims, 1998).

Data is randomly split into train- and test-groups, allowing the SVM to learn on a subset of the data, and then be validated by predicting labels on another, smaller subset (see Sec. 4.1 for a discussion on the effect of training-set size). The classifier was trained against binary vectors representing the target "ground truth" label. Upon subsequent input, it outputs a relative likelihood for each of the possible category labels, and assigns the most probable label to that input.

As a modification to the default labeling, if none of the potential labels had a relative likelihood 0.5 (50% likely) or more, the input was deemed too far away from previous training vectors and therefore an "uncharacterized set entry". This "uncharacterized set" labeling is crucial for identifying labels that are not well characterized in the model so that they can be further analyzed by the expert and retrained into the model, when Human-in-the-Loop training is possible.

4 Results

The results presented in this section enumerate classification performance averaged over twenty trials where the designation of training, testing, and validation samples are randomized between tests. The influence of training set size on performance is detailed, though this is recognized as only a small subset of controllable parameters in this model. A broad-sensitivity study, though outside the scope of this work, is a crucial part of future work (Sec. 5).



Fig. 2. The effects on classification accuracy of the fraction of data used for training the SVM. Mean number of correct and incorrect labellings over 20 trials are shown against the lower axis for each training-fraction setting. Additionally, the overall prediction accuracy is shown against the upper axis with 1σ uncertainty bars.

Information Extraction of Maintenance Logs for Diagnostic Analysis

4.1Training Set Size

6

To determine the robustness of the described method to the availability of training data, the SVM was trained on varying proportions of the data selected at random (see Fig. 2). Each increase in the proportion of training data is analogous to a human manually tagging entries marked by the algorithm as unknown, and/or correcting any observed mislabeling, then iteratively retraining the model. In this way we can simulate Human-in-the-Loop training. Note that as this data-set is finite in size, when an increasing amount (e.g. 80%) is used for training, the total number of validation instances or examples reported in the results must drop (e.g. to 20%) corresponding to the total available entries.

For this work, an "uncharacterized set entry" is a label that has too few (or zero) examples in the training set to be confidently characterized by the classification model. Intuitively, as the percentage of training examples increases, this number of "uncharacterized set entries" will likely decrease somewhat. However, for this data-set, and likely any coming from real world industry maintenance logs, there is a base level of these rarely occurring entries. Thus as important as it is to correctly classify or label those entries that can be characterized by the model, it is equally important to identify those that cannot. This not only lowers misclassification chances, it also allows such entries to be flagged for further investigations by an external human operator, who can then correctly label and add it to the model.

It is important to note that the classifier's ability to correctly identify welldocumented labels (i.e. ones not in the "uncharacterized set") is relatively invariant across training-set sizes. This implies that, once the model has learned which labels are reliable, its confidence level in predicting them — and by extension, its performance — is likely to remain consistent, even as new information is added to the model. In other words, these results imply that once any arbitrary label obtains enough entries to be able to be characterized by the model, the expected correct classification rate of that label is consistent, regardless of what the label actually is. Thus adding additional labeled examples is more about extending the model coverage than improving performance; although intuitively performance should also somewhat improve.

In addition, the model is able to correctly isolate the "uncharacterized set entries" not previously seen in the training data, but this ability decreases with a broader selection of labeled entries in the training set. This indicates that, as more of the feature space is populated, an entry must be further away from any known labels to be identified as unknown. Consequently, a dynamic, rather than static threshold of confidence when classifying "uncharacterized set entries" as outliers may be more appropriate for this type of discrimination. Lowering the confidence criteria for labeling "uncharacterized set entries" from 50% to 40% in one of the tests caused an average increase of 3% in correct classification, but also a drop of 6% in the correct identification of "uncharacterized set" labels. Significant improvement might come from re-defining the confidence threshold on-the-fly, or defining it in a feature-specific manner.

7

As shown in Fig. 2, the increased *total* accuracy at lower levels of training data is driven by the model's increased ability to identify unknown entries. This should not lead one to assume that less training data is a favorable; rather, this reaffirms the need to accurately identify the "uncharacterized set entries" at higher proportions of training data, where the total feature space becomes more populated. As the well-characterized areas-the "known" areas of the feature space-become more dense, the "uncharacterized set" areas become harder to distinguish. Additional investigations into methods for managing and mitigating this effect will be part of on going work, perhaps utilizing more crisp kernel models in the classification algorithm, or other similar techniques

Additional investigations on n-gram parsing (Brown et al., 1992) and rare token exclusion have been performed, but a full review is left out due to space constraints. In brief, it was found that additional complexity added via n-gram parsing was unnecessary, likely due to the domain-targeted language of the data set. Conversely, regulating the minimal word token occurrence frequency did seem to have a significant effect on the model's performance. Removal of highly superfluous or infrequent terms aided classification through dimensionality reduction with indications that, for a given data set, there is an optimal occurrence frequency band to include in the model. Further investigations into the generalization to a broader range of data sets of these findings as well as additional areas of inquiry are left for future work.

$\mathbf{5}$ Conclusions & Future Work

Virtual mountains of historic maintenance logs representing an untold wealth of diagnostic knowledge exist throughout industry. Without proper tools and techniques to analyze and contextualize that data, the usefulness of these maintenance logs is severely limited. Presented in this work is a proof-of-concept algorithmic framework for characterizing one aspect of that data. By categorically labeling the generally free form and fragmented text patterns associated with industrial maintenance logs, historical commonalities and recurring problem areas can readily be identified and targeted for process improvement.

The methodology detailed in this work is shown on a preliminary case study to consistently categorize and label a free form maintenance log entry from a set of known labels with over 70% accuracy. Additionally, the algorithm can correctly identify log entries as unique (or potentially needing better labeling) with over 85% accuracy.

Work in this area is a fertile ground for many avenues of continuing and future research. Especially apparent is the need for a broad overview of available methods for training and classifying natural-language texts in the form of maintenance logs. Automated selection of an optimal model for prediction of labels in a given industry or use case is crucial to ensure the best performance. Comparing other quantitative representations for language, like Word2Vec semantic embedding, will provide an excellent means of discovering maintenance-specific patterns in the text logs. In addition, the efficient utilization of domain-expert

8 Information Extraction of Maintenance Logs for Diagnostic Analysis

knowledge will be crucial in implementing systems like this one, leading to a more dynamic ability to parse data with Human-in-the-Loop system schemes.

The authors believe a set of guidelines for selecting appropriate algorithms based on amount and quality of data, as well as the desired outputs, could accelerate maintenance information utilization. Taking advantage of information hidden in maintenance logs could help bolster productivity, improve maintenance practices, and ultimately save time and money wasted on patching trivial symptomatic problems instead of focusing on the root cause.

References

- BOUTI, A., KADI, D.A.: A state-of-the-art review of fmea/fmeca. International Journal of Reliability, Quality and Safety Engineering 01(04), 515–543 (1994), http://www.worldscientific.com/doi/abs/10.1142/S0218539394000362
- Brown, P.F., Desouza, P.V., Mercer, R.L., Pietra, V.J.D., Lai, J.C.: Class-based n-gram models of natural language. Computational linguistics 18(4), 467-479 (1992)
- Brundage, M.P., Kulvantunyou, B., Ademujimi, T., Rakshith, B.: Smart manufacturing through a framework for a knowledge-based diagnosis system. In: Proceedings of the ASME 2017 International Manufacturing Science and Engineering Conference, MSEC2017, June 4-8, 2017, University of Southern California, Los Angeles, CA (2017)
- Dumais, S.T.: Latent semantic analysis. Annual Review of Information Science and Technology 38(1), 188-230 (2004), http://dx.doi.org/10.1002/aris. 1440380105
- Joachims, T.: Text categorization with support vector machines: Learning with many relevant features. In: European conference on machine learning. pp. 137–142. Springer (1998)
- Juran, J., Godfrey, A.B.: Quality handbook. McGraw-Hill (1999)
- Leskovec, J., Rajaraman, A., Ullman, J.D.: Mining of massive datasets. Cambridge university press (2014)
- Liu, H.C., Liu, L., Liu, N.: Risk evaluation approaches in failure mode and effects analysis: A literature review. Expert Systems with Applications 40(2), 828 - 838 (2013), http://www.sciencedirect.com/science/article/pii/ S0957417412009712
- Shrouti, C., Franciosa, P., Ceglarek, D.: Root cause analysis of product service failure using computer experimentation technique. Procedia CIRP 11, 44 - 49 (2013), http://www.sciencedirect.com/science/article/pii/ S221282711300543X

An improved method for calibrating solid-phase microextraction by direct loading

Mengyan Gong^{1,*}, Dustin Poppendieck¹

¹Indoor Air Quality and Ventilation Group, Engineering Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA

*Corresponding email: mengyan.gong@nist.gov

KEYWORDS

SPME, semivolatile organic compounds, tris(2-chloro-1-methylethyl) phosphate, spray polyurethane foam

1 INTRODUCTION

Solid-phase microextraction (SPME) has been widely used as a sampling technique for a variety of organic compounds due to its advantages of small sampling volume, ease of sampling, and simple injection (Ouyang and Pawliszyn 2008). Recently, Cao et al. (2016) developed an improved method for measuring the emission parameters for semivolatile organic compounds (SVOC), which shortened experimental times. In the method by Cao et al. (2016), the SVOC equilibrium concentration in a sealed chamber was measured using SPME sampling. Hence, calibration of SPME is required, which requires that the absolute amount of SVOC adsorption onto the SPME fiber be measured. To do this, Cao et al. (2016) calibrated SPME using liquid splitless injection and assumed that the SPME injection and liquid splitless injection have the same transfer efficiency. However, it is known that the transfer efficiency for both SPME injection and splitless injection are influenced by factors such as the type of liner, temperature of injector, and cross-sectional area of the space between column and liner, so that transfer efficiencies could be significantly different for the two injection techniques (Ouyang, Chen et al. 2005). Hence, calibrating SPME using liquid splitless injection may cause large uncertainties. Among the loading techniques that have been used for SPME, direct syringe-fiber loading has been shown to have transfer efficiencies greater than 95 % for chemicals with low volatility (Zhao, Ouyang et al. 2007). The present study aims to first investigate the applicability of the syringe-fiber loading approach for calibrating SPME, and then apply it to experiments measuring the Tris(2-chloro-1-methylethyl) Phosphate (TCPP) partition coefficient between spray polyurethane foam (SPF) and air based on the method by Cao et al. (2016).

2 METHODOLOGIES

The structure of SPME is shown in Figure 1. Instead of spiking the liquid calibration solution onto the fiber as in the previous loading method (Zhao, Ouyang et al. 2007), we spiked the liquid onto the stainless steel rod just above the fiber to avoid damaging the easily breakable SPME fiber (100 μ m in diameter). In addition, to eliminate the influence of changing the gas chromatograph-mass spectrometer (GC-MS) response, the internal standard was also spiked onto the stainless-steel rod. Here, we took TCPP as an example to evaluate this method. First, a series of experiments were run to examine the influence of evaporation loss during spiking by exposing the SPME stainless steel in a fume hood for different times. As shown in Figure 2, 1 μ L TCPP liquid and 1 μ L internal standard (triamyl phosphate, TAP) were spiked onto the end of the SPME stainless steel rod after pushing the fiber out. Experiments were run to examine the TCPP loss from SPME fiber when it is exposed to the surrounding fume hood air during the process of spiking TAP. Third, standard curves were run to calculate the SPME adsorption amount for different sampling times. Finally, the partition coefficient between TCPP in SPF and air was calculated based on the method described in Cao et al.(2016).
3 RESULTS AND DISCUSSION

Figure 4a shows the response ratio of two TCPP isomers (TCPP1 and TCPP2) and TAP (internal standard) when the TCPP standard was evaporated from the stainless-steel rod for different times. The standard deviations of the ratio for standards with both high and low TCPP concentration is less than 10 %, except for TCPP2 at lower concentration, which is 15 %. Hence, the TCPP loss from the stainless-steel rod during spiking was minimal. Figure 4b shows the response ratio of TCPP and TAP when SPME samples were first taken for 5 min and then the fiber was exposed to surrounding air for different times. The standard deviation for TCPP1 and TCPP2 is 6 % and 11 %, respectively, which indicates that the TCPP loss from the SPME fiber during spiking was also minimal. The linear correlation coefficients of the standard curve with TCPP concentration in SPF we measured in other experiments (8.3 x 10¹¹ µg/m³ and 2.1 x 10¹¹ µg/m³ for TCPP1 and TCPP 2), the calculated partition coefficient between SPF and air at 25 °C is 1.0 x 10⁷ and 1.2 x10⁷ for TCPP1 and TCPP2, respectively.







Figure 2. Illustration of direct loading of calibration liquid onto SPME stainless steel rod



Figure 3. SPME sampling from a sealed chamber (a) chamber; (b) sampling



Figure 4. Evaluation of TCPP loss during spiking (a) TCPP standard, lower ratios for 0.1 ng/µL standard, higher ratios for 1 ng/µL standard; (b) SPME fiber



Figure 5. SPME adsorption amount of TCPP

4 CONCLUSIONS

The improved method for calibrating SPME by direct syringe-stainless steel rod loading for TCPP has been preliminarily validated. In addition, the method has been successfully used for measurement of the TCPP partition coefficient between SPF and air. However, more tests for

TCPP standards at higher concentrations and the applicability of this method for chemicals with different physical properties are needed to completely evaluate this method.

REFERENCES

Cao, J., X. Zhang, J. C. Little and Y. Zhang (2016). "A SPME - based method for rapidly and accurately measuring the characteristic parameter for DEHP emitted from PVC floorings." <u>Indoor air</u>.

Ouyang, G., Y. Chen, L. Setkova and J. Pawliszyn (2005). "Calibration of solid-phase microextraction for quantitative analysis by gas chromatography." Journal of chromatography A **1097**(1): 9-16.

Ouyang, G. and J. Pawliszyn (2008). "A critical review in calibration methods for solid-phase microextraction." <u>Analytica chimica acta</u> **627**(2): 184-197.

Pawliszyn, J., B. Pawliszyn and M. Pawliszyn (1997). "Solid Phase Microextraction " <u>The Chemical</u> <u>Educator</u> **2**(4): 1-7.

Zhao, W., G. Ouyang and J. Pawliszyn (2007). "Preparation and application of in-fibre internal standardization solid-phase microextraction." <u>Analyst</u> **132**(3): 256-261.

Proceedings of ASME 2017 International Mechanical Engineering Congress & Exposition IMECE 2017 Nov 3–9, 2017, Tampa, FL, U.S.A.

IMECE2017-70404

TOLERANCES AND UNCERTAINTY IN ROBOTIC SYSTEMS

F. Proctor, Marek Franaszek, and J. Michaloski National Institute of Standards

> and Technology Gaithersburg, Maryland

ABSTRACT

The ability to be programmed for a wide range of tasks is what differentiates robots from dedicated automation. Consequently, robots can be faced with often-changing requirements and conditions. Conventional application development based on teach programming takes robots out of production and occupies personnel, limiting robots' effectiveness in these environments. Off-line programming solves these problems, but robot inaccuracy must be compensated by a combination of calibration, compliance, and sensing. This complicates up-front systems engineering and application development, but results in systems that can operate in a wider range of requirements and conditions. Performance can be optimized if application tolerances and process uncertainties are known. If they often change, optimization must be done dynamically. Automating this optimization is a goal of smart manufacturing. With its trend of increasing connectivity between the components of robotic systems both within workcells and to the enterprise, exchanging this information has become more important. This includes tolerance information from design through process planning to production and inspection, and measurement uncertainty from sensors into operations. Standards such as ISO 10303 (STEP), the Quality Information Framework (QIF), the Robot Operating System (ROS), and MT-Connect support this exchange to varying degrees. Examples include the assignment of assembly tasks based on part tolerances and robot capabilities; the automated generation of robot paths with tolerances arising from sensed obstacles; and the optimization of part placement to minimize the effects of position uncertainty. This paper examines requirements for exchanging tolerance and uncertainty in robotics applications, identifies how these requirements are being met by existing standards, and suggests improvements to enable more automated information exchange.

Keywords

Uncertainty, tolerance, error, robot, computer aided process planning

Nomenclature

ADC	Analog to Digital Conversion
CAD	Computer-Aided Design
CAM	Computer-Aided Manufacturing
CAPP	Computer Aided Process Planning
D-H	Denavit-Hartenberg
DME	Dimensional Measurement Equipment
DMSC	Dimensional Metrology Standards Consortium
DMIS	Dimensional Measuring Interface Standard
DoF	Degrees of Freedom
GD&T	Geometric dimensioning and tolerancing
НТМ	Homogeneous Transform Matrix
НТТР	Hypertext Transfer Protocol
ISO	International Organization for Standardization
MBD	Model Based Design
PMI	Product and Manufacturing Information
POI	Point of Interest
QIF	Quality Information Framework
QMR	Quality Measurement Results
ROS	Robot Operating System
STEP	STandard for the Exchange of Product model data
URDF	Unified Robot Description Format

"Tolerances and Uncertainty in Robotic Systems." Paper presented at International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9,

XML	eXtensible Markup Language
XSD	XML Schema Definition

INTRODUCTION

Industrial robots have had much success automating repetitive tasks in structured environments, where their ability to be programmed has enabled their proliferation into a wide variety of applications such as part handling, spray painting, and welding. The predominant method of teach programming is effective but time consuming, and is an impediment to applying robots for short-turnaround jobs. Off-line programming solves this problem, but relies on robot accuracy which is typically much worse than that of machine tools, necessitating the use of sensors to compensate for this inaccuracy. Sensors also enable robots to operate in unstructured dynamic environments, and the continued improvement and cost reductions in sensors, especially vision, have now made it possible to deploy robots into previously hard-to-automate operations and for jobs with small lot sizes and short lead times, and often-changing performance requirements.

These changing requirements and uncertain environments make automated offline programming a necessity, but they also afford an opportunity to optimize manufacturing [1]. Effectively conveying information throughout the design, planning, and production phases is key to achieving this optimization. Until recently, this was a manual process, since product and manufacturing information resided in proprietary systems with no support in standards for preserving the semantic content throughout the export and import steps. The situation has improved, with standards in place for design, process planning, execution, and quality activities that make full semantic exchange possible. This paper will examine the state of these information exchange standards, and identify needs for revisions that can improve the efficiency and effectiveness of the exchange of tolerance and uncertainty information through the manufacturing chain for robots.

USE CASE SCENARIO

In this section, a use case scenario will be presented to provide context for the following sections on information exchange issues. At the outset, systems engineering is done to determine how the overall manufacturing activities will be carried out in the facility, given budgets, the equipment market, and staff capabilities. In this scenario, an assembly workcell has been established that combines industrial robots, dexterous manipulators, auxiliary equipment, and sensors, that together can achieve assembly performance to desired accuracy. This workstation includes two robots, each with dexterous grasping abilities, vision system throughout to determine the actual locations of parts and obstacles, force sensing for insertion of close-fitting parts, and auxiliary equipment for the staging and fixturing of parts as needed.

The process begins with a designer preparing a set of design files for components of an assembly that fulfills customer requirements for form, fit, and function. The design files specify product and manufacturing information, including geometric dimensions and tolerances. Using the design information, an assembly planner determines the order of operations needed to complete the full assembly, including any tooling requirements or other resources necessary to carry out individual steps. For complex assemblies, this may require the assignment of tasks to different resources, including people, depending on their capabilities. In this scenario, a process planner prepares a sequence of jobs for fabricating the assembly components [2, 3]. Robot programming for the assembly tasks is done offline, with no teach programming. Computer-aided design (CAD) files for the assembly components are loaded into an assembly planning system, which generates the sequence of operations, possibly well ahead of the actual assembly event. Because of the potential for robot collisions, the exact sequence of robot motions will be determined during execution, using information from the robot controllers and sensors. This can be done using explicit synchronization elements in the individual robot programs, through the sequential execution of programs through a supervisory controller, or through a single controller that does real-time planning for both robots as if they were a combined unit. In this scenario, a supervisory controller is used for high-level sequencing of programs for each robot.

The components arrive in two sets: a base assembly structure, in this case an engine block, and a kit of parts to be attached to it. The engine block has a single feature with tight assembly tolerances, and must be located so that this feature is accessible to the more accurate robot, and in its zone of highest accuracy. A prior inspection step has measured the actual location of this feature, and the CAD model has been updated. The block is affixed to a pallet that can be shuttled onto a rotary worktable via a mobile cart and conveyor. Once on the rotary table, it can be rotated to its most favorable orientation.

The parts kit is in an area accessible to both robots. The location of parts in the kit is not known a priori, so a vision system is used to determine their identity and location to sufficient accuracy. In general, parts may not be graspable in the orientation in which they are to be inserted, or in the order in which they must be assembled, so robot plans must be generated dynamically to reorient parts or clear out obstructing parts. In this scenario, a shared staging area is provided for these operations.

The location (six-degree-of-freedom (DoF) pose) of parts determined by the vision system will have some uncertainty, as will the location of the robot as it grasps the part, resulting in compounded uncertainty in the part as it is presented to its assembly location. To overcome this, a visual localization step is conducted just prior to assembly, where the robot presents the grasped part to a vision system that determines its actual location. The final robot path for assembly is adjusted by an offset to

2017.

the nominal pose.

A nominal free space robot path is computed that includes tolerances on position and orientation that ensure an envelope of collision-free motion. The tolerances give flexibility to the realtime path planner to optimize motion using machine-specific cost functions, such as minimizing joint motion or energy consumption [4].

For assembly tasks with tight clearances, position control is not sufficient, and force control must be done to ensure proper fit. The inclusion of force control may reduce the need for high accuracy, necessitating only moderate accuracy to bring the part into the range where it can engage the assembly location within the region required to begin the force-controlled insertion.

The parts kit arrives first, and the vision system shows that the parts are present and accessible, but that several need to be reoriented in order to be grasped properly for insertion. The supervisory controller generates a set of programs for the robots, consisting of a series of pick-and-place moves with path tolerances that ensure collision-free motion over their full duration, eliminating the need to synchronize motion during execution. The robots proceed to reorient the affected parts.

The engine block arrives and is shuttled into its nominal location on the rotary table. The vision system determines its actual pose, and the part is rotated so that the region of minimum uncertainty contains the high-tolerance feature, and this lies within the robot's region of maximum accuracy. This is the first assembly operation, and the robot acquires the part from the kit, takes it to the vision system to determine its actual grasped pose, and the offset insertion path is executed to bring the part into contact for the final force-controlled insertion.

Assembly of the other components continues. In most cases, the robots can operate independently, with their paths enveloped in tolerances that ensure collision-free motion. In some cases, however, there is the potential for collision unless the motion is synchronized. In these cases, the programs are sequenced by the supervisory controller, which suspends the activities of one robot while the other completes its task.

Once the assembly process is done, the completed assembly is transferred to a final quality assurance step, where the location of features and connections is verified according to the design requirements.

In the following sections, this use case scenario will serve as a reference example to illustrate the applicability of standards for exchanging tolerance and uncertainty, and issues that could be addressed through revisions to these standards.

TOLERANCES IN DESIGN

As described in the use case scenario, the process begins with a designer preparing a set of design files that define product geometry including geometric dimensions and tolerances. These tolerances signify how much variation is acceptable for satisfactory form, fit, and function. The assignment of optimal tolerances that keep manufacturing costs low while still meeting requirements can be difficult, and in many cases, they are informed by practical experience of the manufacturer. Standard practice is to represent this information in geometric dimensioning and tolerancing (GD&T) callouts, which could be interpreted visually but do not support automated querying, as shown in Figure 1.



FIGURE 1: A part with typical GD&T annotations

Standards for visually representing GD&T are ASME Y14.5-2008 [5] and ISO 1101:2012 [6]. These standards have been used for decades and cover a wide range of variations in product features, such as flatness, perpendicularity, location, circularity, and straightness. It is important to distinguish between the presentation of GD&T information, and the representation. Representation includes the information necessary to fully define the meaning of the tolerances, without resorting to visual depiction.

Standards for exchanging semantic information on nominal product information have been available for many years through the ISO 10303 "STEP" family [7]. STEP is a methodology for describing product information throughout its life cycle. CAD systems exchange STEP information through their import and export facilities, converting between native formats and STEP so that partners in the supply chain can each use software that best fits their needs. The advent of ISO 10303 Part 242 in 2014 standardized an information model for the meaning of GD&T. This improvement allows for the more full automation of information exchange further downstream, into production and quality assurance. An example from this information model is depicted in Figure 2.

```
ENTITY flatness_tolerance;
name : label;
description : OPTIONAL text;
magnitude : OPTIONAL length_measure_with_unit;
toleranced_shape_aspect : geometric_tolerance_target;
END_ENTITY;
```

FIGURE 2: EXPRESS entity showing an example of semantic GD&T for flatness

PROCESS PLANNING

The use of semantic GD&T greatly simplifies the task of exchanging models between CAD systems from different vendors, a problem faced by systems integrators and suppliers who must support different formats from different customers. It also opens the possibility of more fully automating the downstream processes that are influenced by tolerance requirements. Following the use case scenario, from the assembly design, a planner determines the order of operations needed to complete the full assembly, including any tooling requirements or other resources necessary to carry out individual steps depending on their capabilities. This process is variously known as computer-aided process planning (CAPP) or computer-aided manufacturing (CAM) depending on the domain. These systems automate much of the tedious calculating of end-effector motions or tool paths, and provide graphical aids for visualizing results, but rely on interaction from a human expert to guide the process and make determinations on suitable alternatives due to the absence of tolerances and other performance requirements. Even when these CAPP and CAM systems can use this information to automate the planning process, it must be manually entered. Here we see the full value of the exchange of semantic GD&T both in preserving information exchanged between design systems, and streamlining its use in downstream planning processes.

To determine if a robot can achieve the tolerance requirements, and to compute optimal robot motion, robot performance metrics are needed. The most common performance metric for industrial robots is repeatability, or the consistency with which the robot returns to a given point. This is due to the prevalence of the method of teach programming, where the robot is physically brought to a series of poses that are recorded for later playback in a programmed sequence. This method of programming can be time consuming, and requires a person to do the teaching while the robot is taken out of production. While it is cost effective for high-volume applications, it is often faster to do small jobs manually. In contrast, off-line programming uses models of the parts, robot, and work volume to generate sequences of nominal robot poses, relying on the accuracy of the robot to achieve the computed points. This is analogous to the primary method of programming machine tools, using CAD data and a CAM system. However, due to their construction, robots are typically much less accurate than machine tools, and off-line programming must be supplemented with other techniques to increase the accuracy to acceptable levels. Calibration can be done, but errors vary considerably throughout the work volume due to flexing of robot links, so calibration tables must be generated at many locations and orientations. This process is known as error mapping. Because robots typically have low stiffness and will deflect appreciably under loads, error maps are only effective when developed under loaded conditions. If these vary during an application, error mapping may be ineffective. Process sensing is effective under varying loading conditions, because the actual location of the robot is measured and deviations can be adjusted in real time. This requires an increased investment in sensing technology, and possible changes in the process to reduce noise or occlusions. With a combination of calibration and sensing, robot accuracy can be increased to approach its repeatability.

Supplemental sensing technology is often brought to bear in robot applications driven by offline programs, after calibration and error mapping have reduced systematic errors. A hybrid technique is to use a set of taught points at key areas in the work volume, benefitting from high repeatability, and calculating offsets from these taught points based in sensor data from cameras or other vision systems. Using this technique, accuracy errors are reduced to the much smaller neighborhood around the offset. Robots have been successfully used in this way for semiconductor chip placement, with placement accuracy requirements well below the 1-mm level. ISO 9283 [8] specifies methods for determining overall values of repeatability and accuracy that are guaranteed throughout the work volume for a given load. These include accuracy and repeatability for a target point under conditions of varying approach, accuracy and repeatability for motion along paths, position settling time and overshoot, cornering deviations, static compliance, and other metrics. These values are worst-case values, and in many cases the accuracy of a robot is much greater. On the other hand, these metrics are valid only for the test conditions, such as at 20 °C, and performance can be worse under other conditions. It is desirable to know where these regions of higher or lower accuracy lie, so that application programs can be placed accordingly and benefit from higher performance. To do this, a more sophisticated model of robot accuracy is needed. The Open Source Robotics Foundation's Robot Operating System (ROS) provides an information model for robots that helps achieve this, called the Unified Robot Description Language (URDF). URDF allows the description of geometric, kinematic, and dynamic information about the links and joints of a robot. Figure 3 shows a sample representation of a robot description.

The xyz and rpy attributes are the Cartesian and orientation (roll, pitch, and yaw) of the transforms between the links and joints. Other attributes define dynamic properties for mass, inertia, and friction. Note that there is no uncertainty associated with any of these values. Automated process planning could be im-

```
<link name="link1">
  <inertial>
   <origin xyz="0 0 0.5" rpy="0 0 0"/>
   <mass value="1"/>
   <inertia ixx="100" ixy="0" ixz="0"
         iyy="100" iyz="0" izz="100"/>
  </inertial>
  <collision>
   <origin xyz="0 0 0" rpy="0 0 0"/>
   <geometry>
    <cylinder radius="1" length="0.5"/>
   </geometry>
  </collision>
</link>
<joint name="joint1">
  <origin xyz="0 0 1" rpy="0 0 3.1416"/>
  <parent link="link1"/>
  <child link="link2"/>
  <dynamics damping="0.0" friction="0.0"/>
  imit effort="30" velocity="1.0"
       lower="-2.2" upper="0.7"/>
</joint>
```

FIGURE 3: Sample Unified Robot Description Format (URDF) showing how links and joints are represented

proved if this information were available when selecting robots for assembly tasks.

Run-time tolerances are supported by ROS in the form of poses with covariance, where the pose representation is Cartesian location and quaternion orientation. The tolerances (or uncertainty, depending on the context) are expressed using a covariance matrix on the Cartesian location, and the roll, pitch, and yaw equivalents of the quaternion orientation. This covariance approach to orientation is not well suited for uncertainty analysis, a point described in detail in the section on the use of orientation uncertainty in assembly tasks.

Robot path tolerances are also available in ROS, using tolerances on end points.

These tolerances apply to the joints values, not the Cartesian values, and so are dependent on the robot selected. ROS does provide various Cartesian motion planners, such as Descartes [9], but tolerance information is experimental. For example, orientation tolerances using a cone about the tool's directional axis have been used for insertion tasks.

QUALITY ASSURANCE

The Quality Information Framework (QIF) [10] is an ANSI standard sponsored by the Dimensional Metrology Standards Consortium (DMSC) that defines an integrated set of Extensible Markup Language (XML) information models to enable the

```
FollowJointTrajectoryActionGoal
    trajectory_msgs/JointTrajectoryPoint[] points
     float64[] positions
      float64[] velocities
     float64[] accelerations
      float64[] effort
      duration time_from_start
   control_msgs/JointTolerance[] path_tolerance
    float64 position
    float64 velocity
    float64 acceleration
   control_msgs/JointTolerance[] goal_tolerance
    float64 position
    float64 velocity
    float64 acceleration
   duration goal_time_tolerance
```

FIGURE 4: Tolerances on paths and goals for robot trajectories in ROS

effective exchange of metrology data throughout the entire manufacturing quality measurement process from product definition to inspection planning to execution to analysis and reporting. QIF handles feature-based dimensional metrology, quality measurement planning, first article inspection, and discrete quality measurement. QIF is gaining attention as an important quality technology [11–16].

QIF is based on XML, and uses terminology and semantics from the inspection world to represent the various elements in the QIF specification. The QIF information models are contained in files written in the XML Schema Definitions (XSD). The QIF XSD Version 2.0 models consists of six application schema files QIFRules, QIFResults, QIFPlans, QIFProduct, QIFStatistics, and QIFMeasurementResources bundled into a QIF Document. QIF also includes a library of XSD schema files containing information items used by all QIF applications (Auxiliary, Characteristics, Expressions, Features, GenericExpressions, Geometry, IntermediatesPMI, Primitives, PrimitivesPD, PrimitivesPMI, Statistics, Topology, Traceability, Units, and Visualization).

The flow of QIF data starts with generation of CAD + PMI data exported as QIF Model Based Design (MBD) application data. Quality planning systems import the MBD and generate Plans (whats), then import Resources and Rules information and export Plans (whats and hows). Programming systems import Plans to generate Dimensional Measurement Equipment (DME) specific programs, or general instructions to guide inspection. Dimensional measurement equipment executes programs and evaluates characteristics of a single manufactured part or assembly and exports the measurements as Results. Analysis systems, typically performing statistical process control, import single parts Results and generate analysis of multiple part batches as QIF Statistics data.

The role of uncertainty in QIF would be to characterize the

Franaszek, Marek; Michaloski, John; Proctor, Frederick.



FIGURE 5: QIF 2 Architecture

statistical distribution of the error. Underlying all measurement data in QIF is the assumption that the QIF Dimensional Measurement Equipment is of order of magnitude ten times more accurate than the inspected feature. Thus, a feature characteristic of 1 mm would require a DME that measures to .1 mm accuracy. This inspection rule-of-thumb allays many concerns, but is not absolute. In fact, QIF has a separate section on detailing the environment and other inspection factors, (e.g., ambient temperature), that could contribute to quality inspection errors.

QIF provides for inspection measurements to have an attached uncertainty associated with the result. In QIF, the measurement actual values (e.g., the diameter of an instance of CircleFeatureActualType) are elements that can have the optional attributes "meanError" and "combinedUncertainty". Those are attributes of the ActualDecimalType in Units.xsd shown below. Measurement data which correspond to actual value types in QIF are derived from ActualDecimalType.

```
<xs:complexType name="ActualDecimalType">
   <xs:annotation>
    <xs:documentation>
     An ActualDecimalType defines a SpecifiedDecimalType
     with two additional optional attributes: meanError
     and combinedUncertainty. These attributes should
    either both be used or both be omitted.
    </xs:documentation>
  </xs:annotation>
   <xs:simpleContent>
    <xs:extension base="SpecifiedDecimalType">
      <xs:attribute name="combinedUncertainty"
       type="NonNegativeDecimalType">
       <xs:annotation>
         <xs:documentation>
          The optional combinedUncertainty attribute
          is a value expressing the combined
          uncertainty assigned to the
          SpecifiedDecimalType.
```

```
</xs:documentation>
      </xs:annotation>
    </xs:attribute>
    <xs:attribute name="meanError"</pre>
      type="NonNegativeDecimalType">
      <xs:annotation>
       <xs:documentation>
         The optional meanError attribute is a
         value expressing the mean error
         assigned to the SpecifiedDecimalType.
       </xs:documentation>
      </xs:annotation>
    </xs:attribute>
   </xs:extension>
 </xs:simpleContent>
</xs:complexType>
```

Determining the uncertainty for a QIF actual measurement is an optional reporting attribute often viewed as unnecessary. It would be preferable if the vendor of the inspection software could have access to the numerous factors that play into determining the uncertainty, for example, ambient temperature, last calibration, and the model of DME reported benchmarks. If this were the case, taking inspection measurements in an intemperate $30 \,^{\circ}C$ ambient temperature would result in a large uncertainty.

SENSOR UNCERTAINTY

MTConnect is an integration standard to solve the "Island of Automation" problem in the discrete manufacturing industry. MTConnect is an open, royalty-free standard that uses prevalent commercial off-the-shelf technology - XML and HTTP. The MT-Connect intent is to foster greater interoperability between controls, devices, and software applications by publishing data over networks using the Internet Protocol [17]. Over the course of the MTConnect standards development, sensors have gone from an implicit modeling role within devices to an explicit information model. This is especially important as sensors can provide realtime production information to better understand and optimize manufacturing activities in a factory.

A sensor may measure in one dimension, such as temperature or acceleration whose fluctuations are a function of time, or a sensor can be multidimensional, such as an image which can be a function of two or three-dimensional space and time. MTConnect Part 2 V1.2 [18] presents a sensor model that defines sensor data formats and communication interfaces. MTConnect includes sensor information models for sensors of one dimension: acceleration, angular acceleration, angular velocity, amperage, angle, concentration, conductivity, direction, displacement, electrical energy, flow, frequency, fill level, linear force, load, mass, pH, pressure, position, power factor, resistance, rotary velocity, sound level, strain, temperature, time, tilt, torque, volt ampere, volt ampere reactive, velocity, viscosity, voltage, and wattage. At present, MTConnect only provides information models for one-dimensional sensor values.

Franaszek, Marek; Michaloski, John; Proctor, Frederick.

"Tolerances and Uncertainty in Robotic Systems." Paper presented at International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9,

In MTConnect, a sensor is comprised of two major components - a sensing element and a sensor interface. A sensing element provides a signal or measured value. It is modeled as an MTConnect DataItem. Each sensor model includes a sensing element, calibration, signal conditioning, and analog-to-digital conversion (ADC) information [19]. A sensor interface has capabilities, such as signal processing, conversion, and communications, and it is modeled as an MTConnect Component called Sensor. Each sensor interface may have multiple sensing elements, which represent the data for a variety of measured values. Further, when an MTConnect sensor represents multiple sensing element(s), each sensing element is represented by a Channel. A Channel represents one sensing element and can have its own attributes and Configuration data.



FIGURE 6: Sensor Architecture of MTConnect

The MTConnect architecture has an "Agent" that is a web service and acts as a "bridge" between an MTConnect "Device" and a Client Application. An MTConnect Device is a piece of equipment, like a robot, organized as a set of components that provide data. An MTConnect "Adapter" is a process that provides a data stream from a device to the agent. MTConnect defines XML information models in order to exchange standard data items. MTConnect has a so-called "Streams" information model that defines data reporting of Events, Samples, Conditions, and Asset data items through continually-updated channels. MTConnect has a standard configuration information model that provides data as a "probe". Figure 6 shows the duality of MTConnect sensor configuration as two types: A Sensor built into an MTConnect Device (i.e., Robot) and an independent Sensor as a standalone MTConnect Device. In Figure 6, the robot "contains" the sensors and will report sensed values as part of its data reporting.

The MTConnect standard provides an XML configuration report colloquially known as a "probe". The MTConnect probe enumerates the DataItems that will be reported in a "stream". The MTConnect probe represents sensor values as a DataItem sample, which includes XML attributes for sensor type, units, name, and XML id. Also included in the DataItem XML for acceleration is an attribute called statistic which is calculated specific to the sensor DataItem. The statistic attribute indicates that the data has been processed using a statistical operation like average, mean, or root square. Examples of MT-Connect statistic attributes are AVERAGE, MINIMUM, MAXI-MUM, ROOT_MEAN_SQUARE, RANGE, MEDIAN, MODE, and STANDARD_DEVIATION. Below is the XML reported from the temperature (thermistor) sensor when querying the MT-Connect Agent for a system configuration (i.e., probe) [20].

The above XML uses the MTConnect Devices XSD as the Information Model schema to describe each device and its data items available. Below, the XML shows the data reporting as configured using the MTConnect Streams XSD as the Information Model. MTConnect streams describe a time series of data items, including samples, events, and conditions. Below the Streams snippet describes the sensor readings for sensor one, the "temp" temperature sensor:

```
<Temperature dataItemId="temp"
timestamp="2017-03-07T21:17:42.814257"
name="Robotemp"
sequence="839777883"> 19.9
</Temperature >
```

A sensor measured value is rarely observed in isolation from a combination of noise and distortion [21]. In fact, noise and distortion are the fundamental source of the limitations in the accuracy of sensor measurements. For example, the sources of accelerometer noise can be broken down into the electronic noise from the circuitry that is converting the motion into a voltage signal or mechanical noise from the sensor itself. If an MTConnect sensor sample is returned as a data item, especially a sensor that may be noisy and prone to providing outlier values, a quantification of the error would be desirable. For example, temperature sensors typically provide a statement that the thermistor is accurate to ± 3 degrees. To translate this into an uncertainty value

Franaszek, Marek; Michaloski, John; Proctor, Frederick.

(or statistical distribution of the error), we assume that the error mean is zero and that three standard deviations from this mean provides approximately 99.7 % of the error statistical population assuming a normal error distribution. Of note, it is assumed that the vendor in providing a bounds on the sensor accuracy, has used a sufficiently large number of observations to provide a reliable estimate of the accuracy. For the thermistor example, the three-degree bounded limit translates into a standard deviation uncertainty of one degree.

It would be desirable for MTConnect to report the uncertainty associated with any measurement, especially sensors, since noisy measurements or outliers could pass through the data reporting system as ground truth, when in fact the numbers are abnormal and should be discarded or should be filtered. Below is an MTConnect sensor that incorporates uncertainty as an associated MTConnect data item:

```
<DataItem type="TEMPERATURE"
  category="SAMPLE"
  name="Robotemp"
  id="temp" units="CELSIUS" >
        <Source componentId="s1">channel:1</Source>
</DataItem>
<DataItem type="TEMPERATURE"
  statistics="STANDARD_DEVIATION"
  category="SAMPLE" name=" RobotempUnc"
  id="tempuncertainty" units="CELSIUS" >
</DataItem>
```

Now, every sensor temperature sensor measurement will have an associated uncertainty value associated with the reading. Below is MTConnect data stream XML snippet reporting the temperature sensor measurement, but now containing an associated data item to explicitly state the uncertainty of the temperature sensor measurement. For example, in the temperature sensor uncertainty reading, we assume measurement units correspond to those described in the probe XML given above, so the temperature uncertainty is given as a standard deviation of one degree

```
<Temperature dataItemId="temp"

timestamp="2017-03-07T21:17:42.814257"

name="Robotemp"

sequence="839777883"> 19.9

</Temperature >

<Temperature dataItemId="tempuncertainty"

timestamp="2017-03-07T21:17:42.814257"

name="RobotempUnc" sequence="839777883"> 1.0

</Temperature >
```

Now, the uncertainty of the MTConnect data item can also be used to register abnormal changes detected internally by the sensor that may affect its measurements. For example, suppose an acceleration sensor contains an internal temperature monitor and detects that its board's internal temperature is exceeding 40 °C, which adversely effects the sensor operation and its acceleration measurements. In this case, the uncertainty value could be negative, indicating the measures can never be equal to the mean. Although there is no explicit MTConnect facility for expressing multidimensional sensor data such as images, MT-Connect has the ability to incorporate and transport XML data independent of the core MTConnect information models. Using the MTConnect "asset" model, MTConnect agents can pass sensor data as embedded "asset" data. This facility along with asset notification and the "statistic" attribute can form the basis for reporting 2D and 3D sensor data. A brief overview will show the deployment of the MTConnect "asset" mechanism.

MTConnect defines "assets," which use an associative array of key/value stores to store the XML. This allows the ability to collect and report entire XML documents as they change within applications. Below, the XML shows how an AssetChanged tag with an asset type Sensor and READING value that would be updated within the MTConnect XML query to indicate new quality results from an inspection.

```
<AssetChanged dataItemId="multidimsensor_asset_chg"
timestamp="2016-09-08T19:42:16.855924Z" sequence="46"
assetType="Sensor"> Reading
</AssetChanged>
```

For our implementation, the Quality Measurement Results (QMR) XML Schema was used to develop the XML that is then assessable via the Internet with the following query to an MT-Connect agent –

http://xxx.xxx.xxx/asset/INSPECTION?type=Part

where xxx.xxx.xxx is the ip address of the MTConnect server, and which returns a so-called "blob" of otherwise unstructured multidimensional sensor data, outlined by the following XML snippet:

```
<MultiDimSensor timestamp="2011-07-25T13:55:22"
assetId="Reading">
<Reading>
<!-- this is the start of the sensor blob data -->
...
</Reading>
</MultiDimSensor>
```

Multidimensional sensors are affected by noise and distortion, so the uncertainty should reflect this. Although it is possible to embed another sensor data item in the MTConnect system configuration to reflect the uncertainty, it would be easier to incorporate an existing standard such as the Metadata Working Group Standard [22] to handle image data.

In theory, using the MTConnect sequence number to package data, a client could query an MTConnect agent to stream multidimensional data. No tests were done to verify this capability.

Franaszek, Marek; Michaloski, John; Proctor, Frederick.

"Tolerances and Uncertainty in Robotic Systems." Paper presented at International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9,

USE OF ORIENTATION UNCERTAINTY IN ASSEMBLY TASK

If a CAD model of a part is available, any Point of Interest (POI) associated with that part can be determined using sixdegree-of-freedom data acquired by a pose measuring system. Uncertainty of a selected POI is derived from uncertainty of pose measurement. Uncertainty of a part's location propagates homogeneously to its all POIs but propagation of orientation uncertainty may be more complicated and may have directional dependence.



FIGURE 7: Two vector bars mounted rigidly to a part form a local frame which is tracked by a pose measuring system. The location of any POI on a part (marked by an arrow) can be determined from dynamically tracked pose and fixed location of the POI in a local frame. Uncertainty in pose measurement must be propagated to the POI.

Let us assume the j-th noisy pose measurement yields R_{j} rotation and t_j translation. If a location of a POI in the CAD coordinate frame is $\boldsymbol{U} = U \boldsymbol{u}(\boldsymbol{\theta}, \boldsymbol{\varphi})$, where $U = \|\boldsymbol{U}\|$ and a unit vector u points in the direction of azimuth φ and elevation θ , then the location of a POI on the rotated object in the coordinate frame of the pose measuring system is

$$\boldsymbol{U}_i = U\boldsymbol{w}_j + \boldsymbol{t}_j \quad , \tag{1}$$

where a unit vector \boldsymbol{w}_i points to a rotated POI

$$\boldsymbol{w}_i = \boldsymbol{R} \boldsymbol{u}(\boldsymbol{\theta}, \boldsymbol{\varphi}) \quad . \tag{2}$$

Uncertainty in the orientation \mathbf{R}_i is propagated to \mathbf{w}_i and the noisy orientation can be represented as

$$\boldsymbol{R}_j = \boldsymbol{\bar{R}} \,\Delta \,\boldsymbol{R}_j \quad , \tag{3}$$

where the mean orientation \bar{R} is used as the approximation of the unknown true orientation and $\Delta \mathbf{R}_i$ is a small random rotation which can be expressed in axis-angle representation (a_i, ρ_i) as

$$\Delta \boldsymbol{R}_{j}(\boldsymbol{a}_{j},\boldsymbol{\rho}_{j}) \approx \boldsymbol{I} + \begin{bmatrix} 0 & -q_{j}^{z} & q_{j}^{y} \\ q_{j}^{z} & 0 & -q_{j}^{x} \\ -q_{j}^{y} & q_{j}^{x} & 0 \end{bmatrix} , \qquad (4)$$

where I is the identity matrix and

$$\boldsymbol{q}_j = \boldsymbol{\rho}_j \; \boldsymbol{a}_j \quad . \tag{5}$$

From N repeated measurements of noisy orientation \boldsymbol{R}_j , j = $1, \ldots, N$, a covariance matrix of orientation data cov(q) can be calculated and its three eigenvalues $\{\wedge_1, \wedge_2, \wedge_3\}(\wedge_1 < \wedge_2 < \wedge_3)$ and associated eigenvectors $\{e_1, e_2, e_3\}$ can be calculated. Noisy rotations \mathbf{R}_i can also be used to investigate distribution of transformed unit vectors w_i in (2). One way of doing this is to characterize a distribution of μ_i defined as the angle between w_i and the mean unit vector \bar{w} . The spread of this distribution may be characterized by the angular uncertainty σ : smaller σ correspond to a tighter concentration of noisy \boldsymbol{w}_i around the mean direction $\bar{\boldsymbol{w}}$ and larger σ correspond to wider spread of w_i around \bar{w} .

For a large class of pose measuring systems, angular uncertainty depends on direction, i.e., $\sigma = \sigma(\theta, \varphi)$ where azimuth and elevation angles determine direction of unit vector $u(\theta, \phi)$ in (2). Directional distribution of $\sigma(\theta, \phi)$ is closely correlated with the distribution of axes a_i of small noisy rotations $\Delta \mathbf{R}_i$ in (4,5), as can be seen in Figure 8 and Figure 9.

Thus, depending on the directions of eigenvectors of the covariance matrix of the orientation data, different POIs of the measured object will be affected differently. Such situation is shown in Figure 10.

In summary, the covariance matrix of the orientation data needs to be carefully analyzed, as not only the values of its diagonal elements (variances of orientation data) and off-diagonal elements (correlation coefficients) are important.

CONCLUSION

Improvements in sensing and control have enabled robots to be deployed in difficult-to-automate applications characterized

Franaszek, Marek; Michaloski, John; Proctor, Frederick. "Tolerances and Uncertainty in Robotic Systems." Paper presented at International Mechanical Engineering Congress & Exposition, Tampa, FL, United States. November 3, 2017 - November 9,



FIGURE 8: Histograms of axes a_j of small random rotations $\Delta \mathbf{R}_j$ on log scale in (a), -inf indicates empty bins; (b) directional distribution of angular uncertainty σ in [*mrad*]. Based on orientation data obtained with an Optitrack Duo, the length of data N > 50,000. Plotted directions of eigenvectors correspond to eigenvalues $\wedge_1, \wedge_2, \wedge_3$ of the covariance matrix of the orientation data q.



FIGURE 9: The same as in Figure 8 but based on orientation data obtained with another pose measuring system (iGPS).



FIGURE 10: Angular uncertainty mapped onto a CAD model based on the data acquired with: a) iGPS (same as in Figure 9b) and b) OptiTrack Duo (same as in Figure 8b).

by often-changing requirements for operation in uncertain environments. To more fully automate the optimization of these activities, it is important to be able to exchange information about the tolerance on required performance, and the uncertainty in measured performance. Standards for the exchange of this information have been revised with this objective, to varying success. This paper examined the support of these standards using a use case scenario, and showed the results of a case study on the use of orientation uncertainty to optimally place objects in a robot workcell. Future work by the authors will examine the performance improvements achievable in robot path control.

DISCLAIMER

Commercial equipment and software, many of which are either registered or trademarked, are identified in order to adequately specify certain procedures. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

REFERENCES

- Proctor, F. M., Hoorn, G. v. d., and Lipman, R., 2016. "Automating robot planning using product and manufacturing information". *Proceedia CIRP*, 43, pp. 208 213.
- [2] Huang, Y. F., and Lee, C. S. G. "A framework of knowledge-based assembly planning". In Proceedings. 1991 IEEE International Conference on Robotics and Automation, pp. 599–604 vol.1.
- [3] Dini, G., and Santochi, M., 1992. "Automated sequencing and subassembly detection in assembly planning". *CIRP Annals - Manufacturing Technology*, **41**(1), pp. 1–4.
- [4] Proctor, F., Balakirsky, S., Kootbally, Z., Kramer, T., Schlenoff, C., and Shackleford, W., 2016. "The canonical robot command language (crcl)". *Industrial Robot: An International Journal*, 43(5), pp. 495–502.
- [5] American Society of Mechanical Engineers, 2008. ASME Y14.5-2008, Dimensioning and Tolerancing.
- [6] International Organization for Standardization, 2004. ISO 1101:2004, Geometrical Product Specifications (GPS) -Geometrical tolerancing – Tolerances of form, orientation, location and run-out.
- [7] International Organization for Standardization, 1994. ISO 1101:2004, ISO 10303-1:1994 Industrial automation systems and integration – Product data representation and exchange – Part 1: Overview and fundamental principles.
- [8] International Organization for Standardization, 1998. ISO 9283, Manipulating industrial robots - Performance criteria and related test methods.
- [9] ROS Industrial. Descartes. http://wiki.ros.org/descartes. [Online: accessed 8-March-2017].
- [10] ANSI, 2014. ANSI/QIF Part 1-2014: Quality Information Framework (QIF) an integrated model for manufacturing quality information part 1: Overview and fundamental principles version 2.0. Report, Dimensional Metrology Standards Consortium, Inc. (DMSC).
- [11] Kramer, T., Stone, R., Hoffman, M., Hoffman, S., and Rippey, W., 2012. "Design and usage guide for version 0.92 of the Quality Information Framework data model and

Franaszek, Marek; Michaloski, John; Proctor, Frederick.

SP-141

XML schemas". US Department of Commerce, National Institute of Standards and Technology Technical Note 1777.

- [12] Morey, B., 2014. "Metrology takes flight". *Manufacturing Engineering*, 152(2).
- [13] Pippenger, B. "Three-dimensional model for manufacturing and inspection". In Proceedings ASME Turbo Expo 2013: Turbine Technical Conference and Exposition.
- [14] Zhao, Y., Kramer, T., Brown, R., and Xu, X., 2011. Information modeling for interoperable dimensional metrology. Springer Science & Business Media.
- [15] Zhao, Y. F., Horst, J. A., Kramer, T. R., Rippey, W., and Brown, R. J., 2012. "Quality information framework - integrating metrology processes". *IFAC Proceedings Volumes* (*IFAC-PapersOnline*), *14*(Part 1), pp. 1301 – 1308.
- [16] Zhao, Y., Kramer, T., Rippey, W., Horst, J., and Proctor, F. "An integrated data model for quality information exchange in manufacturing systems". In Proceedings of the 37th International MATADOR 2012 Conference, pp. 239 – 242.
- [17] MTConnect FAQ. www.mtconnect.org/institute/mtconnectoverview/faqs.aspx. [Online: accessed 8-March-2017].
- [18] MTConnect. MTConnect Standard Version 1.3.1, Part 2: Components and data items. www.mtconnect.org/standarddocuments. [Online: accessed 8-March-2017].
- [19] Lee, K. B., Song, E. Y., and Gu, P. S. "A sensor model for enhancement of manufacturing equipment data interoperability". In ASME 2012 International Mechanical Engineering Congress and Exposition, pp. 705–712.
- [20] MTConnect Sensors. http://www.mtcup.org/wiki/Sensors. [Online: accessed 8-March-2017].
- [21] Vaseghi, S. V., 2008. Advanced digital signal processing and noise reduction. John Wiley & Sons.
- [22] Metadata Working Group . Specifications. www.metadataworkinggroup.com. [Online: accessed 8-March-2017].

Improvement Strategies for Manufacturers using the MESA MOM Capability Maturity Model

Quanri Li1, Michael Brundage1, Boonserm (Serm) Kulvatunyou1, Dennis Brandl2, and Sang Do Noh3

¹Systems Integration Division, National Institute of Standards and Technologies {quanri.li, Michael.brundage, serm}@nist.gov ²BR&L Consulting, LLC dnbrandl@brlconsulting.com ³Department of Systems Management Engineering, Sungkyunkwan University sdnoh@skku.edu

Abstract. Recently, the concept of smart manufacturing has emerged as a new paradigm, with which manufacturers can pursue their competitiveness in the market. Such a paradigm can be viewed as the convergence of Information & Communication Technologies with human and manufacturing technologies. It is expected to bring a new wave of performance improvements to manufacturing industries. However, manufacturing enterprises need to be prepared to adopt and realize the full benefits from these technologies. MESA (Manufacturing Enterprise Systems Association) created the Manufacturing Operations Management/Capability Maturity Model (MOM/CMM) to help evaluate the maturity and readiness of manufacturing enterprises from the factory operation perspective. However, the model, in its raw form, can be time and resource consuming with 832 questions. It also lacks improvement strategies based on the results. The objective of this work is to restructure the questionnaire to reduce completion time and to outline strategies through which a manufacturing enterprise can derive its improvement plans.

Keywords: Smart manufacturing, Manufacturing operations management, Maturity model, Readiness assessment, Improvement strategies

1 Introduction

As the global competition in the manufacturing industry has grown radically, a large number of new technologies, both hardware and software solutions, have been developed and applied to the industry. As a result, the concept of smart manufacturing has emerged and become a new paradigm that is expected to lead to innovations across many industries. To implement the new paradigm across various manufacturing industries, research efforts are being carried out throughout the world under governmental projects such as NNMI (National Network for Manufacturing Innovation, USA), Industrie 4.0 (German), Horizon 2020: Factories of Future (EU),

Brundage, Michael; Kulvatunyou, Boonserm; Li, Quanri. "Improvement Strategies for Manufacturers using the MESA MOM Capability Maturity Model." Paper presented at International Conference on Advances in Production Management Systems, the path to intelligent, collaborative, and sustainable manufacturing, Hamburg, Germany. September 3, 2017 - September 7, 2017.

etc., aimed at bringing innovation to manufacturing processes, improvement of productivities and quality of products [1].

However, often times, manufacturing enterprises cannot fully benefit from these technologies unless they have achieved a certain level of maturity in their manufacturing operations [2]. For example, if there is no procedure established for production tracking, manufacturers are not able to select and locate necessary sensors in their manufacturing systems. Therefore, a method is needed for evaluating foundational capabilities that support repeatable daily manufacturing operations and continuous performance improvements along with a guideline to help manufacturers move toward the readiness for the emerging smart manufacturing technologies.

In our previous work, we developed the Smart Manufacturing System Readiness Assessment (SMSRA) method that measures the maturities of the information connectivity, organization, performance management, and IT system, all of which are important components for deploying new Information & Communication Technologies (ICTs) in a manufacturing system from the shop floor level to the enterprise level [3]. Manufacturing Enterprise Systems Association (MESA) Manufacturing Operation Management Capability Maturity Model (MOM/CMM), on the other hand, provides a method for evaluating the foundational capabilities that should be in place before or while adopting ICTs particularly at the shop floor level [4]. The model is based on the activity models that are defined in the ISA-95 Part 3 [5]. Although some overlaps exist, MOM/CMM can be viewed as a precursor to the SMSRA. In other words, a factory should already be at level 4 or level 5 of the MOM/CMM before applying the SMSRA.

NIST is working with MESA to pilot the MOM/CMM with manufacturers; however, feedback from an initial discussion with potential pilot participants provides the following conclusions: 1) providing answers to all 832 questions or to even a subset of questions for a single operational area to evaluate the maturity is too timeconsuming and 2) there is no clear direction to how a manufacturing organization can derive improvement plans based on the evaluation result.

In this paper, we present the result of our work to reduce the effort needed to complete the questionnaire and provide an initial guideline for manufacturers to derive an improvement strategy based on the evaluation result.

2 **Overview of MESA MOM/CMM**

Part 1 of ISA-95 defines five hierarchical levels (0 to 4) for processes and systems that make up a manufacturing enterprise system [6, 7]. The MESA MOM/CMM is defined based on level 3 processes, manufacturing operations management (MOM). Four operational areas are defined in level 3 as follows: production operations management, quality operations management, inventory operations management and maintenance operations management. Each operational area consists of a set of activities related to detailed scheduling, dispatching, execution management, resource management, definition management, data collection, tracking and performance analysis [5].

Brundage, Michael; Kulvatunyou, Boonserm; Li, Quanri. "Improvement Strategies for Manufacturers using the MESA MOM Capability Maturity Model." Paper presented at International Conference on Advances in Production Management Systems, the path to intelligent, collaborative, and

sustainable manufacturing, Hamburg, Germany. September 3, 2017 - September 7, 2017.

The maturity level is a measured questionnaire organized per activity [5]. Each activity can have the maturity level 0 to level 5. See Table 1 for example questions in each maturity level (within the actual questionnaire each maturity level may have up to 9 questions). The maturity levels are defined as follows:

Level 0: There has been no evaluation performed.

Level 1: The processes are at initial stage and not documented or formally managed.

Level 2: Some of the processes are repeatable with possibly consistent results and documented.

Level 3: The processes are defined with documented standards for all activities.

Level 4: The processes are defined and documented across all organizational groups.

Level 5: The processes focus on continuous improvement and optimization.

Beside evaluating the maturity from the activity viewpoint, the questions can also be organized and evaluate with respect to 7 aspects namely roles and responsibility, succession plans and backups, policies and procedures, technology and tools, training, information integration and KPI. This will be addressed in future work.

3 **Related works**

Numerous frameworks or models for maturity assessment and/or evaluation of business or processes exist in various industries. A maturity model serves as a mean for assessing and providing an improvement framework. The Capability Maturity Model (CMM) is a software engineering process improvement model developed by Software Engineering Institute (SEI) at Carnegie Melon University, which triggered explosive development of maturity models in diverse fields [8].

One of the first maturity models developed is the Quality Management Maturity Grid (QMMG) by Philip B. Crosby in 1979 [9]. It evaluates the maturity of the organization's service and product quality management processes and how organizational cultures are affected by those processes.

In the field of supply chain management, SCRL-Model (Supply Chain Readiness Level) is a well-defined maturity evaluation model. The model is designed to evaluate supply chain by answering assessment questions in various sections such as inventory, supplier consolidation, supplier/customer relationships, commodity price adaptability, visibility etc. By doing so, a SCRL number is assigned to each section and it yields a level that indicates the maturity of the supply chain and how well it has been prepared. The goal of the model is to increase operational efficiency by identifying and mitigating operational risks in the supply chain. [10]

While the SCRL-Model serves as an assessment tool for the supply chain management field, the Manufacturing Readiness Level has been developed by the US Department of Defense (DoD) for the manufacturing field. According to [11], the MRL defines the current level of manufacturing maturity along with elements such as cost and risks that affect the manufacturing processes.

Brundage, Michael; Kulvatunyou, Boonserm; Li, Quanri. "Improvement Strategies for Manufacturers using the MESA MOM Capability Maturity Model." Paper presented at International Conference on Advances in Production Management Systems, the path to intelligent, collaborative, and

sustainable manufacturing, Hamburg, Germany. September 3, 2017 - September 7, 2017.

Since MESA MOM/CMM consists of multi-dimensional questions to evaluate the maturity levels of activities in manufacturing operations and the aspect at the same time, the approach to assessment is different from the aforementioned models. The models introduced (QMMG, SCRL-Model, MRL) assess the processes or practices in one dimension with associated questions for each level.

Putting the number of dimensions covered by MESA MOM/CMM aside, an investigation on design of questionnaire should be addressed since MESA MOM/CMM consists of series of 832 questions. According to [12], the number of questions in a questionnaire affects the quality of answers to the questionnaires. Specifically, the rate of response to the questionnaire could be increased by reducing the number of questions. Also, to reduce cognitive load on respondents, it is recommended to have a maximum of 20 words per sentence, however, one question can consist of more than one sentence [12]. Also, questions that are worded in negative tones tend to have more cognitive loads on respondents due to longer time for processing. It is recommended to have questions worded in positive tones as often as possible [13]. In the next section, we discuss our work to improve MESA MOM/CMM questionnaire due to these research findings and the initial feedback described in section 1.

4 **Restructure of MESA MOM/CMM**

Currently, MESA MOM/CMM consists of 832 yes/no questions across 4 operational areas, each of which has 8 activities to evaluate maturity, effectiveness, robustness and repeatability of the manufacturing operations and practices. The MESA MOM/CMM can be treated as a benchmark for comparison between different operational areas and activities. However, parsing through and answering all 832 questions can be a daunting task.

One strategy to address this issue is to select a subset of operational areas and activities to focus on beforehand so that only the relevant subset of 832 questions is answered at one point in time. However, the approach does not save time and effort and as manufacturers may not be able to decide where to start with the operational area or subset. Furthermore, having a complete picture of the entire operation will allow a better investment decision to be made.

Therefore, restructuring the questionnaire to reduce the number of questions and words while retaining the original intent of the questionnaire is needed.

The following approaches are used in the restructuring process:

- Ouestions that are composite are decomposed into multiple questions. 1) Composite questions are ambiguous to the interviewee when only one part of the question is true. This improves the clarity and answerability of the question.
- 2) Questions in all levels under an activity are grouped according to overlapping keywords and are converted into multiple choice questions. This reduce the number of questions and words the interviewee must parse.

Table 1 shows example questions for the 'detailed production scheduling' activity within the 'production operations management' operational area. In the table, the keyword "procedure" for detailed production scheduling appears repetitively in level

Brundage, Michael; Kulvatunyou, Boonserm; Li, Quanri. "Improvement Strategies for Manufacturers using the MESA MOM Capability Maturity Model." Paper presented at International Conference on Advances in Production Management Systems, the path to intelligent, collaborative, and

sustainable manufacturing, Hamburg, Germany. September 3, 2017 - September 7, 2017.

1 to 3. Similar repetitions exist throughout the questionnaire. These questions can be grouped together based on those repeated keywords and be restructured into a multiple-choice question.

Table 2 shows an example of the restructured questionnaire. Notice that the question is no longer structured per maturity level, it rather aggregates across levels. A programmatic mapping has been developed such that the answer to the new representation of the question is automatically mapped to the original question and the maturity level can be automatically calculated.

Leve	Question	
1		
1	There are no defined procedures for Detailed Production Scheduling and	
	the processes are not repeatable during times of stress.	
2	Detailed Production Scheduling processes vary across organizational	
	groups, with different processes and procedures used in different groups.	
3	Detailed Production Scheduling processes are defined across all	
	organizational groups, and the organization follows written and controlled	
	policies.	
4	(All Level 3 processes are in place.)	
5	(All Level 4 processes are in place.)	

Table 1. Example of repeated questions

We have restructured questions in 2 activities under 1 operational area. The result is promising, the number of questions was reduced to 14 from 29, and the number of words was reduced approximately by 47% (from 473 words to 248 words) in each activity.

Table 2. Example of grouped questions

Question	Procedures for Detailed Production Scheduling
A	Not defined
В	Defined, but vary across organizational groups
С	Defined across all organizational groups

5 **Improvement strategies**

Three types of improvement strategies are proposed in this research: vertical improvement, horizontal improvement and hybrid improvement.

Since the evaluation work is time-consuming and labor-intensive, there should be a strategic and efficient approach to evaluations. Once an enterprise decides to evaluate their current status of maturity, they can choose an improvement strategy between the 3 different types as shown in Table 3.

Brundage, Michael; Kulvatunyou, Boonserm; Li, Quanri. "Improvement Strategies for Manufacturers using the MESA MOM Capability Maturity Model." Paper presented at International Conference on Advances in Production Management Systems, the path to intelligent, collaborative, and sustainable manufacturing, Hamburg, Germany. September 3, 2017 - September 7, 2017.

Vertical improvement strategy focuses on achieving high maturity levels across activities in a high priority, one operational area at a time. This strategy can work well with manufacturers, which know the particular operational area on which they would like to focus their improvement. If such a decision can be made early, the manufacturers can also pace the maturity assessment one operation at a time as well. Another advantage of this strategy is that it can result in high initial ROI (Return of Investment) compared to other strategies because the resources and investments are more focused.



Fig. 1. Different types of improvement strategies

Horizontal improvement strategy focuses on improving maturity levels of a specific activity across operational areas. With the improvement strategy, an enterprise can achieve balanced activity maturity across operational areas so that the enterprise can seek a way to have continuous and simultaneous improvement in different operational areas. This strategy takes less time to initially complete the maturity evaluation for selected activities. It may however require more capital investments as technologies tend to be modularized for each operational area such as scheduling system vs. inventory management system. The advantage can be that the strategy can lead to a more stable improvement for the entire organization.

Improvement strategies	Focus	Initial time needed	Initial capital investment	Effect
Vertical	High priority operational area	Low	Low	High initial ROI
Horizontal	Activity	Low	High	Balanced factory-wide improvement
Hybrid	Weakest activities	High	Depends	More balanced factory- wide improvement

Table 3. Improvement strategies and their tradeoffs

Brundage, Michael; Kulvatunyou, Boonserm; Li, Quanri. "Improvement Strategies for Manufacturers using the MESA MOM Capability Maturity Model." Paper presented at International Conference on Advances in Production Management Systems, the path to intelligent, collaborative, and sustainable manufacturing, Hamburg, Germany. September 3, 2017 - September 7, 2017.

Hybrid improvement strategy focuses on achieving balanced maturity levels for the entire matrix of activities and operational areas by prioritizing the improvement based on the weakest activity. With this improvement strategy, the enterprise can achieve factory-wide balanced maturity. However, if there are equally weak activities across operational areas or once the maturities across activities are leveled out, the horizontal or vertical strategy must be applied. The drawback is that to use this improvement strategy, a sizable initial effort, larger than those required in the previous two strategies, is needed to evaluate maturity levels of all activities across all operational areas. The capital investment will depend on the variation of the maturities across the matrix.

With evaluation and utilization of different types of strategies, an enterprise can pursue robust and mature manufacturing operations and processes which can yield continuous improvements and effective production. To achieve the goal of optimized manufacturing operations, an enterprise should have evaluated their practices. Fig. 2 shows example evaluation results of current and improved status of their practices in their manufacturing operations. As shown in the charts, in current practices, performance analysis (which is denoted as POM/PA) is the weakest activity among 8 activities under the production operations management. The rest of activities are evaluated no higher than level 3. If an enterprise in this example decides to take the vertical improvement strategy, which is to promote maturity levels of all activities under one operational area, the evaluation result would look like the chart noted "After" in Fig. 3. Since level 3 in MESA MOM/CMM is a threshold where all organizations should aim to achieve initially, one would like to pursue a way to promote maturity level of lowly evaluated activities until they match level 3 requirements before improving other activities already at level 3.



Fig. 2. Example evaluation results of current and improved status

6 Conclusion and Discussion

In this research, restructuring of MESA MOM/CMM and manufacturing operations improvement strategies were introduced. By grouping questions and converting them into multiple-choice questions, the number of words and questions included in the original model is reduced considerably. The preservation of the original intention is verified by a successful programmatic translation of answers in the new structure to the original structure. Our future work lies in the validation that the new structure saves time to answer all questions.

Three types of improvement strategies, each of which serve as a guideline for evaluating and improving the manufacturing operations practices, were proposed. With the improvement strategies and new questionnaire structure, manufacturing enterprises can make a better plan to complete the evaluation and track and improve its manufacturing operation management maturity which is foundational to deploying smart manufacturing technologies.

The MESA MOM/CMM is still in its infancy stage, and we are actively working toward piloting the model with real factories. We are also considering further improving the questionnaire - to make it more respondent-friendly. To this end, we will be exploring gamification techniques. Our hypothesis is that gamifying the questionnaire would grasp evaluators interests on completing the questionnaire and reduce the rate of evaluators' drop-off before the completion of the questionnaire.

Reference

- H. S. Kang, J. Y. Lee, S. Choi, H. Kim, J. H. Park, J. Y. Son, B. H. Kim and S. D. 1. Noh, "Smart manufacturing: Past research, present findings, and future directions," International Journal of Precision Engineering and Manufacturing - Green Technology, 2016
- M. Mcclellan, "The Collaborative Effect," Intelligent enterprises San Maeo, vol. 7, 2 p. 35, 2004.
- 3 K. Jung, B. Kulvatyunyou, S. Choi and M. P. Brundage, "An overview of a Smart Manufacturing System Readiness Assessment," in Intl. Conf. on Advances in Production Management Systems, 2016.
- ANSI/ISA-95.00.01-2013, Enterprise-Control System Integration Part 3: Activity 4 Models of Manufacturing Operations Management.
- D. Brandl, "MESA MOM Capability Maturity Model Version 1.0," White Paper 5. #53, MESA International white paper, 2016.
- 6. ANSI/ISA-95.00.01-2000, Enterprise-Control System Integration Part 1: Models and Terminology
- 7 D. He, A. Lobov and J. L. M. Lastra, "ISA - 95 Tool for Enterprise Modeling," in ICONS: The Seventh International Conference on Systems, Saint Gilles, Reunion, France, 2012.
- K. T. Yeo and Y. Ren, "Risk ManagementCapability Maturity Modelfor Complex 8. ProductSystems (CoPS) Projects," Systems Engineering, vol. 12.4, pp. 275-294, 2009.
- M. Mani, K. Lyons and R. Sriram, "Developing a sustainability manufacturing 9 maturity model," in Proceedings from the IMS Summer School on Sustainable Manufacturing, 2010.
- B. Tucker and J. Paxton, "SCRL-model for Human Space Flight Operations 10

Brundage, Michael; Kulvatunyou, Boonserm; Li, Quanri. "Improvement Strategies for Manufacturers using the MESA MOM Capability Maturity Model." Paper presented at International Conference on Advances in Production Management Systems, the path to intelligent, collaborative, and sustainable manufacturing, Hamburg, Germany. September 3, 2017 - September 7, 2017.

enterprise supply chain," 2010 IEEE Aerospace Conference, pp. 1-9, 2010.

- Manufacturing Technology Program, "Manufacturing Readiness Level (MRL)

 Deskbook,"
 May
 2011.
 [Online].
 Available:

 http://www.dodmrl.com/MRL_Deskbook_V2.pdf.
 [Accessed 14 March 2017].
 [Accessed 14 March 2017].
- D. A. Dillman, M. D. Sinclair and J. R. Clark, "Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupantaddressed census mail surveys," *Public Opinion Quarterly*, vol. 57.3, pp. 289-304, 1993.
- 13. A. N. Oppenheim, Questionnaire design, interviewing and attitude measurement, Bloomsbury Academic, 2000.
- G. Weems, A. Onwugbuzie and D. Lustig, "Profiles of respondents who respond inconsistently to positively-and negatively-worded items on rating scales," *Evaluation & Research in Education*, vol. 2003, pp. 45-60, 2003.
- D. Brandl, "Business to manufacturing (B2M) collaboration between business and manufacturing using ISA-95," *Revue De L Electricite Et De L Electronique 8*, pp. 46-52, 2002.



Laser Metrology and Performance XII

Identification of machine tools linear axes performance using on-machine embedded inertial measurement units

G.W. Vogl¹, A. Archenti², and M.A. Donmez¹ ¹National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, 20899, USA ²KTH Royal Institute of Technology, Brinellvägen 68, 10044, Stockholm, Sweden

Abstract

The current trend in manufacturing industry is from mass production towards adaptive manufacturing systems and cloud manufacturing. Self-learning machines and robot systems can play an essential role in the development of intelligent manufacturing systems and can be deployed to deal with a variety of tasks that can require flexibility and accuracy. However, in order for the machine tool (physical and control system) to deal with the desired task in a cognitive and efficient manner, the system must be "aware" of its capability in order to adjust itself to the desired task. Thus, characterization of machine tool accuracy and capability is necessary to realize this goal.

In this study, data from a machine-embedded inertial measurement unit (IMU), consisting of accelerometers and rate gyroscopes, was used for identification of changes in linear and angular error motions due to changes in operational conditions or component degradation. The IMU-based results were validated against laser-based measurement results, demonstrating that the IMU-based method is capable of detecting micrometer-level and microradian-level degradation of machine tool linear axes. Thus, manufacturers could use the method to efficiently and robustly diagnose the condition of their machine tool linear axes with minimal disruptions to production.

1 Introduction

Precision manufacturing has steadily evolved over the past decades to cope with the complexity of components and the requirements for high quality and low cost production. A typical machine tool has multiple linear axes whose accuracies directly impact the quality of manufactured parts. Yet over a machine tool's lifetime, various faults lead to performance degradation, lowering accuracy and repeatability [1]. Typical sources of errors within linear axes are due to pitting, wear, corrosion, and cracks of the system components such as guideways and recirculating balls [2]. As degradation increases, tool-to-workpiece errors increase, which eventually may result in a loss of production quality and/or a failure [3]. Yet knowledge of degradation is elusive; proper assessment of axis degradation is often a manual, timeconsuming, and potentially cost-prohibitive process.

While direct methods for machine tool performance evaluation are wellestablished [4] and reliable for position-dependent error quantification, such measurements typically interrupt production [5]. An online condition monitoring system for linear axes is needed to help reduce machine downtime, increase productivity and product quality, and improve knowledge about manufacturing processes [6]. Efforts to monitor the condition of linear axes components have utilized various sensors, e.g., built-in rotary encoders [7], current sensors [3], and accelerometers [8, 9]. These attempts at condition monitoring of linear axes have had limited success, partly because of the lack of robustness and defined relationships of signals to axis degradation composed of a wide range of spatial frequencies.

Consequently, efficient quantitative measures are needed to monitor the degradation of linear axes. Recently, accelerometers have been used for dynamic metrology of machine tools [10, 11] and six-degree-of-freedom motion sensors exist within integrated circuit (IC) components [12]. Thus, the use of an inertial measurement unit (IMU) is attractive for on-machine condition monitoring.

2 Methodology

One potential solution for online monitoring of linear axis degradation is the use of an IMU [13]. As seen in the schematic of Figure 1, an IMU is mounted to a moving machine tool component. To diagnose axis degradation, the axis is moved back and forth at various speeds to capture data for different frequency bandwidths. This data is then integrated, filtered, and fused to estimate the changes in the 6-degree-offreedom (DOF) geometric errors of the axis. Because the linear axes are stacked, coordinate transformations may be used with all 6-DOF errors to estimate the errors at the functional point [4]. Ideally, data would be collected periodically to track axis

degradation with minimal disruptions to production. With robust diagnostics and prognostics algorithms, incipient faults may be detected and future failures may be avoided. In essence, IMU data can be used to help optimize maintenance, production planning, and ultimately improve part quality.



Figure 1. IMU-based method utilizing data fusion for diagnostics of machine tool performance degradation.

Figure 2 and Figure 3 provide details of the data fusion schemes outlined in Figure 1 for estimation of translational and angular error motions using data corresponding to three speed regimes (the fast speed is 0.5 m/s, the moderate speed is 0.1 m/s, and the slow speed is 0.02 m/s). As seen in Figure 2, the accelerometer data is integrated twice, low- or band-pass filtered, processed, and summed to yield the net translational motions. As seen in Figure 3(b), the rate gyroscope data is integrated once, low- or band-pass filtered, processed, and summed to yield the net angular motions. However, if the measurement axis is nominally orthogonal to the gravitational vector, then the high-speed rate gyroscope data in Figure 3(b) may be replaced by the slowspeed accelerometer data, as shown in Figure 3(a). The schemes in Figure 2 and Figure 3 rely upon the matching of spatial cutoff frequencies and the exclusion of significant modal excitations.



Figure 2. Data fusion scheme for translational motions via use of accelerometer data. Filter cutoff frequencies are shown in parentheses.



Figure 3. Data fusion scheme for angular motions via use of (a) accelerometer and rate gyroscope data or (b) rate gyroscope data only. Filter cutoff frequencies are shown in parentheses. Wavelengths and frequencies correspond to 1 m of travel for 0.5 m/s (fast speed), 0.1 m/s (moderate speed), and 0.02 m/s (slow speed).

3 **IMU for Industrial Application**

For industrial application, an IMU should be physically small and economical while still satisfying the measurement needs. Consequently, for application on machine tools, an 'industrial IMU' was created that is about 73% smaller than the 'testbed IMU' used within a linear axis testbed at NIST [14]. As seen in Figure 4, the industrial IMU is about 9 cm long and contains a triaxial rate gyroscope and a triaxial accelerometer. The bandwidths and noise properties of these sensors are seen in Table 1.



Figure 4. (a) Isometric view of industrial IMU and (b) top view of industrial IMU without its lid.

Table 1. Properties of sensors in indu
--

Sensor	Bandwidth ^a	Noise
Accelerometer	0 Hz to 400 Hz	69 (μm/s²)/√Hz
Rate Gyroscope	0 Hz to 200 Hz	35 (µrad/s)/√Hz

^a frequencies correspond to half-power points, also known as 3 dB points

4 **Experimentation**

4.1 Setup

The industrial IMU shown in Figure 4 was designed for industrial application. Accordingly, repeated testing of the IMU on machine is required for acceptance testing, to determine how well the IMU-based methodology can detect various types of degradation.

Accordingly, Figure 5 shows an experimental setup of the IMU on a vertical milling machine at NIST. For each dataset, the IMU is attached to the worktable at one of three different locations (A, B, or C) and the Y-axis travels between Y = 0 m and Y = 0.5 m. Hence, IMU data was collected for 50 runs sequentially at each location with motion back and forth along the Y axis, according to the method outlined in Figure 1. The three speeds for data collection are 0.5 m/s (fast speed), 0.1 m/s (moderate speed), and 0.02 m/s (slow speed).



Figure 5. Experimental setup of IMU on machine tool worktable: Pictures of IMU at (a) Location A, (b) Location B, and (c) Location C and (d) schematic of three IMU locations relative to center point P of worktable.

Furthermore, data was collected with a commercial laser-based system ('Reference' system) for measuring the geometric errors of the axis at the three worktable locations. The laser-based system is used for verification and validation (V&V) of the IMU-based results. However, while the Reference system measures the motion of the worktable with respect to the base of the machine tool, the worktablemounted IMU measures the changes in the inertial motion of the worktable.

4.2 Results

The IMU data was processed to yield the estimated translational and angular error motions at the three worktable locations (A, B, and C). In order to estimate the translational error motions (E_{XY} , E_{YY} , and E_{ZY}) at the three locations, the IMU data was processed according to Figure 2. The linear term of the positioning error motion (E_{YY}) was determined based on an additional process, not shown in this paper for the sake of brevity, that utilizes only the slow-speed data. Also, the IMU data was processed according to Figure 3(a) or Figure 3(b) to estimate two angular error motions (E_{AY} and E_{BY}) or the third angular error motion (E_{CY}), respectively.

Assuming that the worktable is sufficiently rigid, the estimated error motions are

translated via homogeneous transformations from each of the three worktable locations (A, B, and C) to the center point P of the worktable. In theory, the estimated error motions at the center point P should be independent of which of the three sets is used for the homogeneous transformation. Figure 6 shows the translational errors estimated at the center point P of the worktable, based on data collected at three locations (A, B, and C). The data from the Reference system is also shown in the figures (as thinner lines) for comparison purposes. At each worktable location, Reference data was collected for five runs, which were averaged to produce the curves seen in Figure 6. The standard deviations of each set of five runs was also used to produce the shaded 95 %-confidence zones in Figure 6. Thus, the shaded zones represent a contribution towards, but not the total of, the measurement uncertainty.



Figure 6. Translational errors (a) E_{XY} , (a) E_{YY} , and (a) E_{ZY} estimated at center point P of worktable based on data collected at three locations (A, B, and C) with the IMU and a commercial Reference system. The Reference data has shaded areas representing measurement expanded uncertainties (k = 2) at 95 % confidence based on 5 runs.

As seen in Figure 6, the estimated translational errors from the IMU data match each other respectively to within 5 µm for the three worktable locations (A, B, and C). Also, the estimated error motions from the IMU match those from the reference system to within about 5 µm. The slight differences may be due to differences in error type (inertial for IMU, while relative for the Reference system) as well as to sources of uncertainty.

4.3 Factors affecting measurement uncertainty

For any type of measurement analysis, it is required that uncertainty and/or factors affecting uncertainty are identified and specified. Measurement uncertainty is a combination of uncertainties due to instrumentation, application of the instrument (e.g., erroneous alignment), environmental influences (e.g., floor vibration), and the object (e.g., repeatability of machine tool).

The most significant contributors to the measurement uncertainty u [15] in this study are the IMU, the reference laser system ($u_{\text{laser}} = 0.7 \,\mu\text{m}$ or 3.0 μrad), and the environmental conditions of the measurement runs. The IMU box was rigidly attached to the machine tool table and the condition of the machine tool operation was such that results are repeatable. The laser-based measurement device was applied for five runs according to manufacturer instructions, in order to compute the standard deviation and indicate repeatable results, as seen by the shaded 95 % confidence intervals in Figure 6. Furthermore, environmental conditions were within normal shop conditions (temperature range, floor vibration limit, etc.).

According to simulations of the data fusion process, the accelerometer used in the industrial IMU will result in uncertainties that could be sufficiently low for machine tool purposes. As seen in Figure 7, an accelerometer noise of 69 (μ m/s²)/ \sqrt{Hz} (for the industrial IMU) should result in a straightness uncertainty of about 5 µm when the results of ten runs are averaged.



Figure 7. Simulated uncertainty for straightness error motions due to only data fusion process with uncertainty sources of accelerometer noise, data acquisition noise, and number of runs used for averaging.

These simulations include only influences from accelerometer noise, data acquisition noise, and the method itself, while excluding uncertainties from other sources such as sensor misalignment, crosstalk, and system vibrations. Figure 7 shows that as the accelerometer noise decreases, the uncertainty decreases to a limit caused by noise of the data acquisition equipment. Thus, smaller uncertainties are possible with a more expensive and accurate IMU [14].

5 Conclusions

An 'industrial IMU' was developed to test the effectiveness of a new IMU-based method for on-machine application. The industrial IMU includes a triaxial accelerometer and a triaxial rate gyroscope. An experiment was conducted in which data was collected from the IMU at three different locations on a worktable for the same Y-axis motion.

The IMU-based results were validated through comparison with the laser-based measurement results, showing that the IMU could perhaps be used to track the changes of error motions due to linear axis degradation. The IMU-based method is capable of detecting micrometer-level and microradian-level degradation of linear axes. Although the results are promising, the IMU-based method must be improved for robustness due to sensor drift. Concerning the uncertainty of the IMU-based method, further investigations are also needed to fully certify the system for online measurement. Nonetheless, verified and validated data from an 'industrial IMU' could provide manufacturers and machine tool operators with near-real-time equipment health, diagnostic, and prognostic intelligence to significantly enhance asset availability and minimize unscheduled maintenance.

Acknowledgements

The authors thank Brian Pries, Travis Shatzley, Dan Falvey, and Jay Brandenburg of the Fabrication Technology Group (NIST) for their outstanding contributions with the experimental setup.

References

[1] Li Y, Wang X, Lin J, Shi S (2014) A Wavelet Bicoherence-Based Quadratic Nonlinearity Feature for Translational Axis Condition Monitoring. Sensors 14(2):2071-2088.

[2] Zhou Y, Mei X, Zhang Y, Jiang G, Sun N (2009) Current-Based Feed Axis Condition Monitoring and Fault Diagnosis. 4th IEEE Conference on Industrial Electronics and Applications, ICIEA 2009, 1191-1195.

[3] Uhlmann E, Geisert C, Hohwieler E (2008) Monitoring of Slowly Progressing Deterioration of Computer Numerical Control Machine Axes. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 222(10):1213-1219.

[4] International Organization for Standardization (2012) ISO 230-1 - Test Code for Machine Tools - Part 1: Geometric Accuracy of Machines Operating under No-Load or Quasi-Static Conditions.

[5] Khan AW, Chen W (2009) Calibration of CNC Milling Machine by Direct Method. 2008 International Conference on Optical Instruments and Technology: Optoelectronic Measurement Technology and Applications, 7160:716010.

[6] Teti R, Jemielniak K, O'Donnell G, Dornfeld D (2010) Advanced Monitoring of Machining Operations. CIRP Annals - Manufacturing Technology 59(2):717-739.

[7] Verl A, Heisel U, Walther M, Maier D (2009) Sensorless Automated Condition Monitoring for the Control of the Predictive Maintenance of Machine Tools. CIRP Annals - Manufacturing Technology 58(1):375-378.

[8] Liao L, Lee J (2009) A Novel Method for Machine Performance Degradation Assessment Based on Fixed Cycle Features Test. Journal of Sound and Vibration 326(3-5):894-908.

[9] Spiewak S, Zaiss C, Ludwick SJ (2013) High Accuracy, Low-Invasive Displacement Sensor (Halids). ASME 2013 International Mechanical Engineering Congress and Exposition, IMECE 2013, 2A:V02AT02A077.

[10] Smith KS, Hocken RJ, 2013, Dynamic Metrology Methods and Systems. United States Patent and Trademark Office: US8401691 B2.

[11] Sato R, Nagaoka K, Sato T, 2015, Machine Motion Trajectory Measuring Device, Numerically Controlled Machine Tool, and Machine Motion Trajectory Measuring Method. United States Patent and Trademark Office: US20110246132 A1.

[12] InvenSense Incorporated, 2016, MPU-6050 Six-Axis (Gyro + Accelerometer) MEMS MotionTracking[™] Device, https://www.invensense.com/products/motiontracking/6-axis/mpu-6050/.

[13] Vogl GW, Weiss BA, Donmez MA, 2015, A Sensor-Based Method for Diagnostics of Machine Tool Linear Axes, Annual Conference of the Prognostics and Health Management Society 2015. Coronado, CA: PHM Society, p. 10.

[14] Vogl GW, Donmez MA, Archenti A (2016) Diagnostics for Geometric Performance of Machine Tool Linear Axes. CIRP Annals - Manufacturing Technology 65(1):377-380.

[15] Knapp W (2002) Measurement Uncertainty and Machine Tool Testing. CIRP Annals - Manufacturing Technology 51(1):459-462.

A review of current machine learning techniques used in manufacturing diagnosis

Toyosi T. Ademujimi¹, Michael P. Brundage² and Vittaldas V. Prabhu¹

¹ The Pennsylvania State University, State College PA 16801, USA ² National Institute of Standards and Technology, Gaithersburg MD 20899, USA tta5@psu.edu, michael.brundage@nist.gov, vxp7@engr.psu.edu

Abstract. Artificial intelligence applications are increasing due to advances in data collection systems, algorithms, and affordability of computing power. Within the manufacturing industry, machine learning algorithms are often used for improving manufacturing system fault diagnosis. This study focuses on a review of recent fault diagnosis applications in manufacturing that are based on several prominent machine learning algorithms. Papers published from 2007 to 2017 were reviewed and keywords were used to identify 20 articles spanning the most prominent machine learning algorithms. Most articles reviewed consisted of training data obtained from sensors attached to the equipment. The training of the machine learning algorithm consisted of designed experiments to simulate different faulty and normal processing conditions. The areas of application varied from wear of cutting tool in computer numeric control (CNC) machine, surface roughness fault, to wafer etching process in semiconductor manufacturing. In all cases, high fault classification rates were obtained. As the interest in smart manufacturing increases, this review serves to address one of the cornerstones of emerging production systems.

Keywords: Artificial Intelligence, Machine Learning, Manufacturing Diagnosis, Fault Detection, Intelligent Maintenance, Industrie 4.0.

1 Introduction

Timely diagnosis of process faults provides a key advantage to help manufacturing companies stay competitive by reducing machine downtimes as more customers require manufacturers to provide the products quickly, at low cost and with high quality. Also, during machine downtimes, most of the time spent is on localization of the fault rather than carrying out the actual remediation, as fault diagnosis is the most challenging phase of machine repairs [1]. This has resulted in companies looking for new ways to improve their fault root cause analysis (RCA) process.

With current improvement in sensor technology, data storage, and internet speeds, factories are becoming smarter and more process data is generated. Research projects are focusing on how to utilize this 'big data' to improve manufacturing competitiveness. One such effort is a project at the National Institute of Standards and Technology (NIST) titled prognosis and health management for smart manufacturing (PHM4SMS)

Ademujimi, Toyosi; Brundage, Michael; Prabhu, Vittaldas. "A review of current machine learning techniques used in manufacturing diagnosis." Paper presented at APMS 2017 International Conference Advances in Production Management Systems (APMS 2017), Hamburg, Germany. September 3, 2017 - September 7, 2017. which is aimed at developing the necessary measurement science to enable and enhance condition-monitoring, diagnostics and prognostics [2]. Part of this effort is the utilization of machine learning techniques to improve fault detection (FD) in manufacturing.

The aim of this paper is to review the recent application of machine learning techniques to manufacturing process diagnosis. This review covers papers published from 2007 to 2017 that utilized machine learning techniques for manufacturing fault diagnosis. This review covers 20 articles. The keywords used in the search are "machine learning application in manufacturing process diagnosis". The search was filtered to focus on artificial neural networks (ANN), Bayesian networks (BN), support vector machine (SVM) and hidden Markov model (HMM) techniques. The rest of the paper reviews findings in each of these prominent techniques and provides conclusion along with future directions of research.

2 **Bayesian Networks**

Bayesian networks are a commonly used machine learning technique for FD. BN is a directed acyclic graph whose nodes represent random variables and their conditional dependencies are depicted by directed arcs linking the nodes [3].

Modeling a problem using BN requires specification of the network structure as well as the probabilities for each node. For generating tree structures, different authors proposed using several tools that depict the cause and effect relationship between these nodes. In generating the tree structure, De et al [4] and Pradhan et al [5] proposed the use of failure mode and effect analysis (FMEA), Pradhan et al [5] and Nguyen et al [6] utilized fishbone diagrams (cause and effect diagram), Pradhan et al [5] utilized faulttree analysis and variation sensitivity matrix, and finite element analysis (FEA) was used by Liu & Jin [7]. The precise construction of the tree structure of a BN from data is an NP-hard optimization problem [8]. Yang & Lee [9] and Correa et al [10] made use of K2 [11] and Chow-Liu [12], algorithms respectively to generate trees from data. Jeong et al [13] extract the cause-effect relationship for the equipment that is diagnosed from the equipment's maintenance manual. Process data obtained from sensors and stored in manufacturing execution systems (MES) or maintenance databases are then used to generate the conditional probabilities of the network.

The areas of application of BN vary across manufacturing industries. In the semiconductor industry, Yang & Lee [9] and Nguyen et al [6] used a BN to evaluate process variable influence on wafer quality to diagnose root cause of defective wafers using historic process data. Other application areas include the automobile industry [7] where BN is used to diagnose fixture fault in a taillight assembly and in machining [10] where BN is used to diagnose surface roughness fault. Data sources include quality management systems (QMS), manufacturing execution systems (MES), recipe management systems (RMS), computerized maintenance management systems (CMMS), and coordinate measuring machines (CMM). Table 1 gives a summary of the data sources, algorithms or methods used to determine the tree structure and case study or area of application for each of the BN papers surveyed.

Ademujimi, Toyosi; Brundage, Michael; Prabhu, Vittaldas. "A review of current machine learning techniques used in manufacturing diagnosis." Paper presented at APMS 2017 International Conference Advances in Production Management Systems (APMS 2017), Hamburg, Germany. September 3, 2017 - September 7, 2017.

BN is a white box model as the graphical representation makes it intuitively easy for the user to understand the interaction between the model variables. It is useful for modeling uncertainty and can be readily used to model hierarchical levels of multiple causes and effects with data from numerous sources, which is typically found in manufacturing systems. The same BN model can be used for both prediction and diagnosis. The main challenge of training a BN is in the construction of the tree structure and several methods including expert opinion have been proposed to mitigate this challenge [14].

Ref.	Tree Structure	Data Source	Case Study/Industry Applied
4	FMEA	QMS Systems: Warranty data, Corrective Action Reports (CAR)	N/A
5	Ontology and failure related knowledge	QMS Systems: 5-why, 8-D & field failure report; MES data	Original Equipment Manufac- turers and their tiered suppliers
6	Cause and effect (fishbone) diagram	Historic Process data in MES, CMMS and RMS databases	Semiconductor manufacturing process
7	Variation sensitivity matrix and FEA	Metrology data from CMM	Taillight assembly process
9	K2 algorithm	Historic Process data from sen- sors	Semiconductor manufacturing process
10	Chow-Liu algorithm	Sensor data from designed experi- ments	Surface roughness in machin- ing
13	Equipment mainte- nance manual	Simulated data stored in a lookup table	Machines in an electronics manufacturing company

Table 1. Summary of BN Applications

3 Artificial Neural Network

Artificial neural network is a non-parametric machine learning algorithm inspired by the functioning of the human central nervous system [15]. The adaptive nature provides a powerful modeling capability suited for non-linear relationships among features.

ANN has been used for many manufacturing FD applications. For complex problems with multiple layers and nodes, the network's training time might be significant. To decrease this training time, Barakat et al [16] developed a self-adaptive fault diagnosis ANN which adjusts the number of nodes according to the network's input parameters and terminates the training process according to a set of criteria. The idea was illustrated in the detection and isolation of disturbances in a chemical reactor simulator. Demetgul et al [17] also proposed an optimal configuration algorithm for neural networks used for FD. The algorithm combined genetic algorithm (GA) and ANN to eliminate the trial and error process for selection of the fastest and most accurate ANN configuration by using a fitness function to keep the number of hidden layer(s) and nodes at the minimum possible. The performance of Demetgul's [17] proposed system in FD was evaluated using experimental data collected from a bottle capping pneumatic work cell.

To add FD capabilities to control charts, Zhao & Camelio [18] integrated a neural network (NN) with a statistical process control (SPC) chart. The authors incorporated
process knowledge and measurements of a single sheet metal part to detect and diagnose fixture location fault in an assembly system as a proof of concept. Zhao & Camelio [18] also applied the SPC and NN in an automotive assembly process by using measurement data from the door subassembly to detect potential sources of variation during the assembly of the door to the vehicle.

The direct use of data from analog sensors or multivariate data sensors for FD applications requires a signal processing technique or dimension reduction techniques in the case of multivariate data in conjunction with a machine learning technique. Hong et al [19] proposed an algorithm that combines principle component analysis (PCA), modular neural network and Dempster-Shafer (D-S) theory to detect fault in an etcher system in semiconductor manufacturing. Process data was acquired by sensors and PCA reduced the dimensionality of the multivariate tool data set [19]. Zhang et al [20] also proposed a critical component monitoring method that utilizes Fast Fourier Transform (FFT) to extract features from sensor signals followed by training an Artificial Neural Network (ANN) with the transformed data to predict machine degradation and identify component faults.

Yu et al [21] used clustering as an unsupervised procedure to obtain informative features from vibration sensor signals attached to the motor housing of a machine to determine the bearing condition. An ANN was used for diagnosing machine faults based on these feature vectors [21]. Fernando & Surgenor [22] utilized an unsupervised ANN based on Adaptive Resonance Theory (ART) for FD and identification on an automated O-ring assembly machine testbed. Sensor data was collected while the machine was operating under different conditions (normal condition as well as faulty conditions) and features extracted from the raw sensor data. ART ANN could achieve excellent FD performance with minimal modeling requirements [22].

ANN's non-parametric nature and its capability to model nonlinear complex problems with high degree of accuracy has made ANN applicable to FD problems. The model is easy to initialize as there is no need to specify the network structure like in the case of BN. However, disadvantages include the "black box" nature which makes it difficult to interpret the model. Also, ANN often cannot deal with uncertainty in inputs and is computational intensive making convergence typically slow during training. ANN is prone to overfitting and requires a large diversified dataset for training to prevent this problem.

4 Support Vector Machine

SVM uses different kernel functions like radial basis function (RBF) or polynomial kernel to find a hyperplane that best separates data into their classes, and has good classification performance when used with small training sets [23]. Successful areas of application of SVM range from face recognition, recognition of handwritten characters, speech recognition, image retrieval, prediction, etc. [24].

Application of SVM exist in fault localization, although it is not as common as BN and ANN [25]. The technique was used by Hsueg & Yang [26] to diagnose tool breakage fault in a face milling process under varying cutting conditions. Kumar et al [27]

Ademujimi, Toyosi; Brundage, Michael; Prabhu, Vittaldas. "A review of current machine learning techniques used in manufacturing diagnosis." Paper presented at APMS 2017 International Conference Advances in Production Management Systems (APMS 2017), Hamburg, Germany. September 3, 2017 - September 7, 2017.

created a MapReduce framework for automatic diagnosis for cloud based manufacturing using SVM as the classification algorithm and validated this with a case study of fault diagnosis using the steel plate manufacturing data available on UCI Machine Learning Repository [28]. Demetgul [23] used SVM to classify 9 fault states in a modular production system (MPS) using data obtained from eight sensors, and experimented with 4 different kernel functions namely RBF, sigmoid, polynomial and linear kernel functions, and got 100% classification rate on all except for sigmoid kernel which had 52.08% classification rate. Decision tree technique developed using QUEST (Ouick, Unbiased and Efficient Statistical Tree), C&RT (Classification and Regression Tree), and C5.0 algorithms were also applied to the same dataset and 100% classification rate was obtained, and 95.83% for Chi-square automatic interaction detection (CHAID) [23]. Demetgul [23] concluded that SVM and decision tree algorithms are very effective monitoring and diagnostic tools for industrial production systems.

SVM is an excellent technique in modeling both linear and nonlinear relationships. Computation time is relatively fast when compared with other nonparametric techniques, such as ANN. Availability of large training datasets is a challenge in machine learning, however SVM tends to generalize well even with a limited amount of training data. Also, NIST is actively developing use cases that are representative of common manufacturing processes to support prognosis and health management research [29].

5 **Hidden Markov Model**

Hidden Markov Model is an extension of the Markov chain model used to estimate the probability distributions of state transitions and that of the measurement outputs in a dynamic process, given unobservable states of the process [30].

HMM has been used in fault diagnostics of both continuous and discrete manufacturing systems. In the continuous case, Yu, [31] proposed a new multiway discrete hidden Markov model (MDHMM) for FD and classification in complex batch or semi batch production processes with inherent system uncertainty. Yu [31] applied the proposed MDHMM approach to the fed-batch penicillin fermentation process which classified different types of process faults with high fidelity. For the discrete case, HMM was applied by Boutros & Liang [32] to detect and diagnose tool wear/fracture and ball bearing faults. The model correctly detected the state of the tool (i.e., sharp, worn, or broken) and correctly classified the severity of the fault seeded in two different engine bearings [32]. In addition to the fault severity classification, a location index was developed to determine the fault location (inner race, ball, or outer race) [32].

As with analogue sensor signal used along with ANNs, Yuwono et al [33] also used HMM along with advanced signal processing techniques to discover the source of defect in a ball bearing. The algorithm was based on Swarm Rapid Centroid Estimation (SRCE) and HMM and the defect frequency signatures extracted with Wavelet Kurtogram and Cepstral Liftering were used to achieve on average the sensitivity, specificity, and error rate of 98.02%, 96.03%, and 2.65%, respectively, on bearing fault vibration data provided by Case School of Engineering, Case Western Reserve University [33].

Ademujimi, Toyosi; Brundage, Michael; Prabhu, Vittaldas. "A review of current machine learning techniques used in manufacturing diagnosis." Paper presented at APMS 2017 International Conference Advances in Production Management Systems (APMS 2017), Hamburg, Germany.

September 3, 2017 - September 7, 2017

HMM is a probabilistic model that is excellent at modeling processes with unobservable states such as chemical processes or equipment's health status, thus a good fit for FD. However, the training process is usually computationally intensive.

Table 2. Pros and Cons of each Technique

Technique	Advantage	Disadvantage
BN	Intuitively easy to understand. Good for modeling uncertainty. Can be used to model hierarchical levels of multiple causes and effects. Can reason in both directions (prediction and diagnosis).	The tree structure makes it rel- atively less easy to initialize. Constructing the tree structure may be challenging.
ANN	Can model nonlinear complex problems with high degree of accuracy. Relatively easier to initialize, no need to specify the network structure like in the case of BN.	The model is not easy to inter- pret and cannot deal with un- certainty in inputs. Computationally intensive making convergence typically slow during training. Prone to overfitting.
SVM	Excellent in modeling linear and nonlinear relationships. Computation time is relatively fast when compared with ANN. Tends to generalize well even with limited amount of training data.	Selection of the kernel function parameters is challenging. Not easy to incorporate domain knowledge. Difficult to understand the learned function.
HMM	Excellent at modeling processes with unob- servable states.	Training process is usually computationally intensive.

6 Conclusion and Future Research Direction

A summary of the different techniques' advantages and disadvantages are presented in Table 2. Most data used was process data acquired using sensors and training was done through designed experiments either in a laboratory or in an industry setting. The authors that proposed using data from QMSs did not validate their proposal with data from real applications because of the difficulty in obtaining real data or the challenge in mining Quality Information System (QIS) data such as corrective action reports, warranty information, or deviation reports, which are mostly in paper form. Also, most of the case studies in the papers were limited to FD in a single machine; therefore diagnosis of a whole factory or manufacturing line consisting of multiple machines was not considered.

BN and HMM techniques are both excellent at modeling fault diagnosis with hierarchical levels consisting of multiple causes and effects. BN requires less computational power than HMM. ANN produced very accurate results and several approaches were proposed to reduce the large training time it required. ANN was also used in conjunction with a signal processing technique for applications with analogue sensor data. Unlike the other models, ANN is a black box and it is not easy to visualize what is occurring in the model. ANN is also prone to overfitting. Although not often used in fault diagnosis in comparison to other machine learning methods surveyed, SVM produced high fault classification rate at less computation time than ANN.

Future work will explore using real QIS data to further improve the diagnosis process as well as extending the single stage diagnosis to multi stages to include the entire manufacturing factory.

References

- 1. Kegg, R.: One-Line Machine and Process Diagnostics. CIRP Annals-Manufacturing Technology 33(2), 469-473 (1984).
- Weiss, B., Vogl, G., Helu, M., Qiao, G., Pellegrino, J., Justiniano, M., Raghunathan, A.: Measurement science for prognostics and health management for smart manufacturing systems: key findings from a roadmapping workshop. In: 2015 Annual Conference of the Prognostics and Health Management Society (2015).
- 3. Charniak, E.: Bayesian networks without tears. AI Magazine 12(4), 50 (1991).
- 4. De, S., Das, A., Sureka, A.: Product failure root cause analysis during warranty analysis for integrated product design and quality improvement for early results in downturn economy. International Journal of Product Development 12(3-4), 235-253 (2010).
- 5. Pradhan, S., Singh, R., Kachru, K., Narasimhamurthy, S.: A Bayesian network based approach for root-cause-analysis in manufacturing process. In: 2007 International Conference on Computational Intelligence and Security, pp. 10-14. IEEE, Harbin (2007).
- 6. Nguyen, D., Duong, Q., Zamai, E., Shahzad, M.: Fault diagnosis for the complex manufacturing system. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 230(2), 178-194 (2016).
- 7. Liu, Y., Jin, S.: Application of Bayesian networks for diagnostics in the assembly process by considering small measurement data sets. The International Journal of Advanced Manufacturing Technology 65(9-12), 1229-1237 (2013).
- 8. Cooper, G.: The computational complexity of probabilistic inference using Bayesian belief networks. Artificial Intelligence 42(2-3), 393-405 (1990).
- 9. Yang, L., Lee, J.: Bayesian belief network-based approach for diagnostics and prognostics of semiconductor manufacturing systems. Robotics and Computer-Integrated Manufacturing 28(1), 66-74 (2012).
- 10. Correa, M., Bielza, C., Pamies-Teixeira, J.: Comparison of Bayesian networks and artificial neural networks for quality detection in a machining process. Expert Systems with Applications 36(3), 7270-7279 (2009).
- 11. Guo, H., Zhang, R., Yong, J., Jiang, B.: Bayesian network structure learning algorithms of optimizing fault sample set. In: 2015 Proceedings of the Chinese Intelligent Systems Conference, pp. 321-329. Springer, Berlin Heidelberg (2015).
- 12. Chow, C., Liu, C.: Approximating discrete probability distributions with dependence trees. IEEE Transactions on Information Theory 14(3), 462-467 (1968).
- 13. Jeong, I., Leon, V., Villalobos, J.: Integrated decision-support system for diagnosis, maintenance planning, and scheduling of manufacturing systems. International Journal of Production Research 45(2), 267-285 (2007).

Ademujimi, Toyosi; Brundage, Michael; Prabhu, Vittaldas. "A review of current machine learning techniques used in manufacturing diagnosis." Paper presented at APMS 2017 International Conference Advances in Production Management Systems (APMS 2017), Hamburg, Germany. September 3, 2017 - September 7, 2017.

- 14. Gheisari, S., Meybodi, M., Dehghan, M., Ebadzadeh, M.: BNC-VLA: Bayesian network structure learning using a team of variable-action set learning automata. Applied Intelligence 45(1), 135-151 (2016).
- 15. McCulloch, W., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics 5(4), 115-133 (1943).
- 16. Barakat, M., Druaux, F., Lefebvre, D., Khalil, M., Mustapha, O.: Self adaptive growing neural network classifier for faults detection and diagnosis. Neurocomputing 74(18), 3865-3876 (2011).
- 17. Demetgul, M., Unal, M., Tansel, I., Yazıcıoğlu, O.: Fault diagnosis on bottle filling plant using a genetic-based neural network. Advances in Engineering Software 42(12), 1051-1058 (2011).
- 18. Zhao, Q., Camelio, J.: Assembly faults diagnosis using neural networks and process knowledge. In: 2007 ASME International Manufacturing Science and Engineering Conference, pp. 565-572. American Society of Mechanical Engineers, New York (2007).
- 19. Hong, S., Lim, W., Cheong, T., May, G.: Fault detection and classification in plasma etch equipment for semiconductor manufacturing e-Diagnostics. IEEE Transactions on Semiconductor Manufacturing 25(1), 83-93 (2012).
- 20. Zhang, Z., Wang, Y., Wang, K.: Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network. Journal of Intelligent Manufacturing 24(6), 1213-1227 (2013).
- 21. Yu, G., Li, C., Kamarthi, S.: Machine fault diagnosis using a cluster-based wavelet feature extraction and probabilistic neural networks. The International Journal of Advanced Manufacturing Technology 42(1-2), 145-151 (2009).
- 22. Fernando, H., Surgenor, B.: An unsupervised artificial neural network versus a rule-based approach for fault detection and identification in an automated assembly machine. Robotics and Computer-Integrated Manufacturing 43, 79-88 (2017).
- 23. Demetgul, M.: Fault diagnosis on production systems with support vector machine and decision trees algorithms. The International Journal of Advanced Manufacturing Technology 67(9-12), 2183-2194 (2013).
- 24. Xiang, X., Zhou, J., An, X., Peng, B., Yang, J.: Fault diagnosis based on Walsh transform and support vector machine. Mechanical Systems and Signal Processing 22(7), 1685-1693 (2008)
- 25. Widodo, A., Yang, B.: Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing 21(6), 2560-2574 (2007).
- 26. Hsueh, Y., Yang, C.: Tool breakage diagnosis in face milling by support vector machine. Journal of Materials Processing Technology 209(1), 145-152 (2009).
- 27. Kumar, A., Shankar, R., Choudhary, A., Thakur, L.: A big data MapReduce framework for fault diagnosis in cloud-based manufacturing. International Journal of Production Research 54(23), 7060-7073 (2016).
- 28. UCI Machine Learning Repository, http://archive.ics.uci.edu/ml, last accessed 2017/04/04.
- 29. Weiss, B., Helu, M., Vogl, G., Qiao, G.: Use Case Development to Advance Monitoring, Diagnostics, and Prognostics in Manufacturing Operations. IFAC-PapersOnLine 49(31), 13-18 (2016)
- 30. Rabiner, L .: A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2), 257-286 (1989).
- 31. Yu, J.: Multiway discrete hidden Markov model- based approach for dynamic batch process monitoring and fault classification. AIChE Journal 58(9), 2714-2725 (2012).
- 32. Boutros, T., Liang, M.: Detection and diagnosis of bearing and cutting tool faults using hidden Markov models. Mechanical Systems and Signal Processing 25(6), 2102-2124 (2011).

8

Ademujimi, Toyosi; Brundage, Michael; Prabhu, Vittaldas. "A review of current machine learning techniques used in manufacturing diagnosis." Paper presented at APMS 2017 International Conference Advances in Production Management Systems (APMS 2017), Hamburg, Germany.

September 3, 2017 - September 7, 2017

 Yuwono, M., Qin, Y., Zhou, J., Guo, Y., Celler, B., Su, S.: Automatic bearing fault diagnosis using particle swarm clustering and Hidden Markov Model. Engineering Applications of Artificial Intelligence 47, 88-100 (2016).

INCORPORATING ABSTRACTION METHODS INTO SYSTEM-ANALYSIS INTEGRATION METHODOLOGY FOR DISCRETE EVENT LOGISTICS SYSTEMS

Timothy Sprock Conrad Bock

National Institute of Standards and Technology 100 Bureau Dr, MS8260 Gaithersburg, MD 20899, USA

ABSTRACT

Analysis models, such as discrete event simulation models, are used to support design and operation of discrete event logistics systems (DELS). The time and expertise required to construct these analysis models can be significantly reduced by automatically generating them from formal models of the systems being analyzed. DELS analysis models can be constructed from system abstractions much more reliably when the system and analysis are specified at compatible levels of abstraction. Formal modeling languages, such as those used in object-orientation, make abstraction explicit, simplifying the mappings between system and analysis models and increasing reusability of the integration. In this paper, we propose fundamental abstractions for DELS and identify corresponding libraries of analysis models. These are used in a systemanalysis integration methodology that incorporates abstraction as an explicit step, providing a path to refine and extend those abstractions and model libraries to generate analysis models.

1 INTRODUCTION

Discrete event logistics systems (DELS) are dynamic systems that transform objects flowing through networks of interconnected resources (Mönch et al. 2011). These include systems such as supply chains, manufacturing systems, transportation networks, warehouses, and health care delivery systems. Traditionally, each specialized kind of DELS has been addressed with its own dedicated research and development. However, these systems share a common abstraction — products flowing through processes being executed by resources configured in a facility (PPRF) — and they appear together in integrated models of engineering enterprises. For example, production systems might integrate storage and fulfillment capabilities as well as material handling and transportation systems, and supply chains might integrate flows between warehouses, transportation systems, and manufacturing or health care facilities.

Industrial and systems engineers use a broad variety of analysis methodologies to support design and decision making in DELS. Many, if not all, of these methodologies require significant time and expertise to manually construct analysis models, such as discrete event simulation. The methodologies would be simpler and less time-consuming with automatic formulation and construction of analysis models from an independent representation of the system. Multiple analysis models could be generated from a single system model rather than constructing each analysis model by hand, as is commonly done.

Analysis models that conform to standard definitions of the system (and abstractions thereof) can be automatically generated and reused through formal system-analysis integration (SAI) methods. Most analysis models commonly developed for DELS are based on some abstraction (simplification, view) of the system. Therefore, formalized abstractions should be a cornerstone of system modeling and analysis methodologies. Reusable abstractions can be captured in system reference models and analysis libraries.

This paper seeks to harvest these abstractions, make them formal, and automate formulation of analysis models from the abstractions.

This paper presents two modeling contributions that use DELS abstractions to simplify integration of system and analysis models and make the integrations reusable in more applications. The first contribution is a modeling methodology that explicitly incorporates abstraction into a model-to-model (M2M) transformation process (Section 2). Generalization is used to formalize the relationship between system models and abstract models. This method improves the reliability of retrieving abstractions from system models. The second contribution of this paper is fundamental abstractions used for SAI in DELS (Section 3). The result is models at multiple levels of detail that support simple and accurate retrieval of abstractions from specialized system models. Examples of conforming analysis models, such as those that could support multi-fidelity simulation methods, are identified for each level of abstraction.

2 SYSTEM-ANALYSIS INTEGRATION MODELING METHODOLOGY

Manually integrating information about DELS with discipline-specific analyses, such as discrete event simulation, is time-consuming and difficult. Analysis models use concepts and formats that are inconsistent with each other and not integrated with formal system models. Model-based systems engineering methodologies provide model-to-model transformation methods to automate the translation between system models specified in standard, formal languages (Estefan 2007). System-analysis integration extends these methods by developing standard models for widely-used engineering analysis techniques and integrating those models with widely-used systems engineering modeling languages. Then standard transformations can be developed to automatically translate between these standard system / analysis models and widely-used analysis tools.

However, significant barriers exist to reusing knowledge and tools for modeling and analysis integration. There is often not enough commonality among systems or analysis tools, requiring separate transformations for each pair of system and analysis models. Knowledge relating system and analysis models is encoded in informal and manual mappings between them. This implicit mapping knowledge is also a challenge for manual analysis modeling methods. For example, how should engineers decide which analysis component to select from a simulation model library to represent a particular system object?

Relying on complex, ad-hoc mappings to automate generation of analysis models only moves complexity from the models to the mappings. One source of complexity in these mappings is that system and analysis models are often specified at different levels of abstraction. Integrating abstraction as an explicit step in the M2M methodology simplifies mappings between system and analysis models by breaking them up across multiple levels of abstraction. The analysis methodology proposed in Thiers (2014) separates the M2M transformation into two steps: one from the system model to an abstract model, and then a second to the target analysis model. The system model is mapped to the abstract model via stereotype application in the Unified Modeling Language (UML) (OMG 2015). Network models are proposed as reusable abstractions for DELS. Stereotype application, however, cannot guarantee the correctness of resulting abstraction, because stereotypes are not available for specialization to domain-specific models once they are applied, and are only supported in UML tools.

The approach proposed here uses generalization relationships to formalize the results of abstraction, rather than stereotype application. Generalization relates classes of things to broader classes that include them. For example, the class of forklifts is generalized by the class of mobile resources. Generalization enables system models to be constructed (specialized) directly from abstractions, rather than mapped to the abstraction after the system model has been constructed, as with stereotypes. The resulting system model naturally conforms to the abstraction, because the abstraction is identified as the broader class. Abstractions can be retrieved correctly and efficiently from detailed system models. Model libraries and taxonomies constructed using generalization can be extended and specialized to incorporate new specific system behaviors and any corresponding analysis models, while retaining access to higher levels of abstraction.

Generalization is supported in almost all modern programming languages, as well as UML, providing many more potential modeling platforms than stereotypes.

2.1 Modeling Method: Formalizing Mappings to Abstractions with Generalization

Frantz (1995) describes a taxonomy of abstraction methods for simulation modeling. Two of these abstraction methods, model boundary modification and model behavior modification, can be formalized using objectoriented languages. The model boundary modification method creates taxonomies of model elements related by explicit simplifications and assumptions. Object-oriented languages formalize this with generalization. The model behavior modification method aggregates components and behavior into systems and other behaviors, respectively (Frantz 1995). Object-oriented languages formalize this with composition. This paper and this section focus on model boundary modification.

Generalization formalizes the mapping between system models and their abstractions. It is shown in UML by a hollow headed arrow going from a specialized class to a more abstract one. For example, the generalization from *Warehouse* to *Node* in Figure 1 specifies that everything fitting the description of *Warehouse* will also fit the description of *Node* (a network abstraction). This formal relationship between two model elements guides construction of the transformation from the system model (of warehouses) to the abstraction (of network nodes).

Methods that transform specialized classes into abstract ones, such as conversion, up-casting, or other substitution methods, require an explicit definition of transformations from detailed properties in system models to less detailed properties in abstractions. These transformations are specified by expressions captured as mathematical equations or UML constraints. For example, consider the problem of selecting locations for new warehouses within a supply chain. Algorithms solving facility location problems (FLP) use a network formulation (abstraction) to select the best set of nodes to optimize the network configuration. Formulating this analysis requires mapping *Warehouse* to the more general class *Node*. It also requires extracting the cost of selecting each node from a more detailed cost model included in *Warehouse* (Figure 1). For example, the *fulfillmentCost* and *orderManagementCost* of warehouses are aggregated to *totalCostToServe*, which then generalizes to the *cost* of a Node.





Generalization is a method to organize things into taxonomies (classifications) by their similarity, defining specialized classes to elaborate differences within broader classes while retaining a relationship to them. Taxonomies constructed using generalization explicitly model the assumptions, extensions, and simplifications embedded in them. Things that are logically similar can be organized by generalization. For example, trucks and forklifts can be generalized to mobile resources that carry pallets, mobile resources in general, or all resources. Classes can be specialized to capture differences between specialized things. For example, machines that execute subtractive manufacturing processes can be specialized into classes of milling machines and turning machines, or further into specific brands of milling or turning machines. A model library of extensible components simplifies the construction of analysis models by reusing similar mappings and transformations. In DELS, similarities between system models reduce the number of model library components that need to be constructed in the analysis language. These similarities can also be exploited to reuse analysis models that apply to the abstraction, rather than the system model, when that level of detail is not necessary or available.

One of the key advantages of this SAI method is that system models are constructed from abstractions, rather than mapped to abstractions after modeling. The mapping is defined as part of system model construction by extending existing reference models using generalization. This approach reduces errors and inconsistencies resulting from ad-hoc abstraction mappings and reduces the time-required to verify correctness of abstractions retrieved from the system model. The abstraction used to construct the analysis model can be extracted correctly and automatically from the detailed system model.

3 ABSTRACT MODELS FOR DELS

Analysis models for DELS are formulated at multiple levels of abstraction, due to variations in the information available about the system and in desired performance of the solution method. A modeling methodology should support construction of abstract analysis models, but also provide a method to increase fidelity of those models. Methods for abstraction, refinement, and reuse of models and model components are not supported by most analysis languages, such as those used to construct simulation and optimization models.



Figure 2: A hierarchy of system and analysis models represent the real-world system at multiple levels of abstraction.

Reusability of the SAI method in the previous section relies heavily on constructing the right abstractions. If they do not apply to a broad enough class of systems, then the return on investment for abstract modeling and integration is not substantially greater than integrating each system individually. A more robust solution is to use several inter-related abstractions of DELS, rather than a single abstraction. The proposed DELS abstractions are multi-layered and designed to maximize reuse of analysis models by harvesting commonalities across this class of systems and their related analyses. Object-orientation provides a method for extending abstractions to specify domain-specific attributes and specialized analysis models and methods. Rather than having a two-step transformation with a single abstract model, a continuum of models can represent the real-world system at many levels of abstraction (Figure 2). This continuum of system abstractions to bridge the gap between the two. This section describes reusable, standard, multi-layered abstractions for DELS and some typical analysis models applicable at each level.

3.1 Network-based Abstractions

Network-based abstractions are common in DELS modeling because of their widely-understood mathematical representations, their suitability to many algorithms, and their applicability to a broad range of (abstract) analysis questions about the structure of and flows through DELS. Basic networks, flow networks, and process (or queuing) network models are foundational abstractions for DELS (Figure 3) (Thiers 2014). Each of these abstractions has libraries of analysis algorithms, such as finding shortest paths and optimal facility locations (Drezner and Hamacher 1995), determining throughput for multi-commodity flow networks (Ahuja et al. 1993), as well as service time and utilization in queueing networks (Walrand 1988, Marzolla 2010).



Figure 3: Basic networks, flow networks, and process (or queuing) networks are foundational abstractions for DELS.

3.2 DELS Behavior

Integrating system and analysis models is simpler when the systems models extend standard abstractions corresponding to widely used analysis techniques. DELS share a common abstraction for their behavior: products flowing through processes being executed by resources configured in a facility (PPRF) (Figure 4). Each element of the PPRF abstraction is elaborated with taxonomies of more detailed abstractions, such as the resource taxonomy in OZONE (Smith and Becker 1997) or the resource/process taxonomy in ISA-95 (ISA 2010) (Figure 5).



Figure 4: DELS share a common abstraction for their behavior: products flowing through processes being executed by resources configured in a facility.



Figure 5: Each element of the PPRF abstraction is elaborated with taxonomies of more detailed abstractions.

Analyzing the complex interactions inherent in DELS behavior is usually done using simulation, such as discrete event. For this class of analysis models and tools, system models constructed from a standard

model library of components specialized from PPRF make it easier to validate simulation models and replicate the results. The application of analysis methods designed for one DELS subdomain to another can be formalized by mapping the behavior of systems from the two subdomains to a common abstraction beforehand. For example, analysis methods for solving the traveling salesman problem (TSP) can be used to construct manufacturing schedules (Lenstra and Kan 1975), and methods for scheduling manufacturing job shops can be applied to scheduling nurses or operating rooms in a hospital (Pham and Klinkert 2008), (see also (Beck et al. 2003) for a comparison of vehicle routing (VRP) and job shop problems (JSP)).

Taxonomies of analysis models for specialized system behaviors should be differentiated from taxonomies of solution methods for making control decisions. For example, extensions to common analysis models, such as flexible JSP (FJSP) or capacitated VRP (CVRP), are constructed from specializations of (extensions to) abstract system behavior models (the PPRF taxonomies). Adding resource flexibility to a process, or capacity constraints to resource (system behavior), does not change the types of control decisions that need to be made (control functions), just how decision is made (analysis methods).

3.3 DELS Operational Control

Operational control manipulates the flow of tasks and resources through a system in real-time, or near real-time. A significant body of analysis models and algorithms exist to select control actions to be taken, including: 1) "Should a task be served?" (*admission*); 2) If so, "When should the task be serviced?" (*sequencing*); and, 3) "By which resource?" (*assignment*); 4) "Which process does the task require next?" (*dynamic process planning/routing*); and 5) "To which state does a resource need to be changed?" (change in *resource capacity* or *capability*). Our model of operational control links a controller, which may use analysis models to support decision making, to actuators in the base system (plant) that execute the prescribed actions (Figure 6) (Sprock 2015).



Figure 6: A model of operational control that links a controller to actuators in the base system that execute the prescribed actions.

Each control function is carried out in the base system by a specialized resource (actuator that executes the selected action) performing a specialized process (how the actuator executes the selected action). Process and resource taxonomies are extended to model control functions in the plant (Figure 7). System specifications use these model library elements to indicate *where* and *when* control decisions are required by the system. Decision support is designed separately, specifying *how* control actions are chosen.





Figure 7: Each control function is carried out in the base system by a specialized resource and specialized process.

A wide variety of specialized system resources can carry out each control function, for example, diverters, robotic arms, or humans can admit tasks and raw materials into the system. However, most analysis models do not require highly detailed specification of every resource in the system. Detailed specifications of actuators that execute control functions are mapped by generalization to common abstractions (specialized resources) that specify the control function for each class of actuators. These abstractions of actuators are used to construct analysis models.

Decision support methods for each control function can also be specialized (refined) from a library of simple methods or can be abstracted (simplified) for analysis models. For example, sequencing tasks in a queue may be specified by a simple static priority rule that can be later refined into a complex decision tree of priority rules. Every control function in the system conforms to the same definition and uses the same abstraction. A standard abstraction of DELS control functions enables consistent and correct models of control across analysis models and tools, such as discrete event simulation.

3.4 Extending and Composing DELS

DELS abstractions can be specialized to represent many kinds of DELS, reusing the libraries described in previous sections as needed. The *Process* element can be specialized into a taxonomy of basic DELS functions, *Make*, *Store*, and *Move*. These processes are allocated to specialized DELS for production, storage, and transportation, respectively (Figure 8). Allocation of a process to a DELS, such as allocating *MOVE* to *TransportationSystem*, adds an operation (function) that executes that process when the operation is invoked. The DELS must provide a behavior that implements the function by defining what process steps and required resources are used to execute that operation.

Many DELS are composed of other DELS. For example, a manufacturing facility might have production systems, such as work cells or production lines, linked by material handling or transportation systems, and buffered by intermediate storage points. Composing DELS out of existing specialized DELS, rather than defining a new monolith, allows domain-specific analysis models to be reused for functionally similar

Bock, Conrad; Sprock, Timothy. "Incorporating Abstraction Methods into System-Analysis Integration Methodology for Discrete Event Logistics Systems." Paper presented at 2017 Winter Simulation Conference, Las Vegas, NV, United States. December 3, 2017 - December 6, 201





Figure 8: The DELS abstraction can be specialized to represent many kinds of DELS. These specializations can be composed to specify complex DELS.

subsystems. Each specialized kind of DELS can be further specialized as necessary to capture domainspecific features, such as differences between manual and automated material handling systems.

4 CONCLUSIONS AND DISCUSSION

System-analysis integration methodologies for DELS automate generation of analysis models, such as discrete event simulation, from formal system models. Development and reuse of these methodologies can be simplified by explicitly incorporating abstraction. The system-analysis integration approach in this paper first models abstractions common to many system and analysis models in the DELS domain, rather than integrating individual simulation or analysis tools. This maximizes reusability of SAI methods, and any corresponding tools. On-going research seeks to extend and refine existing taxonomies to the level of detail required by detailed system models.

This paper presents two contributions to modeling systems and related abstractions that are the foundation for many analysis models. Generalization is used to formalize the relationship between system and abstract models. Multi-layer abstractions can be used to generate many different types of analysis models for a variety of specializations of DELS. Abstractions are captured in libraries of components that are linked formally by generalization and organized into taxonomies under the PPRF abstraction. System models are constructed from the abstractions by selecting appropriate library components from the taxonomy, specializing when necessary, and assembling them. Formalizing the relationship between system and abstract models allows abstractions to be retrieved from system models correctly and automatically.

Some analysis methods, such discrete event simulation, present many challenges to developing standard transformations between the abstract models and widely-used analysis tools and techniques. One significant challenge is the lack of complete, standard simulation model libraries for specifying the product, process, resource, and facility aspects of DELS. The lack of standard descriptions is also apparent when integrating representations of control functions, simulation components that carry out the control functions (actuator blocks), and representations of control behavior (control rules). One solution would be a standard specification, abstraction, or model library for simulation tools used in the DELS domain. Then analysis models could be constructed from the abstraction using nearly one-to-one mappings. On-going research is focused on identifying and resolving SAI capability gaps within and between system models, developing additional multi-layer abstraction models and taxonomies, and identifying the analysis model libraries they should be transformed to in simulation tools.

Bock, Conrad; Sprock, Timothy. "Incorporating Abstraction Methods into System-Analysis Integration Methodology for Discrete Event Logistics Systems." Paper presented at 2017 Winter Simulation Conference, Las Vegas, NV, United States. December 3, 2017 - December 6, 2017

ACKNOWLEDGMENTS

Commercial equipment and materials might be identified to adequately specify certain procedures. In no case does such identification imply recommendation or endorsement by the U.S. National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

REFERENCES

- Ahuja, R. K., T. L. Magnanti, and J. B. Orlin. 1993. Network Flows: Theory, Algorithms, and Applications. Prentice Hall.
- Beck, J. C., P. Prosser, and E. Selensky. 2003. "Vehicle Routing and Job Shop Scheduling: What's the Difference?". In ICAPS, 267-276.

Drezner, Z., and H. W. Hamacher. 1995. Facility Location. Springer-Verlag New York, NY.

- Estefan, J. A. 2007. "Survey of Model-based Systems Engineering (MBSE) Methodologies". INCOSE MBSE Focus Group 25:8.
- Frantz, F. K. 1995. "A Taxonomy of Model Abstraction Techniques". In Proceedings of the 27th Winter Simulation Conference, edited by C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, 1413-1420. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- ISA 2010. "ANSI/ISA-95.00.01-2010 (IEC 62264-1 Mod) Enterprise-Control System Integration Part 1: Models and Terminology".
- Lenstra, J. K., and A. R. Kan. 1975. "Some Simple Applications of the Travelling Salesman Problem". Journal of the Operational Research Society 26 (4): 717–733.
- Marzolla, M. 2010, June14-16. "The gnetworks Toolbox: A Software Package for Queueing Networks Analysis". In Analytical and Stochastic Modeling Techniques and Applications, 17th International Conference, ASMTA 2010, Cardiff, UK, Proceedings, edited by K. Al-Begain, D. Fiems, and W. J. Knottenbelt, Volume 6148 of Lecture Notes in Computer Science, 102–116: Springer.
- Mönch, L., P. Lendermann, L. F. McGinnis, and A. Schirrmann. 2011. "A Survey of Challenges in Modelling and Decision-making for Discrete Event Logistics Systems". Computers in Industry 62 (6): 557-567.
- OMG 2015. "OMG Unified Modeling Language (OMG UML) Version 2.5".
- Pham, D.-N., and A. Klinkert. 2008. "Surgical Case Scheduling as a Generalized Job Shop Scheduling Problem". European Journal of Operational Research 185 (3): 1011–1025.
- Smith, S. F., and M. A. Becker. 1997. "An Ontology for Constructing Scheduling Systems". In Working Notes of 1997 AAAI Symposium on Ontological Engineering, 120–127.
- Sprock, T. 2015. A Metamodel of Operational Control of Discrete Event Logistics Systems (DELS). Ph. D. thesis, Georgia Institute of Technology, Atlanta, GA.
- Thiers, G. 2014. A Model-Based Systems Engineering Methodology to Make Engineering Analysis of Discrete-Event Logistics Systems More Cost-Accessible. Ph. D. thesis, Georgia Institute of Technology, Atlanta. GA.
- Walrand, J. 1988. An Introduction to Queueing Networks. Prentice Hall.

AUTHOR BIOGRAPHIES

TIMOTHY SPROCK is an Industrial Engineer in the Systems Engineering Group under the Systems Integration Division (SID) of the Engineering Lab (EL) at the National Institute of Standards and Technology (NIST) specializing in model-based systems engineering methodologies for discrete event logistics systems (DELS). He holds a PhD in Industrial Engineering from Georgia Tech. His email address is timothy.sprock@nist.gov.

CONRAD BOCK is a Computer Scientist in the Systems Engineering Group under the Systems Integration Division (SID) of the Engineering Lab (EL) at the National Institute of Standards and Technology (NIST)

specializing in formalization of engineering modeling languages. His previous experience includes business process modeling with SAP and Microsoft. Mr. Bock leads efforts on process modeling in UML and SysML at the Object Management Group. He studied at Stanford, receiving a B.S. in Physics and an M.S. in Computer Science. His email address is conrad.bock@nist.gov.

A Study on Performance Evaluation and Status-based Decision for Cyber-Physical Production Systems

Feifan Wang, Feng Ju, Yan Lu

Abstract-In concert with advances in information and communication technology and their application to manufacturing environments, physical entities in factories are acquiring more intelligence via integration with cyber systems. This integration brings about Cyber-Physical Production Systems and leads to smart manufacturing, the next generation manufacturing paradigm. In the new paradigm, high levels of agility, flexibility, and real-time control make it possible to keep the system running efficiently and self-organized. At the same time, however, it becomes difficult in a self-organized and decentralized system to capture the system's status, evaluate the system's performance, and predict the system's future events. In this article, we suggest improvements to smart manufacturing systems where the intelligence from smart entities could be fully utilized without losing system control. To achieve this goal, a solution for integrating schedule-driven production (push systems) and event-driven production (pull systems) is proposed to optimize both material flow and information flow for manufacturing operations. For each entity in a smart manufacturing system, details of decision making are encapsulated and its status is exposed. The status-based decisions filter out unimportant information and make smart manufacturing systems looselycoupled and predictable. A simulation case study based on Devices Profile for Web Services [1] is used to illustrate the effectiveness of such an approach. The case study suggests that status-based decisions could be applied to smart manufacturing and that they can be part of an approach that balances the selforganized control with overall performance. Therefore, we can make full use of intelligent entities in lower levels of a factory while keeping the entire system under control.

Index Terms—Smart Manufacturing Systems, Cyber-Physical Production Systems, Status-based Decision, Devices Profile of Web Services

I. INTRODUCTION

The rapid development in information and communication technology (ICT) provides an opportunity to improve current manufacturing systems. Smart manufacturing systems use ICT and intelligent software applications to optimize the use of labor, material, and energy, producing customized, high-quality products for punctual delivery. Such systems can respond quickly to the changes in market demands and supply chains, relying on the advances in Cyber-Physical Systems (CPS) [2][3]. CPS are integrations of cyber components and physical components in a system to add new capabilities to the original physical systems [4][5]. CPS have broad applications [6][7]. Specifically, CPS designed to achieve automation in production are called Cyber-Physical Production Systems (CPPS) [7][8][9][10]. Many case studies have been conducted in recent years to explore how to make use of CPPS. For example, a service-oriented industrial automation system was introduced by Colombo [11]. It is a customized dynamic assembly system with workstations, conveyors, and lifters. Reboredo [12] constructed a Semantic Field Device Integration Platform in discrete manufacturing that considers the computational limitations of field devices. Some other case studies focus on Digital Factory [13] and Smart Factory [14][15][16], where smart products, smart machines, and augmented operators cooperate with each other to form a flexible system with decentralized control.

To be flexible and agile, smart manufacturing systems apply existing technologies such as radio frequency identification (RFID) and embedded computing to a new design for material flow and information flow. Improvement comes from research on ICT, system architecture, production scheduling, and related standards. Much research has been done to improve the intelligence of devices in manufacturing systems, but limited efforts have been made on the design of material and information flow to achieve performance prediction and decision making. From the perspective of production systems, the following issues need to be investigated: (1) How to leverage local optimal decisions made by self-organized entities to contribute to improving overall performance, and (2) How to predict system performance in real time by incorporating endogenous and exogenous decisions and uncertainties. This paper incorporates previous improvements to smart manufacturing systems and proposes a guideline to address these two issues. Specifically, a solution for integrating schedule-driven production (push systems) and event-driven production (pull systems) is provided to optimize both material flow and information flow for manufacturing operations. For each entity in the system, details needed for decision making are encapsulated and its status is exposed. The status-based decisions filter out unimportant information and make smart manufacturing systems loosely-coupled and predictable. A simulation case study based on Devices Profile for Web Services is used to illustrate the effectiveness of such an approach. The case study suggests that status-based decisions could be applied to smart manufacturing to balance self-organized control with overall performance. Therefore, we can make full use of intelligent entities in lower levels of a factory while keeping the entire system under control.

The rest of the paper is organized as follows: Section II introduces the technical background in the architecture

Feifan Wang and Feng Ju are with School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ 85281, USA.Feifan.Wang@asu.edu, Feng.Ju@asu.edu

Yan Lu is with National Institute of Standards and Technology, Gaithersburg, MD 20899, USA. <code>yan.lu@nist.gov</code>

Please address all correspondence to Professor Feng Ju.

of cyber-physical production systems and their operation strategies. In Section III, details are given to describe a solution for integrating push systems and pull systems, for optimizing both material flow and information flow, and for achieving a high level of manufacturing flexibility and agility. A simulation case study is conducted in Section IV to illustrate how smart manufacturing systems could be loosely coupled and predictable. Section V is devoted to conclusions and future work.

II. BACKGROUND

A. Service-oriented Architecture

SoA is an architecture style that promotes integrations of heterogeneous applications and technologies. SoA has successful applications in software engineering, where software systems are loosely coupled and reusable components act as services to the whole system. Details of each component are encapsulated. Interfaces and descriptions are published through platform- and language-independent standards. Several reasons motivate applications of SoA in manufacturing systems. SoA-based manufacturing follows along the early ideas of constructing a more flexible and reusable physical system, where a production system can be decomposed into several modules, and each module acts as a service provider. Modules collaborate with each other in a manufacturing system to realize a variety of products and handle unexpected changes. The flexibility of these machines can be fully taken advantage of when they are viewed as service providers in flexible manufacturing systems. Another motivating factor for applying SoA in manufacturing systems is to extend the cost-effective software systems integration approach in the cyber world to a cyber-physical world.

There are some existing technologies that support SoA in manufacturing systems. Devices Profile for Web Services (DPWS) [1] is one web service technology that defines a minimal set of web service implementation for device integration. DPWS enables secure messaging, discovery, description, and eventing on resource-constrained endpoints. The basic components in DPWS are clients and devices. Devices encapsulate information processes and publish their services using Web Services Description Language (WSDL) [17]. Clients get access to devices in order to acquire data or call methods. OPC Unified Architecture (OPC UA) from the OPC Foundation is another technology implementing SoA. It provides solutions via OPC UA Client and OPC UA Server for data exchange and data modeling in manufacturing systems [18].

B. Push and Pull Systems

Material flow and information flow are two important aspects that should be considered in manufacturing system design. Generally, manufacturing systems can be classified into two types with regards to material and information flow [19]. The first type is called the push system. Production activities in push systems are triggered by schedules. Schedules make demands of parts in the workstation level consistent with demands of products at the system level. The second type is pull systems, where production activities are triggered by status-changing events. In pull systems, only the last workstation in a production line receives demands of production from the system level. Operations happened in the last workstation change the work-in-process (WIP) in buffer, which triggers operations of upstream workstations. Material flows from upstream workstations to downstream workstations, while demand information flows from downstream workstations to upstream workstations. Both push and pull systems have their strengths and drawbacks. Pull systems perform well in controlling WIP and handling changes to the systems. One strength of push systems is that every activity follows a plan. The traditional way to combine push systems and pull systems is to create a push-pull interface, and make the systems partly pull and partly push. In smart manufacturing systems, more tools are available than in traditional manufacturing systems to optimize both material flow and information flow. However, they call for a new system design approach to make use of those tools to make material and information flow smoothly and efficiently.

III. OPERATIONS IN SMART MANUFACTURING SYSTEMS A. A Flat Architecture

It has been a long time since people first made use of computers to assist production. In product development domains, Computer Aided Design (CAD) was initiated in the mid 1970s. Later, Product Data Management (PDM) software was designed to manage product data like CAD drawings. In production systems, Material Requirements Planning (MRP) acts both as a method and a software application, coming into factories during the 1970s. These technologies enabled Computer-Integrated Manufacturing (CIM). CIM technologies played an important role in manufacturing this past decade. These technologies resulted from a convergence of the virtual world and the physical world in manufacturing systems [7]. They do not, however, fundamentally change how production activities are organized. In the CIM paradigm, computers usually replace humans and take over trivial work that is time-consuming but well structured.

Manufacturing systems handle two types of orders. One type (customer requests, for example) comes from outside the system. The other type originates inside the system (such as instructions to a workstation). To fulfill orders effectively, smart manufacturing calls for convergence between the cyber world and the physical world in manufacturing systems, but it differs fundamentally from traditional CIM-based manufacturing systems. Advances in ICT enable machines in a factory to request and respond to messages. When machines on the shop floor obtain the ability to collaborate and take care of more complicated duties in the production activities, the traditional rigid automation pyramid structure can be partly broken, and a flat architecture can be created. The traditional automation pyramid works when there is a lack of intelligence in lower levels. In that case, entire production processes are centrally controlled. If the system works based on a central plan, it becomes vulnerable to unpredictable variances. In smart manufacturing systems, physical equipment has corresponding cyber nodes. Sensors and RFID enable the status of equipment to be detected and the equipment to receive instructions. Enterprise functions make production orders upon customer's requests assuming a factory as a black box, while the intelligence in lower levels of a factory enables resources to be dynamically allocated to fulfill the orders.

B. A Status-based Decision

A series of decisions is required in a manufacturing system to get an order fulfilled. Smart manufacturing systems distribute decisions to different entities in different levels of the systems. With each entity understanding its environment by communicating with other related entities, it makes a local optimal decision based on that understanding. Then, the statuses of the entities are updated to support future decisions. Entities from higher levels check the statuses of lower-level entities and make decisions on task assignments. Entities from lower levels receive the assignments, find a way to finish the tasks, and then expose the updated statuses. As such, each entity from lower levels to higher levels uses its own intelligence to optimally finish its assigned tasks, and the relationship between said entities is therefore loosely coupled. The structure of the system is shown in Fig.1.

Consider a scenario when a customer places an order for a customized product. Based on the status of the factory, the customer is promised to receive the product within a lead time. The demand for the product is decomposed into demands for different kinds of parts. Production planning is then based on the capabilities and capacities of production modules in the factory. Schedules from production planning enable each part to be finished on time. The production modules receive requests to make different parts. Each part gets admitted into the right module every time the module status is changed to be available. In this way, resources inside the module are assigned to parts dynamically.



Fig. 1. A Status-based Decision C. Integration of Push - Pull Systems

In order to achieve better performances, many efforts have been made integrate push and pull systems. Traditionally, this integration is achieved by dividing a production line into two segments, one push system and one pull system. A part in a smart manufacturing system could carry its own process information. In this situation, the material flow follows the information flow, as seen in Fig.2. From the perspective of an individual module, it is a push system. Modules receive orders from schedules, and the work for each module consists of producing the required parts in time. However, the implementation inside the module follows a pull system style. Parts come into a module based on the status of said module. The processing path for each part is carried on the part and it depends on the status of workstations in the module. The process work of workstations is triggered by different events.

The material and information flows can be achieved by using advanced ICT. For each part, an RFID chip can be installed on the carrier of the part. There is an RFID writer and a reader for each module and workstation. In order to accomplish a scheduled job in time, a module allows the right part to enter the module at the right time and writes an optimal path on the RFID for the part. When the part finds its path in the production module, all the process and path information is stored in the RFID chip. Each machine reads the customized process requirements from the RFID chip and provides a proper processing service. Similarly, a conveyor reads the path information and sends the part to the right workstation.



Fig. 2. Integration of Push - Pull Systems

D. Plug-N-Play

With the support of Plug-N-Play [14], a workstation can be removed during run time from the system without reconfiguration. Similarly, a workstation can be added to the system without reconfiguration. The property of Plug-N-Play in smart manufacturing systems keeps manufacturing systems reliable. In a traditional factory, failures and maintenance schedules of machines might have a big impact on the entire manufacturing system. Unplanned failures of machines may cause impossible schedules and create a large number of WIPs. Any maintenance requires a configuration suitable with the production plan. In smart manufacturing systems, the function of Plug-N-Play can help solve these problems.

Plug-N-Play for smart manufacturing brings a lot of advantages. Via Plug-N-Play, smart manufacturing systems are able to handle unplanned variances. Since details of a module are hidden from other modules, all changes that happen to any workstation are represented by a status change. Status-based decisions from different entities keep the entire production process going normally.

E. Capability and Capacity

Higher levels decision making keeps work going to the right modules that are capable of the processes, while lower levels decision making keeps the workload in balance. Compared with traditional manufacturing systems, smart manufacturing systems enable distributed equipment in factories to share the responsibility for decision making and avoid contradicting decisions. Real-time control in discrete manufacturing becomes possible in smart manufacturing systems. Smart manufacturing systems automatically adjust their behaviors to keep load balance and optimize their performance in real time. However, the performance of a system needs to be predictable in order to make these real-time controls effective. This is motivated by the customer requirements for reliable delivery times. When a customer places an order, the smart manufacturing system should be able to give a lead time with high confidence. The more self-organized the manufacturing systems are, the more difficult it is to predict the performance of the systems. The performance can be cycle time, throughput, or WIP. If the capabilities and capacities of modules are defined as statuses, the module can work just like a single machine. Even though a module is a combination of several machines (and/or operators), it can be viewed as a single machine during the planning. If the capability and capacity of a module can be obtained based on the capabilities and capacities of the machines within the module, the scale of a smart manufacturing system can be reduced.

For each module in a smart manufacturing system, only the status of the module is exposed to higher level functions. The status of a module is calculated from its components. Fig.3 is an example of a module, which is a distributed system. The availability of different processes is based on the availability of machines in the module. The cycle time and throughput of different processes can be calculated based on the statuses of all the machines. Production planning is made based on the calculated capability and capacity of the module, and production control is realized by the self-organized module.



Fig. 3. Capability and Capacity of A Module

Plug-N-Play can be conducted for this case. Consider a situation where Machine 2 is down during the run time. First, Machine 2 is removed from the module. The module can detect the event and change its own status. Then the capability of the module is changed – Process 1 is removed, and Process 3 is changed. The system as a whole is still running normally without any modification and reconfiguration. After Machine 2 gets repaired and connected back to the module, the capacity and capability recover.

IV. CASE STUDY OF SMART MANUFACTURING SYSTMES BASED ON DPWS

In order to illustrate the task completion of smart manufacturing systems, we build a model of a distributed system to simulate a typical module in smart manufacturing systems, as shown in Fig.3. The model contains three layers: a physical system, a cyber system, and a user interface. The physical system is represented by a simulation model with three machines and three kinds of parts. The cyber system is constructed by DPWS through Java Multi Edition DPWS Stack (JMEDS) [20], which supports web service messaging, description, discovery, and evening on resource constrained devices. Several parts wait in a buffer before entering the module. The cooperation of the three machines enables the module to finish several different process tasks. Part A calls for Process 1 and requires a path from Buffer to Machine 1 and then to Machine 2. Part B calls for Process 2 and requires a path from Buffer to Machine 1 and to Machine 3. To finish Process 3, Part C goes from buffer to Machine 1 and then to Machine 2 or Machine 3. This model follows the specifications of DPWS and is programmed using Java. JMEDS is used to support DPWS in the Java environment. JMEDS provides Java classes, like DefaultDevice, DefaultClient, DefaultService and Operation, that contain basic functionalities and can be inherited to create new classes for an application.



Fig. 4. Simulation of a Module Based on DPWS

A. Simulation Model Structure

This model can be analyzed in different layers, as seen in Fig.5. The lowest layer is composed of machines and parts manufacturable by the machines. There is no endpoint in the network for parts, but parts can carry machine-readable information about their required processes. When a part enters a machine, the part triggers operations to be performed by that machine automatically. These events are recorded at the cyber system layer, and the system can monitor the status. The Java class is created to represent parts, and each part in the simulation is an instance of this class. In the lowest layer, machines read the information carried by each coming part and process the part. Additionally, machines record their status in their local databases, publish services through Web Service Description Language (WSDL), and respond to requests from clients. The status of machines includes their capacities and capabilities. The machine class in the simulation extends the DefaultDevice class in JMEDS. When a machine object in the simulation starts, a new thread is created for the object and an IP address is automatically assigned to the machine. The layer higher than the machine layer contains clients as modules. The client class extends DefaultClient class in JMEDS and represents modules. It collects data from machines by sending HTTP requests to services published by machines. A client determines which process is available based on the status of all machines in a module. It also gives the estimation of cycle time for each process. The client makes decisions for the module and gives instructions to machines. The module client encapsulates details in the machine layer and enables the module to act as a single machine. The client also exposes the status of the module to an even higher layer through an interface which constructs data readable by human users.

B. Simulation Run

- Step 1: A client is created to represent a module. The module is set to be a constant-work-in-process (CONWIP) system. The numbers of CONWIP and WIP are initialized.
- Step 2: Machine 1 is created and an IP is automatically assigned to Machine 1. A proper service is added to Machine 1, and proper operations are added to the



Fig. 5. Structure of Simulation Model

service of Machine 1. Machine 1 publishes its service to the network through WSDL.

- Step 3: Machine 1 broadcasts a hello message with meta-data of Machine 1. The client receives hello message from Machine 1 and records the meta-data. The client sets the capability of Machine 1 and indicate that it's available.
- Step 4: Similarly, Machine 2 and Machine 3 are created. The capability of Machine 2 and that of Machine 3 are set. Since both Machine 1 and Machine 2 become available, Process 1 and Process 3 become available. Since both Machine 1 and Machine 3 become available, Process 2 becomes available.
- Step 5: When the processes become available, the first part goes into the module from the buffer. The number of WIP is increased by 1. The client requests status from all the machines and finds the best path for the part. The path information is recorded on the part.
- Step 6: When the part finishes its process on Machine 1, it goes to Machine 2 or Machine 3 based on its path.
- Step 7: When the part finishes all processes and leaves the module, the number of WIP is updated. A finished part triggers a decision to take a new part into the module.

Fig.6 shows a sequence diagram describing the simulation, where only Machine 1 is included and the transfer for parts from one machine to another is omitted. The sequence diagram gives a rough illustration of several Java objects communicating with each other to execute the simulation.

C. Discussion

The case study in this section mimics the behavior of a smart manufacturing system. Issues in communication technologies and production scheduling can be explored further in the future research. For communication technologies, the reliability of existing ICT needs to be analyzed to build a robust manufacturing system, because accidental failures or delays in communication channels might happen with multiple concurrent client requests. The potential consequences

Ju, Feng; Lu, Yan; Wang, Feifan. "A Study on Performance Evaluation and Status-based Decision for Cyber-Physical Production Systems." Paper presented at 13th IEEE Conference on Automation Science and Engineering, Xi'an, China. August 19, 2017 - August 24, 2017.



Fig. 6. Sequence Diagram

of failures and delays should be evaluated. For production scheduling, we call for better measurements and control decisions. More precise methods are needed to calculate the status of a module. It is also valuable to determine the best processing sequences for parts in buffers to better balance the production.

V. CONCLUSIONS

In order to make full use of intelligent entities in lower levels of a factory and at the same time keep the entire system under control, status-based decisions are proposed to balance self-organized controls and overall system performance. Specifically, capabilities and capacities are used as statuses to describe states of machines and modules. Future work will extend CPPS to cyber-manufacturing service models. Requirements on service models will be derived from the use cases. Existing manufacturing information modeling standards and ontologies will be used as bases to model CPPS capabilities, performance, states, and interfaces.

REFERENCES

- F. Jammes, A. Mensch, and H. Smit, "Service-oriented device communications using the devices profile for web services," in *Proceedings* of the 3rd international workshop on Middleware for pervasive and ad-hoc computing, pp. 1–8, ACM, 2005.
- [2] Y. Lu, K. Morris, and S. Frechette, "Current standards landscape for smart manufacturing systems," *National Institute of Standards and Technology, NISTIR*, vol. 8107, 2016.
- [3] Y. Lu, K. Morris, and S. Frechette, "Standards landscape and directions for smart manufacturing systems," in *Automation Science and Engineering (CASE)*, 2015 IEEE International Conference on, pp. 998– 1005, IEEE, 2015.
- [4] H. Song, D. B. Rawat, S. Jeschke, and C. Brecher, *Cyber-physical systems: foundations, principles and applications*. Morgan Kaufmann, 2016.
- [5] E. A. Lee, "Cyber-physical systems-are computing foundations adequate," in *Position Paper for NSF Workshop On Cyber-Physical Systems: Research Motivation, Techniques and Roadmap*, vol. 2, Citeseer, 2006.
- [6] L. Wang, M. Törngren, and M. Onori, "Current status and advancement of cyber-physical systems in manufacturing," *Journal of Manufacturing Systems*, vol. 37, no. Part 2, pp. 517–527, 2015.
- [7] L. Monostori, "Cyber-physical production systems: Roots, expectations and r&d challenges," *Procedia CIRP*, vol. 17, pp. 9–13, 2014.
- [8] K.-F. Seitz and P. Nyhuis, "Cyber-physical production systems combined with logistic models–a learning factory concept for an improved production planning and control," *Proceedia CIRP*, vol. 32, pp. 92–97, 2015.
- [9] H. Kühnle and G. Bitsch, Foundations & Principles of Distributed Manufacturing. Springer, 2015.
- [10] Y. Lu and F. Ju, "Smart manufacturing systems based on cyberphysical manufacturing services (cpms)," in *IFAC world congress*, 2017.
- [11] A. W. Colombo, S. Karnouskos, J. M. Mendes, and P. Leitão, "Industrial agents in the era of service oriented architectures and cloud based industrial infrastructures," *Industrial Agents: Emerging Applications of Software Agents in Industry*, pp. 67–87, 2015.
 [12] P. Reboredo and M. Keinert, "Integration of discrete manufacturing
- [12] P. Reboredo and M. Keinert, "Integration of discrete manufacturing field devices data and services based on opc ua," in *Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE*, pp. 4476–4481, IEEE, 2013.
- [13] J. Wermann, E. C. Moraes, and A. W. Colombo, "A service-oriented architecture implementation in the digital factory of the university," in *International Conference on Industrial Applications of Holonic and Multi-Agent Systems*, pp. 73–83, Springer, 2015.
- [14] S. Weyer, M. Schmitt, M. Ohmer, and D. Gorecky, "Towards industry 4.0-standardization as the crucial challenge for highly modular, multivendor production systems," *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 579–584, 2015.
- [15] M. Loskyll, I. Heck, J. Schlick, and M. Schwarz, "Context-based orchestration for control of resource-efficient manufacturing processes," *Future Internet*, vol. 4, no. 3, pp. 737–761, 2012.
- [16] P. Stephan, G. Meixner, H. Koessling, F. Floerchinger, and L. Ollinger, "Product-mediated communication through digital object memories in heterogeneous value chains," in *Pervasive Computing and Communications (PerCom)*, 2010 IEEE International Conference on, pp. 199– 207, IEEE, 2010.
- [17] E. Christensen, F. Curbera, G. Meredith, S. Weerawarana, *et al.*, "Web services description language (wsdl) 1.1," 2001.
- [18] W. Mahnke, S.-H. Leitner, and M. Damm, OPC unified architecture. Springer Science & Business Media, 2009.
- [19] W. J. Hopp and M. L. Spearman, *Factory physics*. Waveland Press, 2011.
- [20] W. S. for Devices, "Stack: Ws4d-jmeds (java). http://ws4d.org/jmeds.."

SP-187

Evaluating the Effects of Cyber-Attacks on Cyber Physical Systems using a Hardware-in-the-Loop Simulation Testbed

Bradley Potteiger, William Emfinger, Himanshu Neema, Xenofon Koutosukos Institute for Software Integrated Systems Vanderbilt University Nashville, TN 37235 CheeYee Tang, Keith Stouffer National Institute of Standards and Technology Gaithersburg, MD 20899

ABSTRACT

Cyber-Physical Systems (CPS) consist of embedded computers with sensing and actuation capability, and are integrated into and tightly coupled with a physical system. Because the physical and cyber components of the system are tightly coupled, cyber-security is important for ensuring the system functions properly and safely. However, the effects of a cyberattack on the whole system may be difficult to determine, analyze, and therefore detect and mitigate. This work presents a model based software development framework integrated with a hardware-in-the-loop (HIL) testbed for rapidly deploying CPS attack experiments. The framework provides the ability to emulate low level attacks and obtain platform specific performance measurements that are difficult to obtain in a traditional simulation environment. The framework improves the cybersecurity design process which can become more informed and customized to the production environment of a CPS. The developed framework is illustrated with a case study of a railway transportation system.

Keywords

Hardware-in-the-Loop, Model Integrated Computing, Vulnerability Assessment, Resilience, Cyber-Physical System, Experimentation, Testbed

I. INTRODUCTION

Cyber-Physical Systems (CPS) include co-engineered interacting networks of physical and computational components. Such systems typically consist of embedded computers with sensing and actuation capability, and are integrated into a tightly coupled physical system. Because the physical and cyber aspects of the system are tightly coupled, cyber-security is important for ensuring the system functions properly and safely. Cyber-security is further becoming increasingly important as many CPS are becoming distributed and utilizing wired or wireless networks for communications and coordination. Such distribution enables smarter systems with increased functionality but also creates a larger attack surface.

In traditional cyber-security it can be difficult to determine how a cyber-attack will affect the running system, especially given that the same attack will most likely have different effects depending on which subset of the system is being attacked. Given how networks and communications couple the components of a system, determining how the effects of an attack propagate through the system can compound the difficulty of such analysis and of predicting attack severity. These problems are only exacerbated when such networked systems are connected to sensors and actuators which tightly couple the system to the physical world. In that case the attack's effects propagate not only through the cyber and communications parts of the system, but also through the embedded controllers and into the physical world [12] [5].

Our previous work focused on utilizing the Command and Control Wind Tunnel (C2WT) environment to develop synchronized, multi-domain CPS simulations [1]. By combining different CPS models (Network, Physics, User Interface) into a single system of systems simulation, a vulnerability assessment framework was developed to analyze impacts of cyber-attacks on different levels of a CPS. Further, an attackerdefender (Red team vs Blue team) game was implemented to aid in developing cybersecurity strategies. However, in certain cases such as platform dependent vulnerabilities, simulations are limited in their ability to predict system behavior. This is especially true for attacks such as distributed denial of service (DDOS) and code injection attacks that have platform dependent impacts.

To address the difficulties of performing impact analysis on cyber-attacks in CPS solely on simulation techniques, we have extended a hardware-in-the-loop (HIL) CPS testbed, and an associated experimentation software development and deployment platform. The HIL testbed extends the capabilities of the C2WT to emulate the CPS software on embedded computers that are representative of the future production environment.

The work at Vanderbilt is supported by NIST (70NANB13H169). No approval or endorsement of any commercial product by the National Institute of Standards and Technology is intended or implied. Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose. This publication was prepared by United States Government employees as part of their official duties and is, therefore, a work of the U.S. Government and not subject to copyright.

Realistic measurements of the system behavior in the presence of cyber-attacks and defense mechanisms can be obtained, which in turn can be fed back into the simulation environment to provide more accurate results. The testbed and software platform together allow rapid development and experimentation with a variety of CPS, ranging from networked satellites, to airplanes and cars [8] [9]. In this paper we demonstrate how the model based software development framework can be used with the testbed for rapidly deploying cyber-attack experiments for impact analysis in a safety-critical CPS such as a railway transportation network. Railway presents a critical domain in that hazardous material, shipping companies, and the general public rely on this method of transportation for timely, reliable and safe movement.

The paper is organized as follows: Section II discusses our previous research and related work, Section III details the architecture of the HIL testbed, Section IV presents a model driven experiment development software framework, Section V presents a case study based of a railway transportation system illustrating the testbed and software framework, and Section VI provides concluding remarks for the paper.

II. RELATED WORK

Cyber-security for CPS is a rapidly growing field, as researchers are demonstrating critical vulnerabilities in networked CPS such as automobiles [2], implanted pacemakers [4], [6], and home automation systems [12]. However, testing the security of networked CPS is challenging, both with regard to the difficulty and price of setting up a CPS on which to test and with regard to the possible dangers associated with the results of the experiments [12] [13] [14]. This is especially true for larger-scale safety critical CPS, such as transportation networks, power distribution networks, and commercial airlines. For such systems, a real laboratory testbed is impossible, leading to the development of simulation or emulation testbeds which involve a simulation of the CPS. The authors in [12] demonstrate the need for complete testing of CPS, as even non-networked devices such as standard lightbulbs can be compromised and pose a threat to users if a networked home automation controller becomes compromised. Such a threat pathway clearly indicates that the physical characteristics of the CPS and its environment play a critical role in threat analysis.

Since this paper presents the uses of such a testbed and how it can be used to analyze the behavior of a distributed CPS under attack, we will not cover the advances in many of the related fields associated with simulation of physical systems, hardware in the loop simulation, software platforms for experiment development, deployment and analysis, or cyber-attack detection and mitigation. This paper addresses only the goal of analyzing how cyber-attacks propagate through networked CPS through the software and into the physical domain. More details about the design of the testbed, its integration with physics simulation, and the software platform enabling rapid experimentation on the testbed can be found in [8] and [9], respectively. Security research in CPS using testbeds has been an active area of research. A good overview of their architecture, and their application in the Smart Grid domain is provided in [5]. The authors mention nine research applications of security research using CPS testbeds as: 1) vulnerability research, 2) impact analysis, 3) mitigation research, 4) cyber-physical metrics, 5) data and model development, 6) security validation, 7) interoperability, 8) cyber forensics, and 9) operator training. Many of these applications are related, but our main research application according to this classification scheme is #2, viz. impact analysis, which "explores the physical system impacts from various cyber attacks to quantify physical system impact."

Indeed other CPS testbeds have been built providing varying degrees of fidelity with respect to the CPS they are trying to analyze. Supervisory Control and Data Acquisition (SCADA) systems have been explored most extensively [10] [3] since they form the backbone of much of the critical infrastructure.

Many of these alternative platforms for security testing do not incorporate the CPS (either through simulation or in the loop), are tied to a specific system being tested, and/or are prohibitively expensive to replicate or extend in other research labs. The rest of the paper demonstrates how the HIL CPS testbed can provide similar capabilities for impact analysis without being bound to any specific CPS.

III. HIL ARCHITECTURE

To determine, measure, and analyze the effects of cyberattacks on networked CPS, we need a platform for developing tests and experiments which can detect, quantify, and measure the effects of cyber-attacks on real CPS. Such a platform needs to have real embedded hardware that would be used in the CPS, and this hardware must have a way to sense and actuate a physical system. For many systems, building and deploying the real CPS is not feasible or not possible due to financial, logistical, or safety reasons. It is traditional and acceptable to use physics simulators to act as the CPS, using hardware-inthe-loop (HIL) with simulation to provide the computational and communication capabilities of the CPS. Additionally, a complete software development and deployment infrastructure is required to enable rapid, iterative experiment design, deployment, and data collection.

The rest of the paper covers the procedures for developing and deploying security experiments using our model based software development framework. To complete the description of the experimental environment, we first describe the architecture of both the hardware and software platforms used by the HIL testbed.

A. Hardware Architecture

The functions of a CPS testbed for security research are 1) (re-)configurability with respect to CPS and software, 2) accurate behavior of the software with respect to the CPS, 3) accurate behavior of the network with respect to the CPS, and 4) accurate behavior of the sensors/actuators with respect to the CPS. In addition to these concerns, the CPS testbed should

behave similarly to the real system in the case of failures or attacks.

Because of these requirements, the CPS testbed was extended with a Hardware-in-the-Loop Emulation platform, where embedded computing nodes on multiple configurable networks are connected to one or more simulation machines. The simulation machines provide the embedded computers the ability to sense and control the (simulated) physical systems in which they will be deployed. The configurable network through which the embedded computers communicate allows more robust and higher-fidelity emulation of the CPS' network as it would be deployed than if the network were simulated. An architectural diagram of the CPS testbed is shown in Figure 1. For more information about this testbed please see [8].



Fig. 1: Diagram showing the architecture of the CPS security HIL testbed.

B. Software Platform

In addition to the hardware and simulation platform which form the backbone of the HIL CPS testbed, the software platform must provide functions for 1) developing the CPS code, 2) developing the attack and measurement code, 3) configuring the experiment, 4) deploying the experiment, and 5) retrieving the results from the experiment for analysis. To meet these requirements, the ROSMOD software toolsuite is extended [9]. Using this software platform, we can create reusable software components which each provide key functionality for our experiments. We compose these software components together into an experiment for deployment onto the HIL CPS testbed. Further details about the software framework are detailed in the following section.

IV. MODEL BASED SOFTWARE DEVELOPMENT FOR CPS SECURITY ANALYSIS

For the purpose of vulnerability analysis on the HIL testbed, a software framework is used to enhance and streamline the cyber-attack experiment creation process. ROSMOD is a graphical model integrated computing(MIC) tool that utilizes component based design principles with the Robot Operating System (ROS) middleware for representing the distributed nature of a CPS. The ROSMOD framework is extended to support the implementation of cyber-attacks within the component and communication scope for the purpose of exploiting respective vulnerabilities in a component's software. The rest of the section details the process of creating an experiment and the different capabilities that can be included.

A. Experiment Development

In respect to the hardware-in-the-loop testbed, the purpose of the software framework is to enhance and allow for rapidly prototyping experiments for measuring the effects of cyberattacks on a CPS. For accomplishing this task, experiments can be developed representing a system at baseline, as well as under attack to identify potential vulnerabilities and analyze the resiliency and fault tolerance of critical infrastructure. ROSMOD provides the ability to model the specific components of an experiment, as well as the communication network through the ROS protocol. ROS utilizes a publish-subscriber technique where each line of communication is assigned a unique identifier (ROS Topic) which can be used by subscriber nodes to filter out incoming messages. To successfully develop an experiment in the software framework, the following needs to be completed:

- 1) Development of the Software Model
- 2) Development of the System Model
- 3) Development of the Deployment Model
- 4) Execution of the Experiment Interpreter

The first step involves developing the generic component libraries representing the various parts of a CPS. The component in the software library represents a basic building block for insertion into an experiment including timer based code execution, publisher-subscriber functionality for communication, and variable placeholders for allowing components to be customized after being placed into an experiment model. Furthermore, the communication messages and their contents are defined, as well as external libraries to link to simulator application program interfaces(APIs). A compilation interpreter is included to provide the ability to cross compile the library code under the testbed node architecture(Arm V7 architecture) and identify locations of errors within the code in the graphical interface similar to a traditional compiler.

The next step includes defining the architecture of the hosts that the experiment is executed on. This includes defining attributes such as the host name, operating system architecture, and location of a secure shell(ssh) key for accessing the host. Additionally, the model specifies the connection of the host to the ROSMOD server through an internet protocol(IP) address for the purpose of transferring and executing code on the hosts and fetching the results back to the ROSMOD server.

The third step includes taking the generic component building blocks previously built in the software library, and customizing them to produce an implementation model for the experiment. This includes creating component instances and editing the component parameters appropriately through the variable placeholders. The deployment represents the finalized experiment model to be executed on the testbed.

The final step involves executing the experiment interpreter which maps the deployment components to the appropriate host nodes declared in the system model (HIL embedded computers), transfers each respective component binary to the appropriate host, and starts the component processes on the HIL testbed nodes. Additionally, this is the time where the user starts up the physics simulator for synchronization with the respective component processes. Furthermore, after the simulation is complete, the results are fetched from the respective hosts and transferred to the server where they can be analyzed in the ROSMOD graphical environment.

B. Experiment Capabilities

The following features can be included in experiments using the development process described above.

1) Baseline: For the purpose of measuring the effects of attacks as well as the impact of defense mechanisms, the first place to start is with a baseline experiment under normal operating circumstances. In this case, the CPS is implemented as it would operate with no external impacting factors, as in how the operators expect the system to behave on a routine day to day basis. This allows the operators to have data to compare against when trying to decide whether an attack is occurring and have a data-centric method of determining the best course of action to mitigate the situation.

2) Attack Implementation: To effectively identify the most critical vulnerabilities of a system, it is important to test the system under as many circumstances as possible, including under attack. By developing experiments with individual attack implementations, the attack effects on system specific metrics are measured to appropriately quantify the impact of the attack and in turn the criticality of the system vulnerability. This aids the risk assessment process in allowing engineers to have an objective, quantitative metric to judge various vulnerabilities. Various attacks are implemented to effect the confidentiality, integrity, and availability of a CPS. Confidentiality attacks that are implemented include packet sniffing, password/authentication, session hijacking, and side channel attacks. Integrity attacks include replay, spoofing, packet delay, and packet dropping attacks. Availability attacks include denial of service, distributed denial of service, and spamming attacks. Additionally, attacks are implemented that include multiple categories such as buffer overflow and code injection attacks.

3) Attack Linking: To make attacks more realistic, the software framework provides the feature of staging multiple attacks to occur in an experiment. As such, multiple attacks can be linked together to identify propagating vulnerabilities in a system and maximize the impact of an attacker. All of the attacks described above can be combined into a single experiment implementation. Attacks can be modeled both occurring simultaneously and occurring in stages. For the simultaneous category, attacks that are independent can be implemented within the software model with no corresponding communication with one another. For example, with an experiment including a denial of service attack on a respective component and spoofing attack on a separate part of the

network, the two attacks are not correlated to one another and can be implemented in software as occurring through the entire simulation with no communication or timer dependencies. However, it is often the case that attacks are dependent on one another. As such, attacks can be modeled as being dependent on timer information, event information, or both. For example, by accessing timer data in the case of the denial of service and packet spoofing attack, the packet spoofing attack can be deployed to start at the beginning of the simulation while the denial of service attack can be programmed to start at 200 seconds into the simulation. Additionally, with a case such as a replay attack and packet sniffing attack, the replay attack will be dependent on the success of the packet sniffing attack capturing a transmitted packet. As such, event based dependencies can be programmed into the software model by having communications between the respective attacks. For example, once the packet sniffing attack is successful, an event message can be transmitted to the replay attack implementation including the captured packet information. Once this message is received, the replay attack sequence can be initiated and executed.

4) Defense Mechnamisms: The most important benefit to system engineers is the ability to implement and analyze the effects of defense mechanisms on the security and functionality of the CPS. The insertion of a defense mechanism often produces a tradeoff (e.g., performance vs. security) where the benefit has to be weighed against the cost. By measuring the performance overhead of the defense mechanism on the testbed while analyzing the success of preventing the targeted attacks, engineers can make the crucial cost benefit decision when designing their systems. Engineers can implement defense mechanisms dealing with hardening communications (AES 256 message encryption, message authentication, firewall implementations, dynamically changing ip addresses), hardening the component device software (password authentciation, input buffer limits, variable scope enforcement), or implementing monitoring algorithms for detecting abnormalities (threshold detection, gaussian processes, unsafe state detection).

5) Data Collection: After the conclusion of an experiment simulation, it is important to analyze the results to determine the behavior of the system. As such, the software framework provides the ability to fetch data results to analyze through the graphical interface. During every experiment, event times are logged in the host repositories and transferred back to the ROSMOD server at the end. By default these events include the timer based execution of code, publishing of messages, and receipt of subscribed messages. Additionally, custom log messages can be inserted to record experiment-specific events such as when a state is changed in a controller. In the graphical interface, the data can be downloaded in text files, but events can also be observed through time domain plots, allowing for rapidly analyzing the sequence of events in the experiment. Furthermore, the user can generate data from the physics simulator to record the consequential behavior in the physical environment. This allows for analyzing the cyber effects of

the experiment on the physical environment.

V. CASE STUDY

The reference case study is based on a railway transportation system. In this example, there are many railway signals and switches that route trains throughout the rail network. Railway signals have a green or red state and determine whether a train can travel to the next rail segment. In circumstances where a junction exists that connects multiple rail segments, rail switches are used to route trains to the appropriate adjacent rail segment. Each rail switch or signal is controlled by command messages sent through a communications network by a train operator located at a central facility. The communication network is comprised of network switches, routers, and basestations. The communications from the train operator are first transmitted through ethernet to a network switch and then passed to a router which transmits packets wirelessely to basestations in the field. Basestations then filter out and relay command packets to the appropriate associated rail signals and switches at the location.



Fig. 2: The railway communication network

For the CPS cyber-attack experiment, a critical rail segment is selected for attack which serves as a central hub to the rest of the network. The corresponding rail segment as well as its relationship in the communications network and connection to the railway network simulated in Train Director [7] are shown in Figure 3. This example consists of ten rail signals, six rail switches, four basestations, two routers, a network switch and a train operator. These components from the architecture described in Figure 2 are mapped to individual nodes in the HIL testbed for execution of the experiment.

A. Experiment 1: Packet Delay Attack

The method of attack on the railway network consists of a man in the middle attack on the communications to the Basestation 1 software component. The railway network is comprised of actuators (railway signals and switches), as well



Fig. 3: Train Director Railway Network Configuration and Communication Network Relationship

as communication infrastructure (basestations, routers, network switches) that relay command messages from a train operator located at a central facility. The wired network between the train operator, network switch, and router is encrypted and authenticated. Additionally, the wireless communication between the routers and basestations has cryptographic encryption but no authentication, preventing an attacker from spoofing command packets. However, this doesn't prevent an attacker from inserting a malicious node to delay the routing of command messages through the network.

In this experiment, command packets (ROS Messages) from the train operator ROS node consist of an actuator name and goal state string variable. The attacker creates a malicious man in the middle node and intercepts traffic routed through Basestation 1. As such, the attacker has the ability to delay when command packets are transmitted to the Rail Signal 1, Rail Signal 2, and Rail Switch 1 components. The attacker uses this vulnerability to execute a packet delay attack, delaying command packets by 20 minutes to have a sufficient effect on the railway network to keep trains from reaching their destinations on time, but at the same time, decreasing suspicion of a denial of service attack or faulty components by keeping components operational. A diagram illustrating this experiment is shown in Figure 4.

1) Experiment Configuration: Configuration of the experiments entails:

- 1) Configuring the communications network between the software components,
- 2) Mapping the software components to simulated CPS sensors/actuators as required,
- Configuring the parameters of the attack, e.g., amount 3) of time to delay command messages,
- 4) Mapping the CPS software components onto the processing hardware of the testbed,
- 5) Mapping the attacking software components onto the

September 18, 2017 - September 22, 2017.

Emfinger, William; Koutsoukos, Xenofon; Neema, Himanshu; Potteiger, Bradley; Stouffer, Keith; Tang, CheeYee. "Evaluating the Effects of Cyber-Attacks on Cyber Physical Systems using a Hardware-in-the-Loop Simulation Testbed." Paper presented at International Symposium on Resilient Communications Systems, Resilience Week, Wilmington, DE, United States.



Fig. 4: The attack vector used against the railway communication network

processing hardware of the testbed, and

6) Configuring the simulator (Train Director) with the proper rail network and communications interface

The ROSMOD software infrastructure described previously enables all of this configuration from its graphical user interface(GUI), as well as deployment, startup, and shutdown of the experiment. For the experiment, the malicious man in the middle node delays packets for 20 minutes from the time the message was received. There are also rail signals, rail switches, basestations, routers, network switches, and a train operator component in the deployment. Each component is mapped to its own computing node.

Further, the experiment development process using the software framework includes developing a software model, developing a system model, developing a deployment model, and executing the experiment by running the experiment interpreter. To develop the software model, component specific code is written to serve as the basic building blocks for designing the CPS topology. Additionally, external libraries are linked into the framework for accessing simulator APIs. To develop the system model, the HIL testbed architecture is defined including the network connectivity, hostname, operating system, ssh key, and IP address for each node in the testbed. To develop the deployment model, instances of generic component models from the software library are customized to create the experiment implementation model. This includes designing the the CPS topology, as well as implementing cyber-attack instances. To run the experiment, an interpreter is executed to map the deployment model to associated nodes in the HIL testbed, transfer the compiled component binaries to the appropriate hosts, start component processes, run the simulation, and fetch results at the end of execution.

2) *Results:* From the experiment configuration explained above, we want to determine the effects that the packet delay attack (which is a cyber-attack) has on the train arrival times in the network. For the physical system, which is simulated in Train Director, we directly use Train Director to measure

the effects of the attack. The attack alters the behavior of the rail signal 1, rail signal 2, and rail switch 1 components and we wish to determine how these changes affect the train delay in the railway network. We directly capture the state of the railway network from Train Director represented by the time the train arrived late to its destination.

Determining and measuring the effects of the packet delay attack on the software's behavior is not difficult, given that the behavior and structure of the software component are known. However, given the coupling between the software and the physical system, it is not as easy to determine what the effects of the altered behavior will be on the overall CPS as a whole. Clearly, the delayed reception of the train operator command messages will cause any trains at the affected location to be delayed. However, this alteration in behavior does not lend itself directly to prediction of the state of the train delay since it is difficult not only to quantify but to relate to the train routes. With the physical simulator (Train Director) we are able to see how the trains respond to the degraded behavior of signal and switch actuators and measure the effects on the train arrival times. Figure 5 shows the results of the Train Director simulation in respect to the lateness of the respective trains. As can be observed from the figure, the packet delay attack caused considerable delay to a number of the trains. Additionally, the trains that were not on the path were not affected by the attack. Therefore, if an attacker wanted to limit his exposure while maintaining the goal of delaying a critical shipment on this route, this attack could suffice. Furthermore, the delay of trains 4 and 5, which follow the path of train 2, is compounded by the time it takes train 2 to pass through the attacked signals and switches.





In addition to implementing a single attack in an experiment, the software framework provides the ability to create an experiment with multiple attacks. As such, the first experiment in which a packet delay attack is implemented is extended to include a second attack: a denial of service attack on a set of railway switches. In this instance the goal of the attacker is to not only delay trains but to prevent the trains not on a specific path (trains coming from railway signal 1) from being able to reach their destination. As such, the critical shipments on these trains will not reach their customers causing a detrimental effect on the receiving companies operating status. To execute this attack series, the packet delay attack will consist of a malicious man in the middle node effecting the rail signal 1, rail signal 2, and rail switch 1 components while the denial of service attack will focus on compromising the rail switch 2 and rail switch 4 components. Due to the fact that rail switch 2 and rail switch 4 are no longer functional, trains originating from the top and bottom of the map will be stuck and will not be able to travel to their adjacent rail segments.

1) Experiment Configuration: The experimental setup is consistent with the same process followed in Experiment 1. As such, the same software component library and system model are used. However, when developing the deployment model, the user configuration section of the rail switch 2 and rail switch 4 components are customized to represent a denial of service attack in addition to the changes already made to the basestation 1 communications path from the previous experiment. In this experiment the compromised components are a malicious man in the middle node, the rail switch 2 component, and rail switch 4 component. This deployment setup is then transferred to the HIL Testbed hosts for execution through the experiment interpreter.

2) Results: The Experiment 2 results are similar to Experiment 1 in that a set of trains are significantly delayed due to the attacks. In the packet delay attack example, train 1 and 3 arrived on time while the rest of the trains arrived to their destinations late. In Experiment 2, trains 1 and 6 arrived on time while the rest of the trains appear significantly delayed. This can be observed from Figure 6. However, it is important to note that even though there is significant train delay in the plot, this delay is actually unbounded. This experiment was run under a 200 minute experiment time, and therefore the delay is limited to that restriction. In reality, due to the fact that trains 2, 3,4, and 5 all have to travel through paths passing through rail switch 2 or rail switch 4, these trains will never reach their destinations due to the denial of service attacks on the respective rail switches. Therefore, as the execution time for this experiment is increased, the output train delays will subsequently increase.



Fig. 6: Multiple Attack Experiment Train Delay VI. CONCLUSIONS

In this work we've shown how a HIL CPS testbed can be used in security research for determining, measuring, and

analyzing how cyber-attacks affect CPS systems and how the attack propagates through the software into the physical system. We described the architecture of such a testbed, including its hardware and software platforms and developed an example security experiment for a railway network. By measuring both the software behavior and the physical system behavior during both normal operations and under attack, we showed how the coupling between software and the CPS can be measured. This work is part of the first steps towards modeling, analyzing, and predicting CPS behavior during cyber-attacks, and by providing a testbed for systematically running experiments and collecting data for attacks, models and analysis techniques can be developed.

REFERENCES

- [1] KOUTSOUKOS, X., NEEMA, H., MARTINS, G., BHATIA, S., JANOS, S., STOUFFER, K., TANG C.Y., CANDELL, R. Performance Evaluation of Secure Industrial Control System Design: A Railway Control System Case Study. In 4th International Symposium on Resilient Cyber Systems (2016), Chicago.
- CHECKOWAY, S., MCCOY, D., KANTOR, B., ANDERSON, D., SHACHAM, H., SAVAGE, S., KOSCHER, K., CZESKIS, A., ROESNER, F., KOHNO, T., ET AL. Comprehensive experimental analyses of automotive attack surfaces. In USENIX Security Symposium (2011), San Francisco.
- [3] DAVIS, C., TATE, J., OKHRAVI, H., GRIER, C., OVERBYE, T., AND NICOL, D. Scada cyber security testbed development. In Proceedings of the 38th North American power symposium (NAPS 2006) (2006), pp. 483-488
- [4] GOLLAKOTA, S., HASSANIEH, H., RANSFORD, B., KATABI, D., AND They can hear your heartbeats: non-invasive security for FU. K. implantable medical devices. ACM SIGCOMM Computer Communication Review 41, 4 (2011), 2-13.
- [5] HAHN, A., ASHOK, A., SRIDHAR, S., AND GOVINDARASU, M. Cyberphysical security testbeds: Architecture, application, and evaluation for smart grid. IEEE Transactions on Smart Grid, 4.2(2013):847-855.
- [6] HALPERIN, D., HEYDT-BENJAMIN, T. S., RANSFORD, B., CLARK, S. S., DEFEND, B., MORGAN, W., FU, K., KOHNO, T., AND MAISEL, W. H. Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses. In Security and Privacy, 2008. SP 2008. IEEE Symposium on (2008), IEEE, pp. 129-142.
- [7] Train Director Raiload Simulation. http://www.backerstreet.com/traindir/ en/trdireng.php Accessed on April 07, 2017
- [8] KUMAR, P., EMFINGER, W., AND KARSAI, G. Testbed to simulate and analyze resilient cyber-physical systems. In Rapid System Prototyping, 2015. RSP '15. (October 2015).
- [9] KUMAR, P., EMFINGER, W., KULKARNI, A., KARSAI, G., WATKINS, D., GASSER, B., AND ANILKUMAR, A. ROSMOD: a toolsuite for modeling, generating, deploying, and managing distributed real-time component-based software using ROS In Electronics 5.3(2016):53.
- [10] MALLOUHI, M., AL-NASHIF, Y., COX, D., CHADAGA, T., AND HARIRI, S. A testbed for analyzing security of scada control systems (tasscs). In Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES (2011), IEEE, pp. 1-7.
- [11] MITROPOULOS, D., KARAKOIDAS, V., LOURIDAS, P., GOUSIOS, G., AND SPINELLIS, D. Dismal code: Studying the evolution of security bugs. In Proceedings of the LASER Workshop (2013), pp. 37-48.
- [12] OLUWAFEMI, T., KOHNO, T., GUPTA, S., AND PATEL, S. Experimental security analyses of non-networked compact fluorescent lamps: A case study of home automation security. In Proceedings of the LASER 2013 (LASER 2013) (Arlington, VA, 2013), USENIX, pp. 13-24.
- [13] VAN LEEUWEN, B., URIAS, V., ELDRIDGE, J., VILLAMARIN, C., AND OLSBERG, R. Cyber security analysis testbed: Combining real, emulation, and simulation. In Security Technology (ICCST), 2010 IEEE International Carnahan Conference on (2010), IEEE, pp. 121-126.
- [14] VAN LEEUWEN, B., URIAS, V., ELDRIDGE, J., VILLAMARIN, C., AND OLSBERG, R. Performing cyber security analysis using a live, virtual, and constructive (lvc) testbed. In Military Communications Conference, 2010-MILCOM 2010 (2010), IEEE, pp. 1806-1811.

Emfinger, William; Koutsoukos, Xenofon; Neema, Himanshu; Potteiger, Bradley; Stouffer, Keith; Tang, CheeYee. "Evaluating the Effects of Cyber-Attacks on Cyber Physical Systems using a Hardware-in-the-Loop Simulation Testbed." Paper presented at International Symposium on Resilient Communications Systems, Resilience Week, Wilmington, DE, United States.

September 18, 2017 - September 22, 2017.

Maturity models and tools for enabling Smart Manufacturing Systems: comparison and reflections for future developments

Anna De Carolis¹, Marco Macchi¹, Boonserm Kulvatunyou², Michael P. Brundage², Sergio Terzi¹

¹ Politecnico di Milano, Via Lambruschini 4/b, Milano, Italy ² National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA

Abstract. One of the most exciting new capabilities in Smart Manufacturing (SM) and Cyber-Physical Production Systems (CPPS) is the provisioning of manufacturing services as unbundled "apps or services", which could be significantly more flexible and less expensive to use than the current generation of monolithic manufacturing applications. However, bundling and integrating heterogeneous services in the form of such apps or composite services is not a trivial job. There is a need for service vendors, cloud vendors, manufacturers, and other stakeholders to work collaboratively to simplify the effort to "mixand-match" and compose the apps or services. In this regard, a workshop was organized by the National Institute of Standards and Technology (NIST) and the Open Applications Group Inc. (OAGi), with the purpose to identify through parallel sessions - technology and standard needs for improving interoperability and composability between services. The workshop was organized into five working session. This paper documents evidences gathered during the "Smart Manufacturing Systems Characterization" (SMSC) session, which aims at establishing a roadmap for a unify framework for assessing a manufacturer's capability, maturity and readiness level to implement Smart Manufacturing. To that end, the technology maturity, information connectivity maturity, process maturity, organizational maturity, personnel capability and maturity, have been identified as critical aspects for Smart Manufacturing adoptions. The workshop was culminating at providing a coherent model and method for assisting manufacturing companies in their journey to smart manufacturing realizations. This paper shows three different maturity models and tools that, thanks to their complementarity, enable to reflect on the different perspectives required by SMSC. These models and tools are usable together for assessing a manufacturing company's ability to initiate the digital transformation of its processes towards Smart Manufacturing. Therefore, based on their comparison, the ultimate purpose of the research is to come up with a set of coherent guidelines for assessing a manufacturing system and its management practices, for identifying improvement opportunities and for recommending SM technologies and standards for adoption by manufacturers.

Brundage, Michael; Carolis, Anna; Kulvatunyou, Boonserm; Macchi, Marco; Terzi, Sergio. "Maturity models and tools for enabling Smart Manufacturing Systems: comparison and reflections for future developments." Paper presented at IFIP 14th International Conference on Product Lifecycle Management, Seville, Spain. July 9, 2017 - July 12, 2017. Keywords: Smart Manufacturing Systems Characterization; maturity model; manufacturer's capabilities; Industry 4.0; Smart manufacturing readiness

1 Introduction

With the introduction of the Smart Manufacturing (SM) concept, manufacturers are faced with a plethora of technologies that envision different ways to improve their manufacturing systems. Thus, many stakeholders, such as governments, consulting companies or research institutes, have provided their own definitions of SM building on such technologies. Among the definitions, for example, PwC identifies eleven digital technologies that enable the SM or Industry 4.0 framework [1]: mobile devices, IoT platforms, location detection technologies, advanced human-machine interfaces, authentication and fraud detection, 3D printing, smart sensors, big data analytics and advanced algorithms, multilevel customer interaction and customer profiling, augmented reality/wearables and cloud computing. It is interesting to point out the wide scope of applications, i.e. value chains, business models, products and services: enabled by the technologies, SM focuses on the end-to-end digitization of all physical assets and integration into digital ecosystems with value chain partners [1]. It is in fact possible to state that manufacturers' adoptions of digital technologies will have three key impacts [2]: (i) the overall product and asset lifecycles, both within and outside the factory; (ii) the entire ecosystem manufacturing systems, thanks to the full interoperability of systems based on shared vision and standards and to their ability to defy rigid standardized hierarchies but create dynamic structures from their articulated functions; (iii) the ability to create new value-added services for customers and operators from the utilization, production and design processes.

However, this leads to the rise of many new complex enterprise challenges [3][4][5]. Indeed, the integration of so many different disciplines could lead to an increased complexity of the whole SM system, which might limit the obtainable advantages from the digital transformation. Therefore, a company should be ready to cope with this complexity. For this reason, we assume that the impact of the digitalization process on a manufacturing company in terms of obtainable opportunities may differ depending on the maturity level of that company's capabilities. This means that, before starting the transformation process towards SM or Industry 4.0, manufacturing companies should define their transformation roadmap according to the actual maturity level of their capabilities [6]. This requires proper methodologies for maturity assessment applied to SM, with the aim to support companies finding their own way towards the digital transformation, i.e. helping manufacturers in understanding their current capabilities and so their needs for undertaking the digital transformation. Even though such methodologies are emerging, there is no established approach or framework to this end. This paper describes three different but complementary tools for analyzing the maturity level of manufacturing systems and environments within a SM perspective, i.e. DREAMY (Digital REadiness Assessment MaturitY model), SMSRL (Smart manufacturing readiness level) and MOM (Manufacturing Operations Management). Based on their comparison, the future work of this research is to come up with a coherent guidelines for maturity assessment to support the transition towards SM.

2 Smart Manufacturing Systems Characterization and Maturity models and tools

2.1 Smart Manufacturing Systems Characterization as a new concept

SM means many things in terms of applications, e.g. new info technologies in supply chains, in product development, in business to shop floor integration, in the development of smart products, in production equipment [7]. Then, SM is really a convergence of technologies and related capabilities brought from multiple areas and multiple business lifecycles. In order to help and guide manufacturers coping with such a complex system, the Smart Manufacturing Systems Characterization (SMSC) was introduced as a new concept. Indeed, based on the work carried out during NIST/OAGi Workshop 2016 [8], it can be stated that SMSC is defined as the enabler of unbiased models, tools, norms or guidelines to understand and analyze manufacturing systems and environment with an Industry 4.0 perspective, to the final aim of prioritizing the investments in the new technologies manufacturers might launch. This, in turn, helps building an approach whose purpose is to support the identification of opportunities for improvement of the manufacturing systems through SM technologies [8]. In other words, SMSC serve for assessing a manufacturing system and its management practices, for identifying improvement opportunities and for recommending SM technologies and standards for adoption by manufacturers [8].

As overall assumption, the manufacturers need to adopt a progressive introduction of the SM applications, systems, and hardware based on a composition of different technologies [6]. Thus, the SMSC is an essential driver: we consider that the introduction of new technologies depends on understanding the actual readiness of the manufacturer to deploy the new technologies in its manufacturing system(s); this should be assessed/re-assessed to master the maturation process towards SM. In particular, SMSC methods are focused on the assessment of manufacturer's capabilities, and readiness level to implement SM technologies and applications. To this aim, a maturity model appears a relevant "tool".

2.2 Overview on maturity models and tools

In order to understand what maturity models are, here the basics concepts of maturity models are given. To this aim, it is appropriate to provide some definitions since the concept of maturity adopted varied, even within one field of expertise [9].

Maturity can be defined as "the state of being complete, perfect or ready" [10][11][12]. Another slightly different perspective on the concept of maturity is the one given by Maier et al. [9], who stated that the process of bringing something to maturity means bringing it to a state of full growth. In other words, maturity implies an evolutionary progress from an initial to a desired or normally occurring end stage

[13]. This last definition, which stresses the process toward maturity, introduces another important concept, which is the one of *stages of growth* or *maturity levels*.

Before reaching a state of "full growth", an entity (an organization as well as a human being) has to encounter different stages of growth or maturity levels. In particular, the stages an organization passes through have three main distinctive elements [14]: (i) they are sequential in nature; (2) they occur in a hierarchical progression that is not easily reversible; (3) they involve a broad range of organizational activities and structures. To this end, we can state that maturity models can be used as tools for determining manufacturers' readiness level and capabilities also within a SM perspective.

Maturity models in literature have different characteristics: they can be of moderate or high complexity, maturity levels can be described in simple or complex terms, and so on. To this end, Fraser et al. (2002) [16] presented a first clear classification per typology of maturity models. In particular, they distinguish three types of maturity models [16]: (1) Maturity grids; (2) Likert-like questionnaires; (3) CMM-like models.

The maturity grids typically illustrate maturity levels in a simple and textual manner, structured in a matrix or a grid. As Fraser et al. (2002) stated, they are of a moderate complexity and they do not specify what a particular process should look like; they only identify some characteristics that any process and every enterprise should have in order to reach high performance processes [9]. On the other hand, the Likert-like questionnaires are constructed by "questions", which are no more than statements of good practices. The responder to the questionnaire has to score the related performance on a scale from 1 to n. A hybrid model can be defined as a combination of the questionnaire approach with the maturity grid definition[16]. Finally, the CMM-like models (Capability Maturity Model) identifies the best practices for specific processes and measures the maturity of organizations in terms of how many practices are implemented [9]. Their architecture is more formal and complex compared to the first twos. They are composed of process areas organized by common features, which specify a number of key practices to address a series of goals. Typically, the CMM-like models exploit Likert questionnaires to assess the maturity. These models have been improved successively by the Capability Maturity Model Integration (CMMI) [17].

Although a number of different types of maturity models have been proposed in literature, they share some common proprieties, which are shown below [6] [16]: (i) Maturity levels (typically from three to six); (ii) A "descriptor" for each level, which gives a meaningful name to each level; (iii) A generic description of the characteristics of each level; (iv) A number of dimensions or "process areas"; (v) A number of elements or activities for each process areas; (vi) A description of each activity, that has to be performed at each maturity level.

The terms 'readiness' and 'maturity' are relative and related. To this end, we define the term 'smart manufacturing readiness' as the capability or maturity of a manufacturing company 'to' deploy smart manufacturing concepts, and the term 'smart manufacturing maturity' as how well a manufacturing company has employed smart manufacturing concepts or its smart manufacturing capability. To that respect,

some maturity models can be viewed as part of smart manufacturing readiness assessment such as the manufacturing operation management (MOM) maturity; and an example of a smart manufacturing maturity model is the Industrie 4.0 Readiness [25] (although calling itself readiness, it is more of a smart manufacturing maturity). In the following chapter, three different tools for assessing manufacturers' readiness or maturity levels 'to' implement SM concepts are described.

2.3 DREAMY (Digital REadiness Assessment MaturitY model)

The Digital REadiness Assessment MaturitY model is a tool with two main objectives. Firstly, it is aimed to assess a manufacturing company readiness level for starting the digital transformation process, which is an aspect of smart manufacturing concepts. For this reason, according to their main objective [10–12], it has the form of a maturity model based on the inspiring principles of the CMMI framework [18,19]. Secondly, it is a tool for identifying manufacturing companies strengths and weaknesses and related opportunities they can gather from the digital transformation, with the final aim to help manufacturers in defining a roadmap for prioritizing investments [6].

ML 1 Initial	The process is poorly controlled or not controlled at all, process management is reactive and does not have the proper organizational and technological "tools" for building an infrastructure that will allow repeatability / usability / extensibility of the utilized solutions.	
ML2 Managed	The process is partially planned and implemented. Process management is weak due to lacks in the organization and/or enabling technologies. The choices are driven by specific objectives of single projects of integration and/or by the experience of the planner, which demonstrates a partial maturity in managing the infrastructure development.	
ML3 Defined	The process is defined with the planning and the implementation of good practices and management procedures. The management of the process is limited by some constraints on the organizational responsibilities and / or on the enabling technologies. Therefore, the planning and the implementation of the process highlights some gaps/lacks of integration, information exchange, and ultimately interoperability between applications.	
ML4 Integrated and interoperable	The process is built on information exchange, integration, and interoperability across applications; and it is fully planned and implemented. The integration and the interoperability are based on common and shared standards within the company, borrowed from intra- and/or cross-industry de facto standards, with respect to the best practices in industry in both perspectives of the organization and enabling technologies.	
ML5 Digital- oriented	The process is digital oriented and is based on a solid technology infrastructure and on a high potential growth organization, which supports – through pervasive integration and interoperability – speed, robustness and security in information exchange, in collaboration among the company functions and in the decision making.	

Table 1. DREAMY Maturity levels' definition (taken from [20])

To define the DREAMY architecture, it was fundamental to identify the manufacturing relevant processes, within which value-added activities are performed, and that are strategic for the digital transformation [20]. In order to make the

architecture as general as possible, manufacturing company's processes were grouped in five main areas: 1) Design and Engineering; 2) Production Management; 3) Quality Management; 4) Maintenance Management; 5) Logistics Management. Each process area can be considered as a self-contained module and therefore it is possible to add or remove one or more areas in case they are not meaningful in certain industrial situations. Cutting-across to these process areas is the Digital Backbone, within which all the information exchange processes across the process areas are considered [20]. The digital readiness of a manufacturing company is then defined through a scale of maturity levels. These levels describe a proper set of company capabilities, to provide a snapshot of their current abilities. The levels have been based on the inspiring principles from the CMMI framework [19] [18]. In this way, as the five-scale CMMI maturity levels provided a generic model to start from, they have been re-adapted in order to gather the definitions, and so the semantic, of the digital readiness levels [20] (see **Table 1**).

From what defined in the maturity levels, it is clear that, when evaluating the capabilities of a company, not only the technologies used to support the processes have to be considered [20]. From these evidences and considering the objective of the maturity model itself, it was decided to evaluate the digital readiness of manufacturing companies through four analysis dimensions (equivalent to four aspects of analysis): *Process, Monitoring and Control, Technology*, and *Organization* [20].

The DREAMY model is synthesized in Fig. 1.



Fig. 1. DREAMY (Adapted from [20])

The model in its current form can be used for descriptive purposes. That is maturity indexes for each process can be calculated to reflect the as-is situation of a manufacturing company [20]. With further analysis, strengths, weaknesses, and opportunities (prescription [6]) for smart manufacturing concept adoptions can be derived. Going forward the model can be enhanced such that such information can be automatically generated form the model To date, the "factory" is the unit of the analysis considered by the model. However, thanks to its modular structure, future works will be done to include other value-added process areas such as Supply Chain
Management, Sales, Marketing, Customer care, Human Resource Management, etc. in order to extend the scope of the analysis. In addition, due to the high relevance of the topic, which is also seen as one of the enabler of manufacturing companies' digital transformation, also *Skills of Personnel* should be considered as analysis dimension when assessing company capabilities.

2.4 SMSRL

Smart manufacturing readiness level (SMSRL) is an index that measure a manufacturing company's readiness for employing smart manufacturing concepts with the assumption that smart manufacturing is an intensive use of information and communication technologies to improve performances [21]. For this reason, SMSRL bases its readiness model on the factory design and improvement (FDI) processes, i.e., an IDEF0 activity model [22,23]. FDI consists of four high-level activities as shown in **Fig. 2**. Each activity has one more level of decomposition consisting of tasks/processes that should be regularly performed for continuously improving factory operational performances. Information flowed between activities and software functions supporting each activity are captured in the activity model.



Fig. 2. Factory Design and Improvement Activity Model (from [23])

The figure shows software functions grouped into five categories entering the bottom of each activity box SCM for supply chain management, ERP for enterprise resource planning, DM for digital manufacturing, PLM for product life cycle management, and MES for manufacturing execution system. The more tasks performed and managed, software functions deployed, and information digitally flowed, the more ready a factory is for the deployment of smart manufacturing concepts. The contribution of these aspects and dimensions to the smart manufacturing readiness is illustrated as shown in **Fig 3**. Differing ways of computing readiness index are used for C1 to C4. C1 uses the CMMI index qualification. C2 and C3 uses counting measures, while C4 uses incidence matrix-based similarity measure



along with an incidence scoring scheme based on the technology used to enable the information flow. They are viewed independently or averaged into a single SMSRL index.

Fig. 3. SMSRL measurements (from [[21])

Like other models SMSL is largely descriptive. After an assessment, a company can use the model to prescribe goals to improve the readiness, but the model has not yet included guidelines for achieving those goals. The FDI activity model underlying the assessment also focuses on within-factory improvement tasks, not day-to-day factory operation tasks, and has weaknesses on supply chain and logistics operations. The model will also need to be revisited to clearly bound the assessment as smart manufacturing readiness as opposed to smart manufacturing maturity because some of the software functions included may already step into the smart manufacturing arena. For these reasons, SMSRL will benefit from alignment and harmonization with other assessment methods described in this paper.

2.5 MOM Maturity

MESA (Manufacturing Enterprise Systems Association) created the MOM/CMM (Manufacturing Operations Management/ Capability Maturity Model) to evaluate the maturity of manufacturing enterprises' manufacturing facilities [24]. The objective is to determine the policy, procedure, and execution of a manufacturing operation management to be organized, robust, and repeatable. In other words, MOM/CMM does not provide a measure of sophistication of the physical production, but a measure of the capability to streamline operations, particularly in the abnormal events. The MOM/CMM focuses on four main process areas: 1) Production Operations Management, 2) Inventory Management, 3) Quality Test Operations Management, and 4) Maintenance Operations Management. Each process area consists of multiple activities: 1) Scheduling, 2) Dispatching, 3) Execution Management, 4) Resource Management, 5) Definition Management, 6) Data Collection, 7) Tracking, 8) Performance Analysis [24]. Each activity can have a maturity level from level 0 to level 5, which are briefly characterized in Table 2.

The higher the level of maturity, the more likely an efficient organization and fewer problems at the manufacturing operations management level. The maturity levels can be also applied across different aspects, such as 1) roles and responsibilities, 2) succession plans and backups, 3) policies and procedures, 4) technology and tools, 5) training, 6) information integration, and 7) KPIs. The model, in its raw form, can be time and resource consuming to complete with 832 questions and lacks improvement strategies based on the results. However, the model can provide a benchmark for comparison to others in their industry and can aid in understanding where to make improvements. Future work will simplify the questionnaire and map improvement strategies to the results.

Table 2. MOM Maturity level definitions

Level 0	There has been no evaluation performed.
Level 1	Procedures for activities and their executions are at initial stage and not documented
	or formally managed.
Level 2	Procedures of some activities are documented and executed with possibly repeatable
	results in the normal situation.
Level 3	Procedures for activities are defined with documented standards for all activities
	whose executions are possibly supported by software tools and better handling of
	abnormal situations.
Level 4	Procedures for activities are defined and documented across all organizational
	groups; and their executions are repeatable and monitored with software tools
	supports.
Level 5	Procedures for activities are focused on continuous improvement and optimization.

3 Models comparison: building a framework for SMSC and road-mapping its development

From the review of the different methods described in the previous chapters (DREAMY, SMSRL, MOM maturity models), it is possible to state that they provide complementarity in the overall scope of Smart Manufacturing. In fact, MOM maturity model focuses on day-to-day factory operation tasks. Therefore, it can be complemented by SMSRL, which focuses more on assessing the maturity of factory improvement tasks. Both MOM and SMSRL do not include product life cycle and business processes in their scope of analysis, so they can be complemented by DREAMY, which offer a business processes-oriented view also on product life cycles phases. According to their *different but complementary* objectives, DREAMY, SMSRL, and MOM models might be used by manufacturing companies with *different but complementary* purposes, i.e. descriptive and prescriptive and descriptive and comparative respectively.

In the table below summarizes the three models showing their objectives, clarifying their focus, and describing their structures.

Element	DREAMY	SMSRL	MOM
Objective(s)	 To assess a manufacturing company readiness level for starting the digital transformation process To identify strengths and weaknesses and related opportunities manufacturers can gather from the digital transformation, with the final aim to help them in defining a roadmap for prioritizing investments 	To assess a manufacturing company's readiness to employ data-intensive technologies for its performance management.	To determine level of an organization's capability to have mature, robust, and repeatable manufacturing operations [24].
Focus	Manufacturing company / Product and Factory Life Cycles	Maturity of performance improvement tasks/processes, availability of software supports, maturity of information sharing capability, and availability of responsible personnel	Manufacturing Operations Management (MOM) processes
Analysis Dimensions	Process / Execution, Monitoring and control, Organization, Technology	Organization, IT, Performance Management (process execution), and Information Connectivity	Process / Execution
Process Areas	Product and asset design and engineering, Production management, Quality management, Maintenance management, Logistics management, Digital Backbone	(Change) Requirement developments, Basic (rough) design of a new or a change requirement, Detail design, and Test	Production Operations Management, Inventory Management, Quality Test Operations Management, Maintenance Operations Management
Maturity levels	5 (1-5)	6 (0-5)	6 (0-5)
Inspiring framework	СММІ	Factory Design and Improvement Activity model	ISA-95 Enterprise Control Activities
Assessment methods	Interview / case study	Self-assessment	Self-assessment
Model purpose	Descriptive and prescriptive	Descriptive and comparative	Descriptive and comparative
Questions / Answers' type	Questions with normative answers	Yes/No Question, Scoring Question	Yes/No Questions
number of questions	questions 200 scoring	~123 Yes/No questions	832 Yes/No Questions

Table 3. Comparison of DREAMY, SMSRL and MOM models

4 Conclusions

Smart Manufacturing (SM) "recipe" requires enterprises to merge together different "ingredients" to obtain the best results in terms of performance improvements. In particular, due to the high-complexity of the digital transformation process, companies aiming at building SM systems have to be ready for starting the journey and need to be endowed with several capabilities. What are these capabilities and how can they be measured? This paper carries out some reflections of this smart manufacturing system characterization (SMSC), showing three different tools for assessing manufacturing companies their ability to start the digitalization process. Thanks to their comparison, it is now possible to reflect on the different perspectives required by SMSC, and future developments expected for such types of "tools".

First of all, the evidences from the literature, and from the considerations emerged during the workshop organized by NIST and OAGi [8], show that several capabilities are required in terms of organization, process execution and technology. To this aim, models and tools for assessing enterprises' readiness to start the digitalization process to embrace SM should consider all these different aspects and analysis dimensions in order to be effective. For this purpose, we may expect that many, current, and emerging models and tools could be used by the manufacturers, in different aspects and dimensions. Furthermore, it is worth remarking that these models and tools should not support solely at the assessment phase. Instead, they should be enhanced to support the prescription phase of improvements. In addition, with sufficient improvement data, benchmarking can be developed providing the evidence of returnon-investment for the smart manufacturing adoption. This could accelerate the industry adoption. Finally, further studies should deal more with principles, providing an abstract view on the founding concepts, and adequately addressing the differences between other "readiness" and "maturity" models, in order to suggest the most appropriate tool to use in each of the digital roadmap building phase, i.e. for maturity assessment and/or for digital readiness identification.

5 References

- [1] GMIS. Industry 4.0: Building the Digital Industrial Enterprise. Р 2016. https://www.pwc.com/m1/en/publications/documents/middle-east-industry-4-0-survey.pdf.
- E.R. for C. in M. SCorPiuS, Validated sCorPiuS Vision, (2016) 1-17.
- ī31 J. Jager, O. Schollhammer, M. Lickefett, T. Bauernhansl, Advanced complexity management strategic recommendations of handling the "Industrie 4.0" complexity for small and medium enterprises, in: 49th CIRP Conf. Manuf. Syst. (CIRP-CMS 2016), 2016.
- [4] W. Elmaraghy, H. Elmaraghy, T. Tomiyama, L. Monostori, Complexity in engineering design and manufacturing, CIRP Ann. - Manuf. Technol. 61 (2012)793-814. doi:10.1016/j.cirp.2012.05.001.
- J. Jäger, A. Kluth, A. Schatz, T. Bauernhansl, Complexity patterns in the advanced complexity [5] CIRP, management of value networks. in: Procedia 2014: pp. doi:10.1016/j.procir.2014.01.070.
- [6] A. De Carolis, M. Macchi, E. Negri, S. Terzi, Guiding manufacturing companies towards digitalization A methodology for supporting manufacturing companies in defining their digitalization roadmap, 2017.
- [7] [8] D. Brandl. SMART MANUFACTURING SYSTEM CHARACTERIZATION SESSION. (2016). NIST/OAGi, Drilling down on Smart Manufacturing - Enabling Composable Apps, (2016). https://www.nist.gov/news-events/events/2016/04/nistoagi-workshop-drilling-down-smartmanufacturing-enabling-composable (accessed April 18, 2017).

Brundage, Michael; Carolis, Anna; Kulvatunyou, Boonserm; Macchi, Marco; Terzi, Sergio. "Maturity models and tools for enabling Smart Manufacturing Systems: comparison and reflections for future developments." Paper presented at IFIP 14th International Conference on Product Lifecycle Management, Seville, Spain. July 9, 2017 - July 12, 2017.

- [9] A.M. Maier, J. Moultrie, P.J. Clarkson, Assessing organizational capabilities: Reviewing and guiding the development of maturity grids, IEEE Trans. Eng. Manag. 59 (2012) 138-159. doi:10.1109/TEM.2010.2077289.
- [10] J.A. Simpson, E.S.C. Weiner, The Oxford English Dictionary, 1989.
- [11] T. Mettler, A Design Science Research Perspective on Maturity Models in Information Systems, Design. 41 (2009) 1-13. doi:10.2174/97816080506351100101.
- H. Karkkainen, A. Silventoinen, Different Approaches of the PLM Maturity Concept and Their [12] Use Domains - Analysis of the State of the Art, in: 12th IFIP WG 5.1 Int. Conf. PLM 2015, Springer International Publishing Switzerland 2016, Doha, Quatar, 2015: pp. 89-102. doi:10.1007/978-3-319-33111-9 9.
- [13] T. Mettler, P. Rohner, Situational maturity models as instrumental artifacts for organizational design, Proc. 4th Int. Conf. Des. Sci. Res. Inf. Syst. Technol. - DESRIST '09. Artic. No. 22. (2009) 1-9. doi:10.1145/1555619.1555649.
- [14] P. Gottschalk, Maturity levels for interoperability in digital government, Gov. Inf. Q. 26 (2009) 75-81. doi:10.1016/j.giq.2008.03.003.
- [15] M. Van Steenbergen, R. Bos, S. Brinkkemper, I. Van De Weerd, W. Bekkers, The design of focus area maturity models, in: CEUR Workshop Proc., 2010: pp. 17-19. doi:10.1007/978-3-642-13335-0 22.
- [16] P. Fraser, J. Moultrie, M. Gregory, The use of maturity models/grids as a tool in assessing product development capability, in: IEEE Int. Eng. Manag. Conf., 2002.
- C.P. Team, Capability Maturity Model {\textregistered} Integration (CMMI SM), Version 1.1, C. [17] Syst. Eng. Softw. Eng. Integr. Prod. Process Dev. Supplier Sourc. (CMMI-SE/SW/IPPD/SS, V1. 1). (2002).
- [18] M. Macchi, L. Fumagalli, S. Pizzolante, A. Crespo, J.F. Marquez, G. Fernandez, Towards eMaintenance: maturity assessment of maintenance services for new ICT introduction, in: APMS 2010 Int. Conf. Adv. Prod. Manag. Sys-Tems, Cernobbio, Italy, 2010.
- [19] M. Macchi, L. Fumagalli, A maintenance maturity assessment method for the manufacturing industry, J. Qual. Maint. Eng. 19 (2013) 295-315. doi:10.1108/JQME-05-2013-0027.
- [20] A. De Carolis, M. Macchi, E. Negri, S. Terzi, A maturity model for assessing the digital readiness of manufacturing companies, 2017.
- [21] K. Jung, B. Kulvatunyou, S. Choi, M.P. Brundage, An Overview of a Smart Manufacturing System Readiness Assessment, in: Springer-Verlag Berlin Heidelb., 2016.
- K. Jung, S. Choi, B. Kulvatunyou, H. Cho, K.C. Morris, A reference activity model for smart [22] factory design and improvement, Prod. Plan. Control. (2016). doi:10.1080/09537287.2016.1237686.
- B.. Kulvatunyou, Factory Design and Improvement (FDI) Activity Model, (n.d.). [23] https://www.nist.gov/services-resources/software/factory-design-and-improvement-fdi-activitymodel.
- [24] D. Brandl, MESA MOM Capability Maturity Model Version 1.0, (2016).
- [25] Lichtblau, K. et al. Industrie 4.0 Readiness, VDMA, Available online: https://www.industrie40-readiness.de/?lang=en

Brundage, Michael; Carolis, Anna; Kulvatunyou, Boonserm; Macchi, Marco; Terzi, Sergio. "Maturity models and tools for enabling Smart Manufacturing Systems: comparison and reflections for future developments." Paper presented at IFIP 14th International Conference on Product Lifecycle Management, Seville, Spain. July 9, 2017 - July 12, 2017.

9th International Symposium on Steel Structures, November 1-4, 2017, Jeju, Korea

Localized fire tests on steel beams with different end restraints

Selvarajah Ramesh¹, Mina Seif², and Lisa Choe³

1,2,3 National Institute of Standards and Technology, Gaithersburg, MD 20899-1070, USA

¹ selvarajah.ramesh@nist.gov, ² mina.seif@nist.gov, ³ lisa.choe@nist.gov (corresponding author)

Abstract

Two 6.2-m span I-shaped structural steel beams were tested under combined structural and open flame localized fire loads. Two different support conditions for the beam ends were considered: (i) simple support, and (ii) double-angles bolted to laterally braced support columns. A four-point flexural loading scheme was used such that two concentrated forces were applied 2.44 m apart around midspan. The beam specimens were laterally braced at the location of concentrated forces. The midspan of each specimen, i.e. expected plastic hinge zone, was exposed to an open-flame fire using natural gas burners. The measurements included heat release rate and applied forces as well as thermal and structural responses to fire. Since the thermal gradients developed in the fire-exposed cross sections induced thermal bowing, both beam specimens exhibited similar deformational behavior and failure mode regardless of their end conditions. However, the presence of end restraint, provided through the double-angles on the second specimen, decreased its fire resistance (i.e. the failure time and failure temperature).

Keywords: Fire, End restraints, Steel beams, Experimental

1. Introduction

The current fire safety design focuses on improving the fire resistance of unrestrained structural steel members to uniform heating caused by a post-flashover fire. In openplan compartments, however, a fire rarely develops into flashover and could be localized at a certain location. Some researchers (Zhang et al, 2014; Choe et al, 2016a; Agarwal et al, 2014) studied the effects of non-uniform heating on the fire performance of structural steel columns and discussed the effect of thermal gradients on the loadcarrying capacity of columns. There is still lack of technical information needed to evaluate the behavior of restrained structural steel beams subjected to non-uniform heat possibly induced by an open flame, localized fire.

As part of commissioning the new structural fire testing capabilities of the National Fire Research Laboratory (Bundy et al, 2016), a series of localized fire tests on structural steel beams were conducted. This paper focuses on the influence of end restraints on the 6.2-m long, W16x26 structural steel beams tested under localized fire exposure. The fire performance of the beams with two different end supports were evaluated, including (a) simple support (Specimen 1) and (b) double-angle connection (Specimen 2).

2. Test Setup

Details of the test configuration and instrumentation layout are presented in Choe et al. (2016b) and are briefly summarized herein. Fig 1 shows a schematic of the test setup erected on the strong floor. The nominal length of the W16x26 beam specimens (between the centerlines of support columns) was 6.71 m. Specimens were made of ASTM A992 (ASTM, 2015b) steel with a minimum yield and tensile strengths of 340 MPa and 450 MPa, respectively. The width and thickness of flanges was 140 mm and 8.76 mm, respectively; the thickness of web was

6.35 mm; and the depth of the section was 400 mm.

As shown in Fig. 1, the beam specimen was loaded using two box-shaped steel beams placed at 2.44 m apart around mispan. To apply a bending moment to the specimen, the steel rods attached at the ends of the loading beams were pulled by hydraulic actuators placed in the basement underneath the strong floor. Specimens 1 and 2 were laterally braced at the location of the loading beams so that the unbraced length was 2.44 m. For producing a localized fire exposure to the beam specimen, a 1-m² natural gas burner system was placed 1.1 m below the bottom flange of the specimen at its midspan.



Fig. 1. A schematic of the test setup

Details of the beam end supports are shown in Fig. 2. The ends of Specimen 1 were only restrained against the rotation about the longitudinal axis of the specimen (i.e. twisting). The ends of Specimen 2 were bolted to L5x5x5/16 double angles using three 19.1 mm diameter ASTM A325 (ASTM, 2015a) bolts.

Choe, Lisa; Ramesh, Selvarajah; Seif, Mina. "Localized fire tests on steel beams with different end restraints." Paper presented at The 9th International Symposium on Steel Structures, Jeju, Republic of Korea. November 1, 2017 - November 4, 2017.



Fig. 2. Details of the support of (a) Specimen 1 (b) Specimen 2

Fig. 3 shows the instrumentation layout used in the tests. Type-K thermocouples were installed at sections 1 through 5 in the fire-affected zone. The temperaturecompensated string potentiometers were installed to measure the vertical and lateral displacements at midspan. The conventional string potentiometers were used to measure the vertical displacements of the bottom flanges at the two loading points.



Fig. 3. Thermocouple and displacement sensor layout

3. Results and Discussion

Fig. 4 shows the average point load (P_v) applied on the beam specimen at the location of box-shaped loading beams and the heat release rates (HRR) of the natural gas burner over the test duration. Prior to fire loading, each point load on the beam specimen was increased to 89 kN, approximately 67% of the measured ambient capacity, and this load was maintained throughout the test. The specimen was exposed to an open flame fire controlled by regulating the flow of natural gas. The t-square fire was applied using a pre-programed quadratic function of HRR = $4.5 \cdot t^2 + 250$, where t is fire exposure time in minutes and HRR is in kW. Fig. 5 shows a photograph of Specimen 1 during the test. When the specimens could no longer support the imposed loads, the specimens were immediately unloaded, and the burner was shut off. The figure shows that Specimen 1 failed at 37.9 minutes (i.e., 17.5 minutes from ignition of a fire); Specimen 2 failed at 30.5 minutes (i.e., 14.2 minutes from ignition of a fire).



Fig. 4. Measured HRR and average point load (P_y)



Fig. 5. Specimen 1 under structural and thermal loading

Figs. 6(a) and 6(b) show steel temperature-time curves measured in the fire-exposed cross-sections of Specimens 1 and 2, respectively. The magnitudes of temperatures in those figures are averaged values of temperatures in sections 3, 4, and 5 at each thermocouple location shown in Fig. 6(c). As shown, non-uniform temperature distribution was developed through the section depth. The temperature of the exposed bottom flange was directly affected by the HRR versus time relationship. However, the temperatures in the other locations, i.e., the upper portion of the cross section, were similar, and no severe thermal gradient was developed.

Choe, Lisa; Ramesh, Selvarajah; Seif, Mina. "Localized fire tests on steel beams with different end restraints." Paper presented at The 9th International Symposium on Steel Structures, Jeju, Republic of Korea. November 1, 2017 - November 4, 2017.



Fig. 6. Thermal response of (a) Specimen 1 and (b) Specimen 2; (c) locations of temperature measurements

Figs. 7(a) and 7(b) show the displacement versus the bottom temperature responses of Specimens 1 and 2, respectively, where the displacement data includes the vertical displacements at midspan and at the loading points as well as the lateral displacements at the top and bottom flanges. In those figures, the positive values indicate the vertical displacement in the downward direction and the lateral displacement in the south direction. It should be noted that, the discontinuity of the plotted displacement data indicates the sensor failure.

As shown in Figs 7, both Specimens 1 and 2 started sagging as soon as a fire was ignited. Regardless of the beam end conditions, the thermal gradient developed through the section depth, as shown in Fig. 6, caused bowing of the specimen. During a fire, the lower portion of the cross section subjected to an intense heat from fire expanded greater than its upper portion which remained at lower temperature. This nonuniform thermal expansion led the beam specimen to deflect toward the hotter side.

As the critical temperatures were reached, the beam specimens behaved in a complex way such that combined flexural bending (about the strong axis) and lateral torsional buckling occurred simultaneously. The midspan lateral displacements of Specimens 1 and 2 gradually increased when the bottom flange temperature reached about 550 °C and 500 °C, respectively. At those temperatures, the vertical displacements were continuously increasing.

To evaluate the overall fire resistance of the steel beams on a consistent basis, the temperature at a certain vertical displacement limit can be compared. In this study, the vertical displacement of 76 mm was considered as the displacement limit at failure since it was the maximum displacement recorded for Specimen 1. As shown in Fig. 7, the bottom flange temperature of Specimen 1 corresponding to this displacement limit was (663 ± 85) °C and that of Specimen 2 was (552 ± 21) °C.

Given that the same structural and fire loading were applied, Specimen 1 failed at higher bottom flange temperature than Specimen 2. The lower failure temperature observed in Specimen 2 is attributed to the effects of axial restraints (against thermal expansion) provided by double-angle connections attached to laterally rigid columns.



Fig. 7. Deflection responses of (a) Specimen 1 and (b) Specimen 2

It should be noted that the estimated standard uncertainty in temperature data reported in Fig. 6 was 42.5 °C. The estimated expanded uncertainty in the displacement measurement was 0.3 mm with a coverage factor of 2.

Figs. 8(a) and 8(b) show the deflected shapes of

Choe, Lisa; Ramesh, Selvarajah; Seif, Mina. "Localized fire tests on steel beams with different end restraints." Paper presented at The 9th International Symposium on Steel Structures, Jeju, Republic of Korea. November 1, 2017 - November 4, 2017.

Specimens 1 and 2, respectively, after the cool-down phase. The steel beam specimens exhibited permanent deformations in the combination of the strong-axis bending and the lateral-torsional buckling. In both specimens, local buckling modes were observed at the compression flange (i.e., the top flange at midspan).





(b)

Fig. 8. Deflected shapes of (a) Specimen 1 and (b) Specimen 2

4. Summary and Conclusions

Two 6.17-m long W16×26 steel beams were tested under combined structural (flexural) and open-flame localized fire loads. Specimen 1 was simply supported; Specimen 2 was supported by rigid columns via double angle connections. The specimens were loaded using a fourpoint loading scheme to apply a uniform bending moment between two point loads. The midspan of each specimen was directly exposed to an open-flame natural gas fire. Regardless of the beam end conditions, a thermal bowing was observed due to the thermal gradient developed in the fire-exposed cross sections. Both specimens showed combined flexural and lateral torsional behavior at failure. Under the same t-square fire, however, Specimens 1 and 2 failed at a bottom flange temperature of 663 °C and 552 °C, respectively. This discrepancy in the failure temperature was caused by axial restraints (against thermal expansion) provided by double-angle connections.

6. References

 Agarwal, A., Choe, L., Varma, A. (2014). "Fire design of steel columns: Effects of thermal gradients." Journal of Constructional Steel Research Vol 93, 107-118.
 AISC (2010). Steel Construction Manual, 14th edition, American Institute of Steel Construction (AISC), Table 1-1, Chicago, IL

- ASTM International. (2015a). "Standard Specification for High Strength Structural Bolts, Steel and Alloy Steel, Heat Treated, 120 ksi (830 MPa) and 150 ksi (1040 MPa) Minimum Tensile Strength, Inch and Metric Dimensions," Standard F3125/F3125M-15a, ASTM International, W. Conshohocken, PA.
- ASTM International. (2015b). "Standard Specification for Structural Shapes," Standard A992, ASTM International, W. Conshohocken, PA.
- Bundy, M., Hamins, A., Gross, J., Grosshandler W., Choe, L. (2016). "Structural Fire Experimental Capabilities at the NIST National Fire Research Laboratory," *Fire Technology.*, pp. 1-8, doi:10.1007/s10694-015-0544-4.
- Choe, L., Agarwal, A., Varma, A. (2016a). "Steel Columns Subjected to Thermal Gradients from Fire Loading: Experimental Evaluation." Journal of Structural Engineering 142.7: 04016037.
- Choe, L., Ramesh, S., Zhang, C., and Gross, J. (2016b). "The performance of structural steel beams subject to a localized fire." 9th International Conference on Structures in Fire. Princeton, New Jersey, USA.
- Zhang et al. (2014). "Behavior of unrestrained and restrained bare steel columns subjected to localized fire." Journal of Structural Engineering 141.10: 04014239.

NEW CONNECTIONS FOR ENHANCING ROBUSTNESS OF PRECAST CONCRETE FRAME STRUCTURES

Travis Thonstad, Yihai Bao, Jonathan M. Weigand, Joseph A. Main, and H.S. Lew National Institute of Standards and Technology, Gaithersburg, MD

ABSTRACT

A recent experimental and computational study of two full-scale precast concrete moment-frame assemblies under a column removal scenario revealed vulnerabilities in the moment connections arising from (a) local bending of reinforcing bars due to eccentricity in the tensile load path between components and (b) reduction in the ductility of reinforcing bars due to welding. To overcome these issues with the tested prototype connections, three new alternative connections, which incorporate welded link plates, threaded bars, and grouted bars, respectively, are proposed to enhance the robustness of precast concrete frame structures.

Computational analyses of the experimental column removal scenario were conducted for one of the tested assemblies and for geometrically identical assemblies that incorporated the alternative connection details. The robustness index, a normalized measure of the dynamic load carrying capacity of a system under sudden column removal, was used to quantify and compare the performance of the alternative connections based on the analysis results. The assemblies with the welded link plate, threaded bar, and grouted bar connections had robustness indices 50 %, 130 %, and 220 % greater, respectively, than that of the assembly with prototype connections.

This paper describes the three proposed connection concepts, presents results from numerical modeling and outlines plans for component and full-scale connection testing.

Keywords: Disproportionate Collapse, Precast Concrete Buildings, Structural Robustness, Finite Element Analysis, Moment Resisting Connections.

INTRODUCTION

The National Institute of Standards and Technology (NIST) has been conducting a multi-year research program focused on mitigating disproportionate collapse in structural systems. As part of this research, ten-story prototype buildings were designed using various structural systems; steel braced frames; reinforced concrete shear walls; and steel, reinforced concrete, and precast concrete moment frames. Full-scale beam-column assemblies representing portions of steel moment frames,¹ reinforced concrete moment frames,² and precast concrete moment frames^{3,4} were tested under an interior column removal scenario.

Two exterior, two-bay precast concrete moment frames, designed for different Seismic Design Categories (SDCs),⁵ were tested as part of this research. The first frame was designed as an ordinary moment frame (OMF) for SDC B. The second frame was designed as a special moment frame (SMF) for SDC D. Beams of the OMF assembly were deeper and had fewer longitudinal reinforcing bars than those in the SMF assembly. In both assemblies, the same type of connection was used between the precast beams and columns; the prototype connection detail for the SMF assembly is shown in Fig. 1. The beams and the columns were connected via steel link plates, which were welded to steel angles embedded in the beams and steel plates embedded in the columns. Forces in the link plates were transmitted to the beams via anchorage



Fig. 1 SMF prototype connection detail

bars, which were welded to the interior faces of steel angles. The SMF connections incorporated three no. 11 (36M) anchorage bars, while the OMF connections incorporated two no. 10 (32M) anchorage bars.

The test setup for the SMF precast concrete frame assembly is shown in Fig. 2; the test setup for the OMF assembly was identical. Because of space limitations, the spans in the tested assemblies were reduced to 300 in. (7620 mm) from the prototype building span of 360 in. (9144 mm). Longitudinal bracing was provided between the tops of the exterior columns to restrain in-plate movement. Lateral bracing was provided at midspan of the beams and at the center column.

Similar to the cast-in-place concrete moment frame assemblies tested previously,² failure of the precast concrete frame assemblies involved fracture of reinforcing bars at the bottom of the beams near the center column. However, fracture of the reinforcing bars occurred at much smaller vertical displacements of the center column for the precast concrete frame assemblies: 2.49 in. (63.2 mm) for the SMF assembly and 5.66 in. (144 mm) for the OMF assembly.^{3,4} For comparison, the cast-in-place concrete SMF assembly, which had shorter span lengths and beams with higher span-to-depth ratios, was able to withstand vertical displacements in excess of 44.5 in. (1130 mm) prior to failure.²



Fig. 2 Experimental configuration for the SMF specimen (adapted from Main et al.³).

The large displacement capacity of the cast-in-place concrete SMF assembly allowed it to develop significant catenary action prior to failure. This led to an increase in load carrying capacity after the vertical resistance due to flexural and arching action was depleted at smaller vertical displacements. The potential for development of catenary action in the precast concrete frames is limited by the low span-to-depth ratios of the beams (3.1 and 4.8 for the precast OMF and SMF assemblies, respectively, compared with 9.2 for the cast-in-place SMF assembly). Experimental data from a large set of reinforced concrete assemblies showed that significant catenary action was only developed in those assemblies with beams that had span-to-depth ratios exceeding 6.0.⁶ Although the potential for development of catenary action in the precast concrete frames is limited, improvements in the connection design leading to larger vertical

displacements at the center column could be expected to increase the load carrying capacity by further mobilizing the development of flexural and arching action .

In the precast concrete frame assemblies, fracture of the anchorage bars occurred at the end of the flare-bevel-groove welds between the anchorage bars and the embedded angles. Fracture was attributed to flexure of the bars, due to the eccentricity of the tensile force transfer path, in combination with reduced ductility of the bars in the heat-affected zone caused by welding.^{3,4} To determine the reduction in ductility of the bars that contributed to this failure, uniaxial tension testing under concentric loading was performed on a welded bar extracted from a prototype connection,³ as shown in Fig. 3. The effect of welding was shown to result in a decrease in the strain at fracture. In combination with flexure due to connection eccentricities, this reduction in ductility caused the anchorage bars to fracture significantly below their nominal tensile capacity.^{3,4}



Fig. 3 Reduced ductility of welded no. 11 anchorage bar (adapted from Main et al.³)

The robustness of the precast frames under sudden column loss was estimated from the static push-down tests and the applicable gravity loading on the prototype structure using an energy-based approach;³ the procedure is shown schematically in Fig. 4. This estimation assumes that the gravity loads are insufficient to cause collapse and that the frame deforms in a single, predominant mode. In such a system, the kinetic energy is zero at the point of maximum displacement, and the work done by external loads can be equated to the internal energy of the system at a given deformation.

Using this energy-based approach, a robustness index R can be defined as a measure of the robustness of a structural system under sudden column loss:⁷

$$R = \frac{1}{P_G \Delta_u} \int_0^{\Delta_u} P(\Delta) \, d\Delta \,, \tag{1}$$

where $P(\Delta)$ is the force-displacement response of the system obtained from static pushdown loading at the location of the removed column, Δ is the vertical displacement at the removed column, Δ_u is the displacement at the ultimate static load, and P_G is the force in the column to be removed under service-level gravity loading G, which is commonly taken as G = 1.2D + 0.5L, where D is dead load and L is live load.⁷



Fig. 4 Energy-based estimation of capacity under sudden column loss

Robustness indices of 1.11 and 1.02 were obtained for the OMF and SMF precast assemblies, respectively,^{3,8} which means that the estimated ultimate capacities exceeded the service-level gravity loading by 11 % and 2 %. These robustness indices correspond to the prototype span length of 30 ft (9144 mm) and were obtained using computational models, since the experimental assemblies had reduced span lengths of 25 ft (7620 mm). Given the uncertainty in the computational models and the sensitivity to certain material and geometric parameters, as discussed subsequently, the ultimate capacity of these frames may be insufficient to resist collapse under a sudden column removal for the applicable service gravity loads.

To increase the robustness of precast concrete frames under column removal scenarios, three potential strategies were identified for enhancing the performance of the connections:

- 1. Enhance the force and deformation capacity of the welded link plate connection by
 - a. Reducing the eccentricity in the tensile force transfer path, and
 - b. Promoting yielding and plastic deformation of the link plates.
- 2. Replace the welded link plate connection with a threaded rod connection.
- 3. Replace the welded link plate connection with a grouted bar connection.

These modifications were meant to increase both the ultimate load and the corresponding deformation that can be sustained by the connections, thus enabling flexural and arching action to be more fully mobilized in resisting column loss.

ENHANCED CONNECTIONS

The first connection type (Type A) uses a modified link plate detail to attach the precast beams to the precast columns; the second connection type (Type B) employs high-strength threaded steel rods; and the third connection type (Type C) uses large-diameter, grouted reinforcing bars.

WELDED LINK PLATE (TYPE A)

The revised welded link plate connection is shown in Fig. 5. The Type A connection incorporates the minimal modifications to the prototype configuration necessary to reduce the eccentricity of the tensile force transfer path and improve connection ductility. The construction sequence and beam reinforcement are otherwise identical. In a Type A connection, anchorage bars are welded to the *exterior* faces of an embedded WT section in the beam, rather than to the *interior* faces of an angle, as was done in the prototype configuration. This aligns the forces in the link plate and the anchorage bars, thus reducing flexural stresses due to eccentricities. The flexural stiffness of the WT section also reduces flexural stresses in the anchorage bars by limiting the curvature that can develop between the link plate and the anchorage bars. The link plates are reduced in width compared with the prototype connection, so that the expected tensile strength of the link plate is roughly equal to that of the three anchorage bars. A free length is added between the welds, to promote plastic deformation of the link plates by spreading the yielding over an increased length. By reducing the distance between the welds on the embedded column plate, adequate weld length parallel to the beam centerline can be maintained.



ig. 5 Type A connection deta

5

THREADED ROD (TYPE B)

The threaded rod connection is shown in Fig. 6. The Type B connection eliminates welding of critical connection components in the field. The beams and columns are connected via threaded rods, which were designed to have a tensile capacity approximately equal to that of the three anchorage bars in the prototype design. The threaded rods run through ducts within the beam, are anchored by a nut that bears against an embedded rectangular steel hollow structural section (HSS) in the beam, and terminate at a standard steel coupler embedded in the column. The threaded rods are tightened to wrench-tight.



Fig. 6 Type B connection detail

Axial forces in the threaded rods are transferred to the beams through bearing of the rectangular HSS section on the beam concrete and through flexural bars connected to the rectangular HSS section opposite the threaded rod. These flexural bars could be connected to the HSS section using threaded ends with nuts, or by welding. The HSS section is shown flush with the front face of the beam in Fig. 6, to allow continuous longitudinal reinforcement to pass on the opposite side. This is a consequence of the prototype geometry, and for other beam geometries, the HSS section could be embedded at mid-width in the beam. This arrangement would allow reinforcement to pass by on either side and would result in a symmetric beam section.

GROUTED BAR (TYPE C)

The grouted bar connection is shown schematically in Fig. 7. The Type C connection avoids the use of welded link plates, in favor of grouted reinforcing bars in ducts. The beams and column are connected using large diameter reinforcing bars, designed to have a tensile capacity approximately equal to that of the three anchorage bars in the prototype design. These bars run continuously through the ducts in both the beams and columns. These ducts, as well as the beam-column interface, are then grouted to complete the connection. To slide the bars through the column and into the adjacent beam, troughs are cast into the beam extending a short distance into the span. The corrugated steel ducts are located at the center of the beam cross-section, to further reduce eccentricities of tensile force transfer path.



Fig. 7 Type C connection detail

Tension forces from the anchorage bars are transferred to the beam through bond. Previous bond tests of large-diameter bars grouted in ducts have shown that a bonded length of roughly 10 bar diameters is sufficient to reach the ultimate capacity of the bar under monotonic loads⁴. To increase their elongation at the connection, the bars are locally debonded through the column and in the beam, near the grouted beam-to-column interface.

NUMERICAL MODELING

The performance of each modified connection type was evaluated through high-fidelity finite element modeling of the connections. Each connection type was evaluated under two loading scenarios: (1) direct tension applied to a single connection and (2) increasing vertical displacements applied at the center column of a two-span frame assembly. The modeling approach used in this proof-of-concept study followed that of Main et al.,³ this approach was experimentally validated against the precast concrete OMF and SMF assembly tests. A summary is given here for context; detailed information about the modeling approach is provided by Main et al.³

The concrete components were modeled using 8-node constant stress solid elements, the embedded steel components were modelled using 8-node fully integrated solid elements, and the reinforcement was modeled using beam elements. The mesh size of the concrete elements was roughly 2 in. (51 mm), with a maximum aspect ratio of approximately 2.5. Stirrups and discontinuous bars were modeled assuming perfect bond; grouted bars in the Type C connection were assumed to have degrading bond stiffness, calibrated using experimental data from bar pull-out tests.⁹ A segment-based contact model prevented interpenetration and enabled bearing of the components within the assembly. A coefficient of friction of 0.3 was assumed.

Based on previous comparisons of computational models with experimental measurements from the tested assemblies,³ the estimated uncertainty in the ultimate capacities obtained from the computational models is ± 8 %. This uncertainty corresponds to test specimens for which measured material properties were available. Sensitivity studies were also performed to investigate the influence of factors that would vary in construction practice, and the ultimate capacities were found to be particularly sensitive to the ductility of the anchorage bars, as influenced by welding, and the initial air gap between the beams and columns, which affects the development of arching action.³ Realistic variations in these factors were found to alter the ultimate capacities of the assemblies by as much as 35 %.³ By reducing the influence of welds on connection failure and by eliminating air gaps through grouting of connection interfaces, the enhanced connections proposed in this study are expected to have reduced sensitivity to such construction-related factors, thus reducing the uncertainty in the performance of the enhanced connections. This aspect of the connection performance will be investigated in future sensitivity studies. Because the focus of this proof-of-concept study is on the relative performance of alternative connection types, material and geometric properties are assumed in all models to be consistent with those of the tested prototype specimen. The only factor that is varied between models is the type of connection, to focus on the potential performance improvements of the alternative connection concepts.

MATERIAL PROPERTIES

The material properties for the components were identical for all the analyzed models where applicable. A continuous surface cap model¹⁰ and an isotropic elastic-plastic material¹¹ were used to capture the nonlinear behavior of the concrete and steel materials, respectively. The

compressive strength of the concrete was taken as 5785 psi (39.9 MPa), and the material model parameters were taken as the recommended values for this compressive strength from FHWA-HRT-058-063.¹⁰ The material properties used for the steel components are summarized in Table 1. In all the steel materials, the elastic modulus was taken as 29 000 ksi (200 GPa).

Table 1 Steel material model parameters					
Component	Yield Stress	Ultimate Stress	Plastic Strain at		
Component	ksi (MPa)	ksi (MPa)	Ultimate Stress		
Plate	42.0 (290)	84.2 (581)	0.4300		
Embedded Steel Sections	42.0 (290)	84.2 (581)	0.4300		
No. 4 bars	74.2 (491)	95.8 (661)	0.0826		
No. 9 bars	66.6 (459)	97.6 (673)	0.0924		
No. 11 bars	68.0 (469)	95.0 (655)	0.0833		
No. 14 bars	67.5 (465)	95.3 (657)	0.0969		
Threaded rod	105.0 (724)	125.0 (862)	0.0457		

CONNECTION UNDER DIRECT TENSION

The fracture of the reinforcing bars observed in the tests of the prototype specimen was primarily due to the tensile forces transferred via the link plates. Therefore, it is expected that increasing the tensile capacity of the connections and the corresponding tensile deformation at fracture should lead to improved structural performance, relative to the tested assembly, under a column removal scenario.

The performance of the alternative connection details was evaluated by considering a single connection with the end of the beam subjected to increasing horizontal displacement, as shown in the insert of Fig 8. The tops of the column segments were prevented from moving by two lines of rigid contact elements, and the bases of the column segments were held fixed. The tension-load-displacement curves of the prototype and alternative connections are shown in Fig. 8.

The peak loads of the assemblies with the alternative connections were comparable to each other and were roughly 40 % larger than the peak load obtained for the prototype connection. This was expected, because the tensile capacity of the threaded rod and large diameter bars were selected to be comparable to the total capacity of the three anchorage bars in the Type A and prototype connections, and the prototype connection sustained premature fracture of the anchorage bars ^{3,4}. The displacements corresponding to the peak loads were significantly larger for the assemblies with the alternative connections: roughly 200 %, 1000 %, and 700 % larger than that of the prototype connection for the Type A, Type B, and Type C connections, respectively.



Fig. 8 Numerical analyses of connections under uniaxial tension.

TWO-SPAN FRAME ASSEMBLY

To further evaluate the performance of the alternative connection details, frames with the same geometry as the tested prototype SMF specimen, but with three alternative connections concepts, were analyzed under an interior column removal scenario. The model of the prototype SMF specimen had been previously validated against experimental data,³ and was used as a starting point for the subsequent models. Fig. 9 shows the load-displacement responses of the frame assemblies with the alternative connection types.



Fig. 9 Numerical analysis of frame assemblies under column removal.

The assembly with Type A connections had a similar initial stiffness and similar failure mode to the prototype assembly; the anchorage bars at the bottom of the beams fractured near the center column stub. However, the peak load was about 80 % greater than that of the assembly with prototype connections. The peak load of the assembly with Type B connections exceeded the capacity of the assembly with Type A connections by roughly 20 % and exceeded the capacity of the assembly with prototype connections by roughly 120 %. The larger stiffness for the Type B connection resulted largely from grouting the interface between the precast beams and columns, which was not done in the frames with the prototype or Type A connections. Grouting the beam-to-column interface in the Type A connection would be expected to result in similar increases in stiffness, associated with an earlier onset of arching action. The failure mode of the assembly with Type B connections differed from the failure mode observed in the other two assemblies. Following the peak vertical load, the resistance of the connection gradually decreased due to distributed concrete damage within the span of the beam between the connections.

The peak load of the assembly with the Type C connections exceeded the vertical load carrying capacity of the assemblies with the prototype, Type A, and Type B connections by roughly 190 %, 60 %, and 30 %, respectively. This increase in load carrying capacity was largely attributed to additional arching action. However, it should be noted that developing significant arching action in the modeled assembly relies on the behavior of the column segments under large shear forces, which was not specifically validated in the previous precast frame studies,^{3,8} since failure of the connections occurred prior to the development of significant arching action. The eventual failure of the assembly with Type C connections was governed by fracture of the grouted bars running through the center column segment.

Table 2 quantifies the improvement to the robustness of the precast frames under a sudden column loss using the robustness index, given in Equation 1. The robustness indices of the alternative connection concepts were normalized to the prototype connection for direct comparison.

Connection type	Normalized robustness index			
Prototype	1.0			
Type A	1.5			
Type B	2.3			
Type C	3.2			

Table 2 Normalized robustness indices of modeled precast frames.

The assembly with the Type A connection, representing the minimal modifications necessary to address the shortcomings of the prototype connection identified by Main et al.,^{3,8} had a 50 % increase in robustness under a sudden column loss. The models of the assemblies with Type B and Type C connections, which both eliminated the link plates and included grouted beam-to-column interfaces, had 130 % and 220 % larger robustness indices than the prototype frame, respectively. These analyses show that the proposed alternative connection designs have the potential to enhance the robustness of a precast concrete frame under an interior column removal scenario. Experimental validation is needed to verify these increases in performance.

PLANNED EXPERIMENTAL TESTING

Experimental testing is currently underway to verify the performance of the new alternative connection details and to further validate key aspects of the high-fidelity finite element models, including bond and bearing of the steel components within the beam. These tests will also be used to further refine the connection concepts, as appropriate; and develop reduced-order modeling approaches that can be implemented in system-level numerical simulations.

Uniaxial tension tests on components of the precast frame assemblies are underway. These tests will focus on the tensile force transfer path from the beam to the column, which plays a critical role in the connection performance. First, full-scale components of both the prototype and the Type A connection detail will be tested, to validate the behavior of the improved welded link plate detail. Second, scaled partial-depth assemblies will be tested to estimate the relative performance of the new connection details in terms of the ultimate strengths and deformation capacities of the connections in tension.

Testing of full-scale, cantilever, beam-to-column assemblies that utilize the new connection concepts is planned. The setup for these tests is shown schematically in Fig. 10. These assemblies will incorporate the enhanced connections, and will include any improvements that are identified through the component testing.



Fig. 10 Full-scale connection test setup.

Axial and rotation demands will be applied quasi-statically to the cross section at the end of the cantilever beams in the connection assembly; a similar approach has been used previously to test steel gravity connections.¹² Two actuators will be aligned with the beam axis and one (or two) will be oriented transversely. The applied displacements and rotations will be determined by assuming an antisymmetric response for the hypothetical connection at the other end of the beam, which requires zero axial displacement and zero bending moment at mid-span. The actuators oriented axially along the beam centerline will apply the forces due to compressive arching action (at small rotations) and catenary action (at large rotations). The benefit of this arrangement is that an arbitrary degree of longitudinal restraint can be applied to the connection, to assess the influence of both flexural and arching action on frame performance under a column loss scenario.

SUMMARY AND DISCUSSION

Connections between precast concrete components must facilitate rapid on-site assembly and may be required to provide sufficient structural robustness to prevent disproportionate collapse under a column removal scenario. A recent experimental study on two precast concrete frame assemblies under a column removal scenario revealed vulnerabilities in the precast concrete moment frame connections arising from the eccentricity of the tensile force transfer path between components and a reduction in the ductility of reinforcing bars in the heat-affected zone due to the welding.^{3,4,8}

Three new alternative connections were proposed to enhance the robustness of precast concrete frame structures. One connection (Type A) represented the minimal modifications to the original connection required to reduce the eccentricities in the tensile force transfer path. The other connections, one with a threaded rod connection (Type B) and one with a grouted bar connection (Type C), were designed to both eliminate the eccentricity of the force transfer path and avoid the requirement of welding critical connection components on-site.

Numerical models of the previously tested SMF assembly,^{3,4} with connections replaced by the proposed connection types, were used to evaluate the robustness of the proposed connection designs. The performance of each of the frames was compared to an experimentally validated model on the SMF assembly with the prototype connections. The robustness index, a measure of the dynamic load carrying capacity of a system in comparison with service gravity loads, was used to quantify increases in structural robustness of the assemblies with the alternative connections. The assemblies with the Type A, Type B, and Type C connections had robustness indices 50 %, 130 %, and 220 % larger than that of the assembly with prototype connections. The proposed connection types show potential for enhancing the robustness of precast concrete frame structures. Full-scale experimental investigation of these new connection types is planned.

ACKNOWLEDGEMENTS

Financial support for this work from the Precast/Prestressed Concrete Institute (PCI) is gratefully acknowledged. An Advisory Group organized and chaired by Roger Becker of PCI provided valuable input in the development of the new connections presented in this paper, including key suggestions for improving constructability. The following precast concrete experts served as members of the Advisory Group: Craig Barrett (LEAP Associates International, Inc.), Ned Cleland (Blue Ridge Design, Inc.), Harry Gleich (Metromont Corp.), S.K. Ghosh (S.K. Ghosh and Associates, Inc.), Sam Kakish (Enterprise Properties, Inc.), Walter Korkosz (The Consulting Engineers Group, Inc.), Alex Mihaylov (Vector Structures), Clay Naito (Lehigh University), Spencer Quiel (Lehigh University), Perry Schram (Pennoni Associates, Inc.), Kim Seeber (Seaboard Services of Virginia, Inc.), and Aaron Vnuk (Precast Services, Inc.).

DISCLAIMER

Official contribution of the National Institute of Standards and Technology; not subject to copyright in the United States.

REFERENCES

- 1. Sadek F., Main J.A., Lew H.S., Robert S.D., Chiarito V.P., and El-Tawil S. "An experimental and computational study of steel moment connections under a column removal scenario." *NIST Technical Note 1669*, National Institute of Standards and Technology, 2010.
- 2. Lew H.S., Bao Y., Sadek F., Main J.A., Pujol S., and Sozen M.A. "An experimental and computational study of reinforced concrete assemblies under a column removal scenario." *NIST Technical Note 1720*, National Institute of Standards and Technology, 2011.
- Main J.A., Bao Y., Lew H.S., Sadek F., Chiarito V.P., Robert S.D., and Torres J.O. "An experimental and computational study of precast concrete moment frames under a column removal scenario." *NIST Technical Note 1886*, National Institute of Standards and Technology, 2015.
- 4. Lew H.S., Main J.A., Bao Y., Sadek F., Chiarito V.P., Robert S.D., and Torres J.O. "Performance of precast concrete moment frames subject to column removal: I. Experimental study." *PCI Journal*, 2017 (in press).
- 5. American Society of Civil Engineers (ASCE). "Minimum Design Loads for Buildings and Other Structures" *SEI/ASCE 7-10*, 2010.
- 6. Weigand J.M, Bao Y., and Main J.A., "Acceptance criteria for nonlinear alternative load path analysis of steel and reinforced concrete frame structures." *Proc., Structures Congress 2017*, 2017.

- 7. Bao Y., Main J.A., Noh S. "Evaluation of Structural Robustness against Column Loss: Methodology and Application to RC Frame Buildings." *Journal of Structural Engineering*, 2017.
- 8. Bao Y., Main J.A., Lew H.S., and Sadek, F. "Performance of precast concrete moment frames subject to column removal: II. Computational analysis." *PCI Journal*, 2017 (in press).
- 9. Steuck, K.P., Eberhard, M.O., Stanton J.F. "Anchorage of Large-Diameter Reinforcing Bars in Ducts," *ACI Structural Journal*, V. 106 No. 4, 2009, pp. 506-513.
- 10. Federal Highway Administration. "Evaluation of LS-DYNA concrete material model 159" *Publication No. FHWA-HRT-058-063*, 2007
- 11. Hallquist J. "LS-DYNA keyword user's manual" *Livermore Software Technology Corporation*, 2007.
- Weigand J.M. and Berman J.W. "Integrity of Steel Single Plate Shear Connections Subjected to Simulated Column Removal." *Journal of Structural Engineering*, V. 140, No. 5, 2014.

Identification of Industrial Robot Arm Work Cell Use Case Characteristics and a Test Bed to Promote Monitoring, Diagnostic, and Prognostic Technologies

Brian A. Weiss and Alexander Klinger

National Institute of Standards and Technology, Gaithersburg, Maryland, 20899, USA brian.weiss@nist.gov alexander.klinger@nist.gov

ABSTRACT

The National Institute of Standards and Technology (NIST) is performing research to advance the state of the art in monitoring, diagnostic, and prognostic technologies (collectively known as prognostics and health management (PHM)) to enhance decision-making at the factory floor to promote smarter maintenance and control strategies. One specific thrust in this hierarchical research is focused at the work cell level. A robot system is the focus of this research level where the manufacturing community would benefit from measurement science (e.g., performance metrics, test methods, reference datasets, software tools) to design, deploy, verify, and validate Prognostic, Health Management, and Control (PHMC) technologies aimed at a robot system work cell. NIST's identification of representative manufacturing robot work cell use cases will provide the foundation for which it will construct its own physical test bed. The test bed is designed to emulate the chosen robot system use case and afford sufficient flexibility to add, subtract, or upgrade components and capabilities to be commensurate with common industrial practices. This paper presents various use case options that NIST has considered and highlights the one that will be the foundation of the physical test bed. Additionally, the initial test bed design is introduced.

1. INTRODUCTION

Manufacturing factory floor operations are becoming increasingly complex, yet more efficient and/or flexible with the inclusion of advanced technologies. Keeping factory floor technologies at efficient operations requires thoughtful maintenance and control practices (Lee, Ghaffari, & Elmeligy, 2011) (Jin, Siegel, et al., 2016). One such factory

floor level technology that may bring greater efficiency, yet add more complexity and maintenance (dependent upon what system or technology the robot is replacing) is the industrial robot system. They are becoming more widely used within many manufacturing environments including those that build automobiles, aircraft, and consumer-electronic goods (DeVlieg, 2010) (Kahan, Bukchin, Menassa, & Ben-Gal, 2009) (Kusuda, 1999) (Zwicker & Reinhart, 2014). Maintenance practices are critical to keeping industrial robot systems running at necessary efficiencies and accuracies to enable manufacturing process productivity and quality targets. This effort defines a robot as an industrial robot and a robot system as an industrial robot system according to definitions specified in ISO Standard 8373 (International Organization for Standardization, 2012) . These definitions are:

- 2.9 Industrial Robot - automatically controlled, (2.4), multipurpose reprogrammable (2.5),manipulator (2.1), programmable in three or more axes (4.3), which can be either fixed in place or mobile for use in industrial automation applications
- 2.15 Industrial Robot System system comprising industrial robot (2.0), end effectors (3.11) and any machinery, equipment, devices, external auxiliary axes or sensors supporting the robot performing its task.

Prior efforts have provided case studies to determine the current state of maintenance practices within manufacturing environments (Jin, Weiss, Siegel, & Lee, 2016) (Helu & Weiss, 2016). Both large and small- to medium-sized manufacturers (SMMs) were examined. In short, the case study findings illuminated that both large manufacturers and SMMs would benefit from both advanced maintenance and control strategies with respect to their manufacturing operations, including those featuring robot systems.

The U.S. National Institute of Standards and Technology (NIST) has started the Prognostics, Health Management, and

Brian A. Weiss et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Control (PHMC) project to generate measurement science products (e.g., performance metrics, test methods, reference datasets, software tools) to promote the design, deployment, verification, and validation of monitoring, diagnostic, and prognostic technologies at the factory floor level to enhance decision-making to yield smarter maintenance and control strategies (National Institute of Standards and Technology, 2017). This paper documents NIST's efforts to further develop use cases and construct a test bed to enable the development of performance metrics, test methods, reference datasets, use case scenarios, and software tools for an industrial arm robot work cell. Section 2 presents NIST's research focus with respect to industrial arm robot system PHMC. Section 3 highlights several use cases and key characteristics that have been identified through discussions and site visits with various manufacturers. Section 4 details the initial test bed use case configuration. Section 5 provides information on major components being used in the initial test bed configuration. Lastly, Section 6 concludes the paper and discusses future efforts.

2. RESEARCH FOCUS

NIST's PHMC project is influenced by input from manufacturing stakeholders including end-users, technology integrators, technology developers (both hardware and software), academic institutions, and other government organizations (Pellegrino, Justiniano, Raghunathan, & Weiss, 2016) (Brian A. Weiss et al., 2015) (Jin, Siegel, et al., 2016). Part of this input is coming from SMMs (Helu & Weiss, 2016). NIST's research approach can be broken down into three levels:

- Component Level Machine Tool Linear Axes Diagnostics and Prognostics (Vogl et al., 2017)
- Work Cell Level Health and Control Management for Robot Systems (Qiao & Weiss, 2017) (Brian A Weiss & Qiao, 2017)
- System Level Manufacturing Process and Equipment Monitoring (Helu & Hedberg, 2015)

Manufacturers have successfully deployed industrial robots in a variety of configurations to accomplish a wide range of tasks. In the past, it was common for a robot to perform the same operations throughout its lifespan. In today's evolving economy where manufacturers need to be responsive to everchanging consumer demands, robots are often reconfigured to perform a wide range of tasks over their lifespan (Link, Oliver, & O'Connor, 2016). Or a robot may be called upon to vary its operations on a regular basis. For example, a robot may be asked to use multiple tools during its "work day" where it could be installing vehicle seats in one instance with one type of end effector and then placing a windshield onto a car frame with another type of end effector (Wired, 2013). These changing demands make it challenging to ascertain the robot arm's health. Likewise, the integration and change-over of technology, whether planned or unplanned, makes it challenging to determine the health degradation of the overall robot system or work cell.

Installed robot systems must be robust, especially when used in high-volume applications, and must undergo preventative (or predictive) maintenance to ensure their performance reliability. Robot systems are becoming more complex, especially with the inclusion of additional sensors (to offer more awareness and intelligence to better respond to faults and failures), safety systems (to promote more operations where robots and humans work in close proximity to one another), and end-effector variants (this is highlighted when a robot is capable of quick-changing tools allowing it to increase its operations). More industries are utilizing robotics to perform a wide-range of operations. Monitoring may be leveraged to minimize the occurrence of faults and failures. This is becoming more critical as a robot system increases the diversity of its operations. Data captured across multiple cycles (of differing activities or parameters) is less likely to support "apples-to-apples" comparisons.

NIST's work cell research requires the identification of numerous use cases that can be represented in test methods that are realized within a physical test bed. NIST is constructing such a test bed where the initial configuration should be complete and active by late 2017. The construction of this test bed will promote development of measurement science including performance metrics, test methods, reference datasets, and software tools that will ultimately support open standards and guidelines.

3. USE CASES AND CHARACTERISTICS FOR ROBOT SYSTEMS

Designing appropriate use cases for which to build the health and control management of robot systems research upon is critical to generate output that is relevant and valuable to industry. Considerable effort has been put forth to understand numerous industrial arm robot system use cases that are currently active in industry. This section highlights several of those use cases including those used by large and/or SMMs. This effort took shape through numerous case studies that the NIST research team undertook. Several case studies were documented as a joint effort between the NIST, the University of Cincinnati, and the University of Michigan -Ann Arbor (Jin, Siegel, et al., 2016) (Jin, Weiss, et al., 2016). Likewise, this research was also complemented by case studies that focused on SMMs (Helu & Weiss, 2016). A use case for the overall project has been developed that brings together the component, work cell, and system research levels. This use case involves a mix of machine tools, robot systems, and inspection equipment that must interact together to perform a process (Brian A Weiss, Helu, Vogl, & Qiao, 2016). This use case would support a mix of diverse manufacturing technologies and is represented in Figure 1.



Figure 1. Overall Use Case for NIST PHMC Research (Brian A Weiss et al., 2016)

The remainder of this section outlines numerous use cases that the PHMC team aims to represent in its research to ensure relevance to the manufacturing community's needs. By no means do the highlighted use cases represent an exhaustive list; these use cases are indicative of common robot work cell activities and/or present challenges making advanced monitoring, diagnostics, and prognostics advantageous to their overall effectiveness.

3.1. Time-Based vs. Event-Based Tasks

Every robot task and sub-task is executed based upon a specific lapse in time, the occurrence of an event, or combination thereof. For this effort, time-based tasks can be defined as a robot waiting a set period of time between tasks or from an idle state to an active state. An example of this could be a robot holding a freshly-glued panel in place for a specific time to ensure sufficient adhesion between the set surfaces. An event-based task is defined as a task that is performed only after the robot has received an input, whether it be from an operator, sensor, or other automation signal before proceeding to its next task. An example of an eventbased task would be a robot waiting for an input from a proximity sensor alerting the robot that a box is present and ready to be moved into another location. A hybrid task is one that is governed by both the completion of an event or lapse of a specified period of time. For example, a robot that is holding two glued parts together, may have to hold the pieces together for a minimum amount of time until it releases its grip and repositions itself to hold two more pieces together. However, if the subsequent parts are not ready to be mated, then the robot could idle until a 'parts ready' signal is received. The NIST robot work cell use case will feature a combination of time-based and event-based tasks to ensure that a range of variations are represented in the course of verifying and validating PHMC technologies that inform on the health of the work cell. For example, many manufacturers, both large and small, leverage industrial robot

systems to move objects from one position to another position within a given work volume. Industries, including those within the automotive and aerospace supply chains, and consumer-electronic goods, use industrial robot arms for a range of material handling operations such as part transfer, packaging, and palletizing. These operations include both time-based and event-based tasks.

3.2. Positioning vs. Compliance vs. Allowance

All industrial robot operations feature some type of movement given that an industrial robot is defined to include a manipulator and is programmable (i.e., controllable) in at least three degrees of freedom. Given this definition, all industrial robots are capable of being positioned at specific locations within their work volume. The accuracy of this positioning capability is governed by numerous factors, including a robot's physical specifications (e.g., motors, gears, joint encoders) and controller (at minimum, the robot's specific controller).

In many operations, robot positioning is supplemented by additional active or passive motion capabilities to increase the robot's accuracy and/or increase task success. Positional allowance in this context is defined as the acceptable positional relationship deviation from nominal between two components. In every interaction between a robot's end effector and a work piece, and a work piece and fixturing, there is positional deviation allowance (acceptable deviation from nominal relationship). For example, a robot is to insert a round peg into a round hole with twice the peg's diameter (e.g., the peg has a diameter of 2 cm while the hole has a diameter of 4 cm). The nominal insertion position of the peg is the center of the hole. However, the robot may not achieve the position precisely, where it can be off center by less than the radius of the peg. The insertion is still successful because there was allowance (clearance in this case) in the hole that the peg was inserted in.

When the positional allowance between a tool and a workpiece is less than the accuracy of the robot performing the positioning, or information provided to the robot for positioning (e.g., assumed positions, imprecise sensor information), compliance is required to ensure task success. There are two forms of compliance: passive and active. Active compliance is typically enabled by external sensors providing position feedback (e.g., force, vision, distance sensors) and actuation positioning actuators (possibly the robot drives themselves) and can be programmed to automatically occur within a given operation. Active compliance can be considered micro-positioning once a robot has positioned itself within a designed area. Passive compliance is enabled by flexible components which typically undergo elastic deformation and cannot be programmed (e.g., spring-loaded break away, flexible

brushes). Compliance is often used to augment robot positioning to create an allowable relationship between the end effector and workpiece or workpiece and fixture when the workpiece is being positioned by the robot.

A machine tending operation can be an example of a robot positioning task that highlights both active compliance and allowance (see Figure 2). The machine tending task requires a robot to pick up parts from an incoming conveyor belt and place them on a fixture within a machine tool. In this specific example, an industrial robot arm, with an attached parallel gripper, is mounted to a rigid base where incoming parts, the machine tool's work fixture, and the finished part's outgoing mode of transport (to the next station) are within the robot's work volume. Material handling operations are typically not high-precision operations in that the opening and/or operations of the end-effector and the interaction between part and fixtures provide allowance for positional deviation from nominal. The part in question is a 50 cm cube and needs to be picked by the parallel gripper which opens to 60 cm. Incoming parts (e.g., cubes) on the conveyor belt are spaced approximately 100 cm apart and are resting in a random orientation. The robot knows the approximate (assumed) location of the part on the incoming conveyor belt and after each part is removed from the conveyor belt, it automatically moves 100 cm forward. The machine tending operation begins with the robot moving to a position above the assumed position of the part. Once the robot and conveyor are in position for the next pick, an external vision system (compliance sensor) measures the part's precise position and orientation (more precise then the assumed position) and automatically provides this information to the robot's controller so the robot may refine its location, complying to the part location. Once the robot is positioned based on the vision system's information, the robot can position the gripper jaws around the part and close to complete the pick operation. Because the part can slide freely on the conveyor top while the gripper jaws are closing, the lateral position allowance in the gripping direction is +/- 5 cm from center, and imperfect positioning and orientation below the allowance will not cause the pick to fail. This external sensor's feedback offers an automated means of compliance to increase the accuracy of the task and reduces the allowance needed. Even with the vision system's additional information, the +/- 5 cm of allowance provides an additional margin of error for the task to be successful simply by closing its jaws.

Compliance and allowance are seen in other operations. The use of vacuum grippers offer another means of allowance in material handling operations. A vacuum gripper may have a suction area of $x \text{ cm}^2$ while the top of the box it is picking up has an area of y cm^2 where y > x. It is likely optimal for the gripper to pick the box from its center, yet the gripper can still be successful in picking up the box if it is slightly off-center (the robot system may not be successful in moving or placing



Figure 2. Machine-tending Operation (Fotolia)

the box if its off-centered position causes it to become unbalanced within the gripper).

Industrial arm robot systems are supporting more precise operations, when paired with active compliance end-effectors and/or sensors that promote a finer degree of positioning as compared to the robot's movement alone. In some instances, this capability is allowing the overall work cell to outperform human counterparts in terms of accuracy, repeatability, and reliability. Another example of an active compliance endeffector would be a complementary drilling tool mated to the end of a robot. The robot moves the drill to an exact location near the surface it needs to drill and the end-effector drill would provide the necessary plunging movement and force to produce the desired hole. When the position of the hole must be more precise than the robot is, the end-effector may be capable of further positioning of the drill with onboard actuators (e.g., servos and lead screws) (once the robot has placed the drill in an approximate location) and performing the drilling/plunging task.

Overall, many common manufacturing industrial robot operations are enabled by some combination of positioning, compliance, and allowance. Welding, gluing, 3D-printing, fastening, and mating operations highlight many successful examples of industrial arm robots, external axes, external sensing, supporting automation, and controllers working in harmony. However, each manufacturing process and constituent components mentioned in this section are subject to some form of health degradation over time. If left unchecked, this health degradation will compromise the success of the operation, impacting quality, productivity, etc. when allowances are exceeded and limits of compliance are reached.

3.3. PHMC Robot Work Cell Use Case

All NIST use cases involve an end-effector on the robot arm, the robot controller, a human-machine interface (HMI, which may be directly integrated with the robot controller), and some level of safeguarding. Some of the robot work cell

configurations will include additional elements that range from a programmable logic controller (PLC), sensors (e.g., camera systems, or other sensors measuring part presence or specific location information), external axes, and supporting automation (e.g., conveyor belt bringing parts into the work cell or taking away completed parts and actuated fixtures). The presence, or lack thereof, of these elements is influenced by many factors, including complexity of the desired process, number of robots, variability of the process, number of system inputs and outputs (sensors and actuators), and proximity of human operators.

The initial NIST PHMC use case consists of two robots working together to accomplish a specific operation (Brian A Weiss et al., 2016). One robot is configured for material handling (see Figure 3) and is used to transport parts into and out of the second robot's work volume. The second robot performs a simulated precision operation (e.g., the robot touching the center of a part with a tool tip that leaves a mark on the part). This precision operation is intended to represent a precise manufacturing operation (e.g., welding, machining). A supervisory PLC is used to monitor and coordinate the movement between robots. When running nominally, the PLC will manage the status of each part in the process and command both robots to take particular actions. Each robot controller contains actions that may be executed by the PLC. This includes Cartesian waypoints and motion profiles.



Figure 3. PHMC Material Handling Robot

The development of this test bed will allow researchers to develop PHMC methods on real equipment and verify that methods requiring physical components such as test artifacts can be integrated into a real process and perform as intended. The test bed will also enable the examination of hardware resilience under fault and failure conditions, provide the ability to stress specific elements within the work cell, and simulate various faults and failures in certain components to understand how these errors propagate through the system and impact the overall work cell.

Edge cases may also exist in a real system that would be hard to implement in simulation even if they are known. The availability of information and limits of communications technologies are real constraints in implementing PHMC in existing systems. By using a physical test bed, all of the limits of the work cell components must be dealt with, ensuring a solution that is applicable to manufacturers.

4. TEST BED CONFIGURATION

The robotics test bed at the NIST is a reconfigurable work cell with multiple robots performing different tasks. This work cell allows NIST to evaluate the use case presented in Section 3 and future iterations of the use case. Figure 4 shows the PHMC robot work cell, still under construction.



Figure 4. PHMC Robot Work Cell (Under Construction)

The initial configuration for the test bed includes manufacturing relevant equipment, including two industrial robot arms, a task-specific tool on each robot, fixtures to hold test parts and artifacts, a supervisory PLC, a human-machine interface, safety systems, representative parts and test artifacts for the robots to interact with, and optical tables to fix the equipment to in a controlled way. Future configuration plans include the addition of a conveyance system(s) and a vision system(s).

One robot will be configured for material handling operations with a gripper type end-effector. The second robot will be configured to perform a precision operation with the chosen end-effector and test parts. Parts being moved and operated on inside the work cell will feature geometries that make them representative of parts typically found in the use cases outlined in Section 3. It is anticipated that test artifacts, engineered geometries designed to aid in test methods currently under development, will also be present in the work cell.

The material handling robot will perform pick and place operations, including moving parts from an input area to inprocess work fixtures. Once parts are placed in/on the work fixtures, the second robot will interact with the part in a specified precise manner. When the precision operation is completed, the material handling robot will then move the completed part to an output. The general layout for this can

5

be seen in Figure 5. A PLC will be used to monitor all work cell input and output sensors beyond integrated robot and end effector sensors and coordinate these tasks Specific sensors are still being determined for integration and monitoring within the work cell and could include presence detection and distance measurement sensors. It is anticipated that in addition to moving the representative parts around and performing simulated high accuracy tasks, the material handling and precision operation robots will also interact with test artifacts during the execution of test methods.

A HMI consisting of a digital monitor(s) and physical buttons will be built to provide a means of operator control and feedback. The HMI will contain operator controls and provide system status information. System status information will include process control information as well as health state information that is obtained when executing test methods under development.



Figure 5. PHMC Robot Work Cell (Upon Completion)

Numerous configuration constraints exist on the test bed outside of the PHMC project's control. One of these constraints is the physical size of equipment being used. The robots and all other equipment must safely fit into the available lab space. This limited robot selection to the smallmid size range. Other constraints such as door openings and elevator size were taken into consideration when selecting optical tables and other large pieces of equipment.

Though the robots that will be used in the PHMC lab are relatively small (in the industrial robot domain), the use case, robots, and processes are scalable to other sizes of equipment found in industry. Using small robots will allow for optimal physical access during integration and development. Researchers in the lab will have physical access to all components of the work cell without the need for additional access and safety equipment such as ladders and fall protection. Likewise, smaller benchtop robots would force researchers to work in limited space, adding additional complexity in integration, reducing flexibility in end effector and sensor selections and configurations.

5. TEST BED COMPONENTS

Each of the following test bed components are actively being specified or integrated (where the specification and procurement has been completed). It is important to note that each of the following components is a potential source of faults or failures within the work cell. These potential faults and failures are not discussed in this publication, yet will be discussed in future work.

5.1. Industrial Robots

Two lightweight, industrial, six degrees of freedom (6DOF) robotic manipulators comprise the robotic component of the test bed. One robot has a rated lift capacity of 3 kg and a reach of 0.5 m, while the other has a rated lift capacity of 5 kg and a reach of 0.85 m. These robots are chosen because they are (1) 6-DOF articulated arm robots used in industry, (2) flexible robots that can be integrated quickly in the PHMC robotics test bed, allowing the focus to be on test method development, (3) the physical size of the robots work well in the laboratory environment. (e.g., larger robots would not be able to exercise a full range of motion in the PHMC Laboratory, while smaller robots would make integration of additional sensors more challenging), (4) used in multiple labs at NIST and an in-house expertise is established allowing for technology transfer and sharing of resources between research projects, and (5) designed to be used safely among humans with minimal additional safety requirements, limiting the need for expensive safety equipment, costly integration, and complex operating procedures.

5.2. End-Effectors

The On Robot RG2 Gripper was selected as the end effector for the material handling operation in the initial test bed configuration (Figure 6 shows the RG2 mounted to the larger robot within the PHMC test bed). The RG2 is relatively low cost and comes with a software package for the robots making it quick to integrate and operate, limiting integration time and complexity. The RG2 is an electrically actuated gripper that is grip force and/or position controllable. This allows the gripper to be used in both time-based control and

event-based control schemes. The gripper is capable of being controlled though the robot controller or the PLC.

Figure 6. Material Handling Robot with an RG2 Gripper

A fixed end effector was selected for the simulated precision operation in the initial configuration (this end effector is still in the design stages). The end effector will have feature(s) that will allow for test method verification. These features may include visual fiducials that can be tracked, or specific geometry that interacts with parts and artifacts. Additional end effectors may be integrated into the PHMC Robotics Test Bed in future configurations.

5.3. Fixture(s)

Fixtures for the test bed will be manufactured at NIST to work with the parts and artifacts being used. Extruded polymer 3D printers will produce the test bed fixtures. This will allow for quick turnaround on design variations necessary in the development of test methods.

5.4. PLC/Process Controller

A Beckhoff PLC was selected to be the PHMC Robotics Test Bed process controller. The PLC was chosen for supervisory process control over using the robot controllers (to communicate with each other directly) because it is a flexible hardware and software device that can be reconfigured for use in future test bed configurations. A PLC also allows for greater flexibility in the amount and types of sensors and data that can be monitored and recorded throughout a process. The selected platform is sufficient to control multiple test bed configurations while running NIST-developed software without the need for additional computers, wiring, and communication integration. The PLC will also serve as the historian, recording process information, raw sensor data, and output data from PMHC test methods under development.

5.5. HMI

The initial configuration of the test bed will include a simple HMI consisting of a few buttons and switches used to start and stop processes, acknowledge faults, and select operating mode, along with a visual display which will be used to visualize system status. This display may also display data related to PHMC test methods being developed on the test bed. In the initial configuration, the visual display will be managed by the PLC. Future plans include the incorporation of a standalone HMI integrated with the test bed through the PLC.

5.6. Safety Systems

Industrial safety equipment including safety-rated sensors and a safety PLC will be used to protect those working around the test bed. The safety sensors will include safety laser scanners, safety floor mats, emergency stops, and visual indications of safety status. The system safeguards the work cell in compliance with the specifications in ISO 10218-2:2011 (American National Standards Institute, 2007).

5.7. Additional Work Cell Elements

Other elements are present within the work cell, are in the process of being added, or are being planned for future configurations.

- Optical tables - currently present within the work cell. Optical tables were chosen as the mounting platform for the test bed because they provide a stable base, and can be used to support a convenient common coordinate system for the robots and sensors.
- External Sensors in the process of being added and being planned for future configurations. In the initial configuration of the test bed, only sensors which are needed to perform the use case and collect health information deemed valuable are used. This includes monitoring of sensors embedded in the robots and end effectors and operator controls. As the work cell becomes more "intelligent," more sensors will be added. Future plans include additional sensing to allow the material handling robot to accurately pick a randomly placed part on a moving conveyor. This may include a vision system or other part detection sensors.
- Conveyor System(s) being planned for future configurations. A conveyor system will be used to input raw parts and output finished parts to and from the work cell.

6. CONCLUSION

The fusion of emerging technologies with innovative manufacturing operations has increased the complexity of processes and equipment found on the factory floor. To maintain their competitive advantage, manufacturers must

Klinger, Alexander; Weiss, Brian. "Identification of Industrial Robot Arm Work Cell Use Case Characteristics and a Test Bed to Promote Monitoring, Diagnostic, and Prognostic Technologies." Paper presented at Annual Conference Of The Prognostics And Health Management Society 2017, St. Petersburg, FL, United States. October 2, 2017 - October 5, 2017

SP-233

minimize both the planned and unplanned downtime of their manufacturing processes and supporting equipment. PHMC is offering the manufacturing community a way of enhancing monitoring, diagnostic, and prognostic capabilities to stay "one step ahead" of faults and failures.

To promote greater viability and adoption of PHMC technologies, NIST is conducting research at the component, work cell, and system levels. NIST's PHMC research of robot work cells focuses on current industry challenges in maintaining the health of industrial arm-based automated processes. NIST has identified representative manufacturing robot work cell use cases. These use cases are serving as the foundation for NIST to construct its own use case and physical test bed. The initial configuration has been specified and is being integrated. In parallel, additional capabilities (e.g., external sensors) are still being specified and are expected to be integrated in the future. This will provide a sufficiently flexible and expandable test bed where PHMC methods can be developed, verified, and validated and reference datasets can be collected. NIST is actively constructing the test bed and is expected to have the initial configuration complete and operational in late 2017. As the test bed is being constructed, test methods are being designed that will leverage the test bed's capabilities and support the further development of the use case. These test methods will be presented in future work.

ACKNOWLEDGEMENT

The authors would like to acknowledge Marco Capraro, Jared Kaplan, Josh Chin, Jonathan Garner, and David Schmitt. Each of them supported the fabrication of the PHMC Lab and robot work cell test bed in a variety of ways.

NIST DISCLAIMER

The views and opinions expressed herein do not necessarily state or reflect those of NIST. Certain commercial entities, equipment, or materials may be identified in this document to illustrate a point or concept. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

REFERENCES

- American National Standards Institute, I. (2007). Robots for Industrial Environment - Safety requirements - Part 1 - Robot Terms and definitions (Vol. 10218-1-2007, pp. 44). Ann Arbor, Michigan: Robotic Industries Association.
- DeVlieg, R. (2010). Expanding the use of robotics in airframe assembly via accurate robot technology. SAE Int. J. Aerospace, 3(1), 198-203.
- Helu, M., & Hedberg, T. (2015). Enabling smart manufacturing research and development using a

product lifecycle test bed. Procedia Manufacturing, 1.86-97.

- Helu, M., & Weiss, B. A. (2016). The current state of sensing, health management, and control for small-tomedium-szed manufacturers. Paper presented at the ASME 2016 Manufacturing Science and Engineering Conference, MSEC2016.
- Jin, X., Siegel, D., Weiss, B. A., Gamel, E., Wang, W., Lee, J., & Ni, J. (2016). The present status and future growth of maintenance in US manufacturing: results from a pilot survey. Manufacturing Rev., 3, 10.
- Jin, X., Weiss, B. A., Siegel, D., & Lee, J. (2016). Present Status and Future Growth of Advanced Maintenance Technology and Strategy in US Manufacturing. International Journal of Prognostics and Health Management, 7 (Special Issue on Smart Manufacturing PHM)(Sp5), 18.
- Kahan, T., Bukchin, Y., Menassa, R., & Ben-Gal, I. (2009). Backup strategy for robots' failures in an automotive assembly system. International Journal of Production Economics, 120(2), 315-326. doi:http://dx.doi.org/10.1016/j.ijpe.2007.09.015
- Kusuda, Y. (1999). Robotization in the Japanese automotive industry. Industrial Robot, 26(5), 358-360. doi:10.1108/01439919910283786
- Lee, J., Ghaffari, M., & Elmeligy, S. (2011). Selfmaintenance and engineering immune systems: Towards smarter machines and manufacturing systems. Annual Reviews in Control, 35(1), 111-122. doi:10.1016/j.arcontrol.2011.03.007
- Link, A. N., Oliver, Z. T., & O'Connor, A. C. (2016). Economic Analysis of Technology Infrastructure Needs for Advanced Manufacturing: Advanced Robotics and Automation. NIST, Gaithersburg, MD) NIST GCR, 16-005.
- National Institute of Standards and Technology. (2017). Prognostics, Health Management, and Control (PHMC). Retrieved from http://www.nist.gov/el/isd/ks/phmc.cfm
- Pellegrino, J., Justiniano, M., Raghunathan, A., & Weiss, B. A. (2016). Measurement Science Roadmap for Prognostics and Health Management for Smart Manufacturing NIST Systems. Advanced Manufacturing Seriess (AMS).
- Qiao, G., & Weiss, B. A. (2017). Accuracy Degradation Analysis for Industrial Robot Systems. Paper presented at the ASME International Manufacturing Science and Engineering Conference, Los Angeles, California.
- Vogl, G. W., Pavel, R., Archenti, A., Winnard, T. J., Mennu, M. M., Weiss, B. A., & Donmez, M. A. (2017). Identification of machine tool geometric performance using on-machine inertial measurements. Paper presented at the 6th International Conference on Virtual Machining

Process Technology (VMPT 2017), Montreal, Canada.

- Weiss, B. A., Helu, M., Vogl, G. W., & Qiao, G. (2016). Use Case Development to Advanced Monitoring, Diagnostics, and Prognostics in Manufacturing Operations. Paper presented at the IMS2016 -Intelligent Manufacturing Systems, Austin, Texas.
- Weiss, B. A., & Qiao, G. (2017). Hierarchical Decomposition of a Manufacturing Work Cell to Promote Monitoring, Diagnostics, and Prognostics. Paper presented at the ASME 2017 International Manufacturing Science and Engineering Conference (MSEC2017), Los Angeles, California.
- Weiss, B. A., Vogl, G. W., Helu, M., Qiao, G., Pellegrino, J., Justiniano, M., & Raghunathan, A. (2015). Measurement Science for Prognostics and Health Management for Smart Manufacturing Systems: Key Findings from a Roadmapping Workshop. Paper presented at the Annual Conference of the Prognostics and Health Management Society 2015, Coronado, CA.
- Wired (Producer). (2013, August 1, 2017). How the Tesla Model S is Made. Retrieved from https://www.youtube.com/watch?v=8_lfxPI5ObM
- Zwicker, C., & Reinhart, G. (2014). Human-robotcollaboration system for a universal packaging cell for heavy electronic consumer goods Enabling Manufacturing Competitiveness and Economic Sustainability (pp. 195-199): Springer.

BIOGRAPHIES



Dr. Brian A. Weiss has a B.S. in Mechanical Engineering (2000), Masters in Engineering (2003), and Ph.D. in Mechanical Engineering (2012) from the University of Maryland, College Park, Maryland. He is currently the Project Leader of the Prognostics, Health Management, and Control (PHMC) project

within the Engineering Laboratory (EL) at the National Institute of Standards and Technology (NIST). Prior to his leadership role in the PHMC project, he spent 15 years conducting performance assessments across numerous military and first response technologies including autonomous unmanned ground vehicles; tactical applications operating on Android devices; advanced soldier sensor technologies; free-form, two-way, speech-to-speech translation devices for tactical use; urban search and rescue robots: and bomb disposal robots. He is currently a member of the American Society of Mechanical Engineers (ASME). His efforts have earned him numerous awards including a Department of Commerce Gold Medal (2013), Silver Medal (2011), Bronze Medals (2004 & 2008), and the Jacob Rabinow Applied Research Award (2006).



Alexander Klinger has a B.S. in Mechanical Engineering (2013) from Lehigh University, Bethlehem. Pennsylvania and a Masters of Engineering in Mechanical Engineering (2017) from the University of Maryland, College Park, Maryland. He is currently a mechanical

engineer working on the Prognostics, Health Management and Control project within the Engineering Laboratory (EL) at the National Institute of Standards and Technology (NIST). Prior to his role at NIST, he was a mechanical engineer at the United States Postal Service (USPS) Engineering Headquarters, working on automation and robotic systems. He has also worked in various engineering roles in the manufacturing and entertainment industry.

Joint Special Interest Group meeting between euspen and ASPE Dimensional Accuracy and Surface Finish in Additive Manufacturing, KU Leuven, BE, October 2017 www.euspen.eu



Investigation of complementary use of optical metrology and x-ray computed tomography for surface finish in laser powder bed fusion additive manufacturing

Jason C. Fox^{1*}, Felix Kim¹, Zachary Reese², Christopher Evans², John S. Taylor²

¹National Institute of Standards and Technology ²University of North Carolina at Charlotte

*jason.fox@nist.gov

Abstract

The development of additive manufacturing (AM) has allowed for increased complexity of designs over traditional manufacturing; however, increased design complexity leads to greater difficulties in post process finishing of the part surfaces. Additionally, uncertainty surrounding the quality of the complex as-built surfaces hinders wide scale adoption of AM. As such, a strong understanding of the as-built surface texture is required to help determine the quality of the resultant part. Complex geometries and internal surfaces create challenges for conventional surface finish metrology, but X-ray Computed Tomography (XCT) has emerged as a candidate since it does not require direct line of sight to the surface being measured. In this work, a comparison of XCT and optical measurements are performed on additively manufactured samples with two main goals: to determine the capability of XCT measurements as a tool for surface finish metrology and to determine if XCT measurements can provide insight into locations of overhangs and undercuts, which are difficult to identify and assess through conventional surface finish metrology and may have a drastic effect on part performance. Samples made from nickel alloy 625 and 17-4 stainless steel were built using a commercially available laser powder bed fusion (LPBF) system. Laser confocal and XCT measurements are performed on the samples and compared to each other and to scanning electron microscope (SEM) images of the parts. The results of this work support the premise that XCT provides useful information about the topography of a AM component. This can complement measurements made using optical methods such as confocal microscopy and also deliver results on porosity, undercuts, and internal surfaces where optical methods cannot reach. However, more work is required to relate XCT capabilities to the Fourier domains more commonly used to describe optical instruments and the associated ISO filter specifications.

Additive manufacturing; laser powder bed fusion; x-ray computed tomography; surface finish;

1. Introduction

Additive manufacturing (AM) has emerged as a key technology for production of high-value and complex parts that reduces time-to-market and cost to manufacture [1]. A key benefit to the AM process is the capability to create lattice structures and highly complex, topology optimized parts that reduce weight. Laser powder bed fusion (L-PBF) in particular has generated a great deal of interest due to the fine focusing optics, layer thicknesses of 20 µm to 100 µm, and the process space (beam powers and travel velocities) over which the system operates that allows for fine detail compared to other AM technologies [2]. Despite this advantage, however, a limiting factor affecting widespread adoption of AM is the as-built surface topography of finished parts. As part complexity increases, the ability to alter the surfaces through a secondary operation (i.e., machining or polishing), either in situ or ex situ, decreases [3]. Thus, improvements in the as-built surface texture has been cited as a key need [1].

Previous studies have focused on optimization of parameters specifically for surface texture as well as determining the effect of as-built surface texture on the resultant part properties. However, most AM literature has focused on the arithmetic mean roughness (Ra) [4], which provides little information as to the physical mechanisms creating features on the surface [5,6]. This, in turn, makes it difficult to develop the functional correlations between measured surface data and part performance (e.g., fracture, fatigue).

In addition to a stronger understanding of the as-built surface texture of AM parts, development of non-destructive evaluation (NDE) techniques such as x-ray computed tomography (XCT) for surface texture measurement is needed. Traditional surface metrology equipment often requires line of sight or adequate clearance to contact the surface with a stylus, which can be difficult to achieve given the highly complex AM geometries [7]. Kerckhofs et al. carried out an initial investigation on feasibility of studying surface roughness with XCT. They presented a high resolution XCT image at 1.5 µm/voxel resolution[8]. Townsend et al. presented a methodology to extract areal surface texture and compare with a focus variation microscopy measurement. The XCT acquisition setting provided 17.3 µm/voxel resolution [9]. Thompson et al. have also aligned the XCT data to conventional measurements (coherent scanning interferometry (CSI) and focus variation (FV) microscopy) and provided a comparison. Two different XCT systems were used which provided resolution of 5.7 μ m/voxel and 5 μ m/voxel respectively[10]. Additionally, Thompson et al. showed a comparison of confocal microscopy (CM), CSI, and FV to XCT data, achieving a resolution of 4.69 µm/voxel, and suggesting the integration of results to "improve understanding of the behaviour and performance of complex surface topography measurement" [11].

The purpose of this research is to investigate the complimentary use of optical metrology techniques and XCT with the primary focus being the benefit that the volumetric XCT data can provide when aligned with the data from conventional

Evans, Christopher; Fox, Jason; Kim, Felix; Reese, Zachary. "Investigation of complementary use of optical metrology and x-ray computed tomography for surface finish in laser powder bed fusion additive manufacturing.

Paper presented at European Society for Precision Engineering and Technology Special Interest Group, Leuven, Belgium. October 8, 2017 -October 11, 2017.
surface metrology equipment. Samples made from 17-4 stainless steel (17-4SS) and Inconel 625 (IN625) are measured using a laser confocal microscope and XCT system. The data from the two systems is then aligned to pinpoint locations on the laser confocal microscope data close to near surface defects to provide more context on numerical analysis.

2. Methodology

Surface height measurements were acquired with a laser confocal microscope system (Zeiss LSM 800¹). All measurements were performed with a 10x objective and 0.5x tube lens or a 20x objective lens. The former creates a lateral resolution of 1.25 µm and the latter creates a lateral resolution of 0.31 µm. Measurements were post-processed with Digital Surf's ConfoMap software to remove outliers and level the surfaces. No other filtering was applied to the data.

X-ray Computed Tomography (XCT) scans were acquired with North Star Imaging (NSI) CXMM 50 metrological CT system. The system is equipped with a 225-kV reflection source including a tungsten target, a rotary stage, and a flat panel detector (127 $\mu m/\text{pixel})$ (Figure 1). A source voltage of 160 kV and current of 80 µA was used. A copper filter (4 mm-thick) was applied to reduce the effect of characteristic X-ray peaks as simulated by TASMICS software (Figure 2) [12]. The average X-ray energy was predicted as 110 keV, which was sufficient to penetrate the 5 mm × 5 mm sample dimension of Inconel 625. The sample was mounted on a plastic sample holder instead of a metal one to reduce potential scattering effect from metal. The voxel size of the XCT data was calibrated based on a calibration object as 10.9 µm for a geometric magnification of 11.65 in each direction. A typical filtered backprojection algorithm supplied by the vendor was used to reconstruct the dataset [13]. XCT acquisition parameters are shown in Table 1.



Figure 1. (a) XCT system setup, and (b) an example image of the sample



Figure 2. Simulated X-ray spectrum of 160 kV and 4 mm-thick copper filter for an air Kerma of 1 mGy

¹ Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards

Table 1. XCT acquisition parameters

Parameter	Value
Voltage	160 kV
Current	80 μΑ
Exposure time	2 s
Filter material/thickness	Cu/4 mm
Number of projection	1000
Source-to-detector distance (SDD)	492.76 mm
Source-to-object distance (SOD)	42.29 mm
Voxel size	10.9 μm

Parts for the analysis were built on the EOS M270 system at the National Institute of Standards and Technology (NIST). Three parts were built for this analysis: two using the commercially available EOS StainlessSteel GP1 (corresponds to US classification 17-4SS [14]) and one EOS NickelAlloy IN625 (corresponds to classification unified numbering system UNS N06625 [15]). It should be noted that the material used for the build was powder reclaimed from prior builds using an 80 µm sieve. It is assumed that the condition of the powder can have a large effect on the surface quality of parts being built and analysis of the powder is currently underway.

For the two parts built using the 17-4SS, the first was built entirely with the default parameters defined by the manufacturer, which will be referred to as "sample 60". The second was built with a beam power of 195 W and beam travel speed of 700 mm/s for the contour passes only (outermost portion of the part) with the rest of the parameters set to the default defined by the manufacturer and will be referred to as "sample 195". This high power contour setting was chosen because prior work by the authors had shown that these settings create a greater variety of surface features [6]. The part built using the IN625 was built with the default parameters defined by the manufacturer and will be referred to as "sample 625". Samples built for the analysis are shown in Figure 3.



Figure 3. Sample 60 (left), sample 195 (center), and sample 625 (right). Scale is in mm.

The samples were built with the same orientation in the build chamber relative to the recoater blade. Surfaces of the samples will be referred to as follows: the top surface is parallel with the build plane and contains the part labels "60", "195", and "625", seen in Figure 3. Side surfaces will be referred to using primary intercardinal directions when looking down at the label on the top surfaces. For example, in Figure 3, the southeast (SE) and southwest (SW) surfaces of each part can be seen and the northeast (NE) and northwest (NW) surfaces cannot be seen.

and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

Evans, Christopher; Fox, Jason; Kim, Felix; Reese, Zachary. "Investigation of complementary use of optical metrology and x-ray computed tomography for surface finish in laser powder bed fusion additive manufacturing.

3. Experiment Results

3.1. Analysis of optical metrology data

Height measurements were taken from the surfaces of the samples using a laser confocal microscope. Each surface was scanned over a 5 x 5 stitch using a 10x objective and 0.5x tube lens to create a 5.89 mm x 5.89 mm view of the surface with 1.25 µm point spacing. An example data set, leveled and outliers removed using the commercially available ConfoMap software, can be seen in Figure 4.



Figure 4. SW surface of sample 60 scanned using a 10x objective and 0.5x tube lens. "NM" indicates non-measured points.

Additionally, each surface was scanned over a 3 x 3 stitch using a 20x objective and 1.0x tube lens to create an 895 µm x 895 µm view of the surface with 0.312 µm point spacing. The scan center was positioned at (2.5,-3) mm assuming the upper left corner of the surface is (0,0). An example data set, leveled and outliers removed using the commercially available ConfoMap software, can be seen in Figure 5.



Figure 5. SW surface of sample 60 scanned using a 20x objective Center of the scan is positioned at (2.5,-3) mm assuming the upper left corner of the surface (seen in Figure 4) is the position (0.0).

3.2. Analysis of XCT data

The reconstructed XCT data were imported into VGStudioMAX 3.0 software and the 3-2-1 registration technique available in the software was used to align the orthogonal faces to the same coordinate system and minimize part-to-part variation during the analysis [16]. Then, a stack of images was exported such that the cross-sectional image shows both SW and NE edges of the sample. The XCT data set can be also converted to a mesh after the surface is determined from XCT grayscale images. The software provides an advanced local adaptive surface detection algorithm which can determine the surface at a sub-voxel level [16]. Lifton et al. [17] estimated the uncertainty of dimensional measurement based on a similar surface detection method to be at sub-pixel level in case of a 2D image. The determined surface can be exported to an STL file for a comparison with other measurement results. Example surface representations are shown in Figure 6.



Figure 6. Example XCT-based surface representations of different AM processing parameters

For the initial characterization, full resolution grayscale XCT images representing cross sections of the samples were converted to binary using a 50 % digital level threshold instead of using the more complex sub voxel interpolated surfaces. The sub voxel interpolated surfaces will be used for the future analysis. The uncertainty of the current surface detection is estimated to be at around a voxel level, but a more comprehensive simulation study would improve the estimation. The binary images are then converted to a "top down" measurement, analogous to how an optical metrology system would obscure the underside of overhangs. This is accomplished by searching through each column of pixels to determine the point at which the image data changes from a value of zero to a value of one. Figure 7 shows examples of the XCT image, binary threshold, and resultant profile highlighted in red.



Figure 7. XCT data (top), converted to binary (middle), with conversion to surface profile (bottom).

Using this method, a 2D array of surface heights can be created and imported into the same commercially available software used to analyse data from the laser confocal system. An example of this data can be seen in Figure 8.



Figure 8. SW surface of sample 60 derived from XCT data. This region is the same general area scanned by the laser confocal system using the 20x objective (Figure 5).

Evans, Christopher; Fox, Jason; Kim, Felix; Reese, Zachary. "Investigation of complementary use of optical metrology and x-ray computed tomography for surface finish in laser powder bed fusion additive manufacturing.

Although prominent particles can be visually identified, the current resolution prevents significant statistical analysis of the XCT data. However, higher resolution scans of the samples are currently underway.

4. Comparison of Results

Images of the surface were taken via scanning electron microscope (SEM) for qualitative comparison. Figure 9 shows an example image from the SW surface of sample 60. The image shows a 1 mm x 1 mm area of the surface that encompasses the area previously scanned with the laser confocal microscope using the 20x objective (Figure 5).



0.5 mm

Figure 9. SW surface of sample 60 taken via SEM. Image is a 1 mm x 1 mm area that encompasses the laser confocal data scanned using the 20x objective (Figure 5).

While the current resolution of the XCT data is limited, it has the distinct advantage of being able to detect pores and undercuts on and near the surface. To detect pores and undercuts, the binary images were used to create a profile in a similar fashion as described in section 3.1. However, instead of the "top down" approach, the profile is determined from approximately 500 μ m below the surface to the surface. This "bottom up" profile is then compared to the "top down" profile and any variations between the two profiles are tracked as pores and undercuts. An example of this methodology is presented in Figure 10 and the resultant pore and undercut locations are presented in Figure 11.



Figure 10. Example of "top down" (red) and "bottom up" (green) profiles for tracking pores and undercuts.



Figure 11. Identification of near surface pores and undercuts on the SW surface of sample 60 using XCT data.

From the identification of pores and undercuts in Figure 11, surface data near these defects can be extracted. The XCT data and surface data taken with the 20x objective were aligned manually by visually aligning prominent features in the false color height maps from Figure 5 and Figure 8 using commercially available image processing software. Pore locations are highlighted in Figure 12. A closer view of the surface data near the defect location in the upper left corner of Figure 12 is shown in Figure 13. While this methodology allows us to align and identify regions where subsurface defects exist, it is still inconclusive if trends in the surface data from the laser confocal system can identify these defects. As such, further analysis of the surface data is ongoing.



Figure 12. Surface data for the SW surface of sample 60, taken via 20x objective on the laser confocal system, with defect areas (pore and undercut locations) highlighted by white boxes.

Evans, Christopher; Fox, Jason; Kim, Felix; Reese, Zachary. "Investigation of complementary use of optical metrology and x-ray computed tomography for surface finish in laser powder bed fusion additive manufacturing."



Figure 13. Surface data taken via 20x objective on the laser confocal microscope for the upper left defect area from Figure 12.

5. Conclusions

In this work, IN625 and 17-4SS samples were built on a commercially available L-PBF system and analyzed via laser confocal microscope and XCT. The data was then manually aligned for complementary analysis. Pores and undercuts were identified using XCT data, which allowed for more stringent analysis of the surface near these features. However, at 10.9 µm/voxel the current resolution of XCT prevents strong quantitative analysis of the surface data from XCT or relevant comparison to the laser confocal microscope surface data. To resolve this issue, higher resolution XCT scans at an anticipated resolution of $\approx 2 \mu m/voxel$ are currently underway. Additionally. the higher resolution will allow for identification of smaller pores and undercuts. While quantitative analysis has been performed on XCT data with larger voxel resolutions, it has required the use of sub-voxel interpolation, which is currently not included in this analysis. Additionally, the work has focused on top surfaces as opposed to side surfaces of the AM parts, which generally contain larger scaled surface features and ease resolution requirements for a strong quantitative analysis. With an increased focus on quantitative analysis forthcoming, an investigation of uncertainty will be required and is planned for future work.

Future work will also include the classification of defects found in the data (i.e., subsurface pores vs undercuts on the surface). and identification of the depth of pores. Additionally, advanced pore detection techniques and image analysis, such as those presented in Kim et al. [18], automated alignment techniques, such as those presented in Thompson et al. [11], and sectioning of the samples for comparison to the XCT and laser confocal microscope data are in progress. It is anticipated that this type of analysis will create avenues to get richer information from the surface data and add a potential methodology for feature based metrology.

6. Acknowledgements

This work was performed in part in the NIST Center for Nanoscale Science and Technology (CNST) Nanofab Facility.

References

- NIST, US Department of Commerce 2012 Measurement [1] Science Roadmap for Metal-Based Additive Manufacturing. Gaithersburg, MD.
- Yadroitsev I 2009 Selective laser melting: direct [2] manufacturing of 3D-objects by selective laser melting of metal powders Saarbrücken (Lambert Acad. Publ)

- Pyka G, Burakowski A, Kerckhofs G, Moesen M, Van Bael [3] S, Schrooten J 2012 Surface modification of Ti6Al4V open porous structures produced by additive manufacturing Adv. Eng. Mat. 14 363-70
- [4] Townsend A, Senin N, Blunt L, Leach R, Taylor J S 2016 Surface texture metrology for metal additive manufacturing: a review Precision Engineering 46 34-47
- [5] Fox J C, Moylan S P, Lane B M 2016 Preliminary Study Toward Surface Texture as a Process Signature in Laser Powder Bed Fusion Additive Manufacturing. Proceedings of the 2016 ASPE Summer Topical Meeting: Dimensional Accuracy and Surface Finish in Additive Manufacturing, Raleigh, NC
- [6] Fox J C, Moylan S P, Lane B M 2016 Effect of Process Parameters on the Surface Roughness of Overhanging Structures in Laser Powder Bed Fusion Additive Manufacturing Procedia CIRP 45 131-4
- [7] Thompson A, Maskery I, Leach R 2016 X-ray computed tomography for additive manufacturing: a review Meas Sci Technol 27 072001
- Kerckhofs G, Pyka G, Moesen M, Van Bael S, Schrooten J, [8] Wevers M 2013 High-Resolution Microfocus X-Ray Computed Tomography for 3D Surface Roughness Measurements of Additive Manufactured Porous Materials Adv. Eng. Mater. 15 153-8
- [9] Townsend A, Pagani L, Scott P, Blunt L 2017 Areal surface texture data extraction from X-ray computed tomography reconstructions of metal additively manufactured parts Precision Engineering 48 254-64
- [10] Thompson A, Körner L, Senin N, Lawes S, Maskery I, Leach R 2017 Measurement of internal surfaces of additively manufactured parts by X-ray computed tomography. 7th Conference on Industrial Computed Tomography Leuven, Belgium
- [11] Thompson A, Senin N, Giusca C, Leach R 2017 Topography of selectively laser melted surfaces: A comparison of different measurement methods CIRP Annals - Manufac. Technol. 66(1) 543-6
- Hernandez A M, Boone J M 2014 Tungsten anode [12] spectral model using interpolating cubic splines: Unfiltered x-ray spectra from 20 kV to 640 kV Med Phys **41** 042101
- [13] Feldkamp L A, Davis L C, Kress J W 1984 Practical conebeam algorithm J. Opt. Soc. Am. A 1(6) 612-9
- [14] Material data sheet - EOS StainlessSteel GP1 for EOSINT M270 2009. https://scrivito-public-cdn.s3-eu-west-1.amazonaws.com/eos/public/5f84f5d2c88ac900/05fb1 582834a38c85ef6dd859733a230/EOS_StainlessSteel-GP1 en.pdf (accessed December 7, 2015).
- [15] Material Data Sheet - EOS NickelAlloy IN625 2011. http://ip-saas-eoscms.s3.amazonaws.com/public/d1327facdca0e32a/373 a60ec4f5c891b7dbcdf572e37d3b0/EOS_NickelAlloy_IN6 25 en.pdf (accessed June 13, 2017).
- Volume Graphics GmbH, VGStudioMax 3.0. [16]
- Lifton J J, Malcolm A A, McBride J W 2015 On the [17] uncertainty of surface determination in x-ray computed tomography for dimensional metrology Meas. Sci. Technol. 26 035003
- [18] Kim F H, Garboczi E J, Moylan S P, Slotwinski J A 2017 Investigation of pore structure in cobalt chrome additively manufactured parts using X-ray computed tomography and three-dimensional image analysis Additive Manufacturing 17 23-38

Evans, Christopher; Fox, Jason; Kim, Felix; Reese, Zachary. "Investigation of complementary use of optical metrology and x-ray computed tomography for surface finish in laser powder bed fusion additive manufacturing.

Hybrid Datafication of Maintenance Logs from AI-Assisted Human Tags

Thurston Sexton, Michael P. Brundage, Michael Hoffman, and KC Morris Information Modeling & Testing Group National Institute of Standards & Technology Gaithersburg, MD, 20899, USA Email: thurston.sexton@nist.gov

Abstract-One of the main challenges of applying AI to certain datasets derives from the datasets themselves being unstructured, unclear, and ambiguous. Furthermore, the insights that are to be gained reflect the quality of the data itself; if the data is skewed, so will be the insights. This problem is not unique to AI technology. People looking back at logs of past events often struggle to understand what was recorded, and to put together a timeline amongst a range of actors. AI technology can help humans sort the data out, but it does not provide the same insight often found in the background knowledge of human participants. This contextual weakness has made unstructured data hard to process. In our work, we have studied typical manufacturing maintenance logs to explore whether and how we can apply AI technologies to gain more insight from this-often vast and under-used-data-source. Our approach combines AI techniques for NLP, machine learning, and statistical processing with human contextual knowledge to quickly develop structured semantics reflecting unique datasets.

Keywords-Natural Language Processing; Intelligent manufacturing systems; Predictive maintenance; Tagging;

I. INTRODUCTION

The typical manufacturing maintenance log consists of composed notes from operators or maintenance technicians on problems encountered, along with some information about the solutions applied. Often, much of the information is left out, and the records contain numerous abbreviations and jargon. Some organizations are better than others about trying to classify these records to track maintenance history, but more often than not the problem details are still described in common language. Maintenance logs are typically only used when there is a recurring or "seismic" problem, at which point a full-fledged investigation is undertaken. Then, it is up to the investigator to try and make sense of what has been recorded. That person brings a good deal of background knowledge to this task, specifically a good understanding of the jargon in these records.

Our research has focused on developing a methodology for extracting useful knowledge from these records using emerging technologies for natural language processing (NLP) and statistical analyses to identify trends in the data. In this work we "datafy" the maintenance logs, making them more useful for statistical insight. We approach this problem through a hybridization-machine learning techniques, augmented with human guidance-to give meaningful structure

to the data. Early results have been promising and are reported on here. First we provide an overview of the problem, followed by our approach and preliminary findings.

A. Why Datafication of Maintenance Logs?

Datafication has been defined as "the transformation of social action into online quantified data" [1]. In the example we present, we take the "social action" of a maintenance technician creating some historical record of events (for the eventuality that it becomes useful to another person), and transform it into data useful specifically for computation. In other words, it is the process of structuring data in a way that adds value to it, by facilitating the transmission of human contextual knowledge used in understanding this data.

Maintenance management in a manufacturing facility is an important part of the manufacturing process, since reducing machine downtime leads to increased productivity. When a machine breaks down and the subsequent maintenance workorder is issued, the associated information is either manually written down, input into a database, or created using maintenance management software. Meanwhile, during the machine downtime, the time spent diagnosing the issue is often larger than the time spent carrying out the repair of the machine [2]. In other words, getting at the knowledge within this work-order data, and using it to diagnose and address problems faster, will mean real improvements to a manufacturing process.

The problems encountered in using this data reflect the human-centric nature of the maintenance activity, and include:

- 1) Technicians often describe problems informally, leading to inconsistencies and inaccuracies in the data.
- 2) Certain maintenance data, such as the actual root cause of a problem, is not always being collected.
- 3) Once data is collected, it is not often subsequently used for future diagnosis.

The first two of these factors can be attributed to a reliance on human background or contextual knowledge. The latter factor is due to a lack of clear and simple procedures for doing so-also a human interaction issue. Our research addresses both aspects of the problem.

The framework presented in [3] shows how these factors can be systematically addressed. In this paper, we explore

Brundage, Michael; Hoffman, Michael; Morris, Katherine; Sexton, Thurston. "Hybrid Datafication of Maintenance Logs from AI-Assisted Human Tags." Paper presented at 2nd Symposium on Data Analytics for Advanced Manufacturing, Boston, MA, United States. December 11, 2017 - December

14. 2017.

 Table I

 EXAMPLE OF TYPICAL INPUT FIELDS IN WORK-ORDER DATA

Input field	Input type
Issue ID	Integer
When work-order was issued	Time/Date
Description of Problem	Raw Text
Asset ID	Protected String
:	
Resolution of Problem	Raw Text
When work-order was completed	Time/Date

how to apply AI techniques to structure the data in a way that will be more easily understandable and processed—this is the datafication of these maintenance logs. The resulting data set contains maintenance work orders that have been consistently tagged for clarity and analysis.

B. The Structure of Maintenance Logs

To understand the problem of datafication for manufacturing maintenance logs, let's consider the data more closely. Maintenance logs, especially when we consider the workorders themselves, contain basic natural language descriptions of the maintenance issues. They're often organized into tabular form, and include some kind of contextual data, as illustrated in Table I.

In other cases, data comes directly from sensors, organized according to some format dictated by the device. While some standards are emerging, at present the devices can vary greatly and the sensor data is essentially a structured set of jargon. In practice, sensor data can be used to facilitate the creation of work orders by monitoring a variety of physical signals created by a machine. Work orders can sometimes be generated automatically if certain sensor measurements indicate the machine is not functioning as expected. These work orders can be used in the same manner as those that are created manually and contain similar data.

For the purposes of this paper, we focus on the humangenerated, natural-language maintenance data—specifically, on "Description of Problem" and "Resolution of Problem" fields commonly found as in Table I. These fields are often a free text description. For example, Technician 1 might describe a "hydraulic leak" as "leaking hydraulic fluid at Machine H1", while Technician 2 might write "hyd leak at cutoff unit of H1." Both of these descriptions represent the same overall problem of "hydraulic leak" located at "Machine H1", but described in two different ways. The same applies for the eventual problem resolution text. These inconsistencies makes it impossible to automatically search a database to find something as simple as "common resolutions for all cases of hydraulic leak".

In an ideal scenario, a maintenance technician would be able to search instances of "hydraulic leak" and find suggested solutions from previous instances. Locating these logs would support the maintenance technician as an augmentation to his existing experience, hopefully aiding to quickly solve the problem. Additionally, this knowledge could give maintenance managers a more accurate way to track problems and corresponding repairs throughout the facility.

The common assumption behind recording these logs is that a human will refer back to them; as such a human is assumed to interpret them. That assumption, however, makes it very difficult for machine computation in any meaningful way. Many manufacturers have recognized this problem, and try to enforce more rigor in these descriptions of problems and solutions, perhaps using interfaces like drop-down menus to control vocabulary. Computerized Maintenance Management System (CMMS) solutions, such as IBM Maximo¹, address some of these issues through drop down menus and enforced categorization [4]. Many other commercial CMMS products, such as Maintenance Connection or eMaint CMMS, offer similar approaches to maintenance management. However, in our conversations, the problems are not completely resolved through these systems; the controlled vocabularies do not fully cover the situational complexity, and much of the useful data for diagnostics is still covered in the comments. The interfaces to these systems are often cumbersome to use and do not commonly succeed in altering technician recording behavior or eventual data structuring. The analysis that can be performed by these CMMS platforms—while certainly powerful-is often limited by the inconsistent structure of the human-generated data. Furthermore, small and medium enterprises (the bulk of the US manufacturing base) often do not have the wherewithal for these solutions.

As a workable solution to these problems, we propose to take the free text descriptions found in these logs and apply the concept of *tagging* to provide context for such data. This allows maintenance managers and technicians to properly analyze the data—perhaps even as a preparation mechanism for available CMMS systems—and thereby obtain diagnostic and prognostic aid from a trove of previous solutions.

II. HYBRID DATAFICATION OF MAINTENANCE LOGS

Underlying the task of datafication is a fundamental need to *classify* the content of available unstructured data, which is to say, provide structure for it. Our challenge is to classify the data from the manufacturing maintenance logs into the diagnostic framework shown in Fig. 1. Each word or token in a maintenance log will be fit into one of these categories. Using this structure we can then identify meaningful patterns in the data. Classification in this context is approached,

"Hybrid Datafication of Maintenance Logs from Al-Assisted Human Tags." Paper presented at 2nd Symposium on Data Analytics for Advanced Manufacturing, Boston, MA, United States. December 11, 2017 - December

14, 2017.

¹Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

generally, in one of two ways: manually, given some framework within which users are allowed to operate and make classification decisions; and via automation, generally using Natural Language Processing (NLP).

Due to the nature of manufacturing maintenance as a natural-language domain, neither of the above are particularly satisfactory approaches. In our experience, the scale of available natural-text maintenance log data exists in the $10^3 - 10^5$ (number of entries) range—this is far too labor-intensive to manually classify under some standardized framework, especially considering the need for timecrunched, domain-expert employees to divert time toward such a menial task. On the other hand, NLP is typically suited toward large corpuses of documents, not the short, technical issue descriptions found in maintenance logs.² Complicating the problem further, these logs are full of highly domain-specific technical terms and user-specific lingo. The classification is very difficult to automate without customized NLP libraries tailored to the jargon, further adding to the burden on manufacturers trying to implement datafication techniques.

Our approach is to hybridize the two paradigms, centering around the use of data *tagging*. "Tags" are simply annotation terms for some resource - in this case, maintenance workorders — and are generally chosen without a controlled vocabulary or boundary [5]. Tagging is very flexible ---being a direct reflection of the tribal knowledge and domain vocabulary of specific users, it can adapt quickly to new domains by simply adding and removing tags. Additionally, tags can be used to reconstruct robust dynamic hierarchies of concepts, sometimes referred to as "folksonomies" [6].

Rather than separating data-points into discrete bins or strict relations, a user of a tagging system simply assigns characteristics directly onto data instances . However, while being more intuitive for humans than strictly unambiguous taxonomic classification, this "fully-manual tagging" methodology is still labor intensive. For this reason we propose a hybrid system, augmenting a human's ability to

²NLP has been applied to short documents with many instances (i.e. Twitter[®]) with general success; however, the size of our datasets are not nearly of that scale, and the language is generally much more technical.

tag natural-language maintenance documents, while preserving the flexibility and robustness of a folksonomy. This is accomplished through easily implemented, foundational techniques from NLP as a way to optimize a human tagger's time investment. In this way, we hope to encourage adoption and understanding of techniques that balance human usability with analytic efficacy for the analysis of largely unused maintenance data.

A set of tags meant to give structure to maintenance logs should in essence list qualities associated with the 1) problem, 2) solution, and 3) items (the objects directly relevant to the issue such as machine, resources, parts, etc.) Consequently, we have adopted a "meta-classification" for the tags themselves, namely, the enumerated classes above. Fig. 1 shows an example tagging in this style, which is accomplished via the process outlined in Section II-A below and in Fig. 2.

A. Importance-based Vocabulary Tagging

It follows from the above discussion that imposing a protected vocabulary when tagging items would be counterproductive; however, a vocabulary list functioning as a lookup table, able to match important concepts to their "tag" representation (without repeated human input), is very wellsuited to technical domains like manufacturing maintenance. NLP allows us to extract an initial vocabulary list which can then be validated and refined by experts. NLP extracts a list of "tokens" ³ that are ranked in order of importance. For the purpose of this study, we turn to a widely used importance metric called Term Frequency & Inverse Document Frequency (TF-IDF), which balances the frequency of occurrence of some word within a document (i.e. issue) and the frequency in the overall corpus of documents. [7]. Passing this list of tokens to a set of domain experts for "meta-classification" is highly efficient for quickly building a list of common vocabulary, as we demonstrate in the following case-study. The job of the experts is to determine

³In the context of NLP, tokens are either individual words, or combinations of words (n-grams) that have a discrete meaning. For example, "ice" "cream", and "ice cream" might all be tokens, with "ice cream" being a 2-gram.



Figure 1. Illustration of a maintenance issue tagging process, mapping from two raw-text inputs-here an issue description and corresponding resolution-to a set of categorized tags. Note the correction of "replaces" to "replace".

"Hybrid Datafication of Maintenance Logs from Al-Assisted Human Tags." Paper presented at 2nd Symposium on Data Analytics for Advanced Manufacturing, Boston, MA, United States. December 11, 2017 - December

^{14. 2017.}



Figure 2. A flow chart illustrating the process from raw-text work-order documents and expert annotation to tags. The key is to pass users a ranked list that is already in (approximately) "tag-like" form, via various pre-processing techniques and ordered by a TF-IDF "importance" heuristic. In the keyword-extraction step, a series of binary decisions determines if incoming tokens are already classified and whether to mark them as Tags or Unknowns.

if the tokens fall into one or a subset of the following categories:

- 1) Item, Problem, or Solution named-entity (NE) tag
- 2) Redundant or un-useful tokens
- 3) Stop-word and Ambiguous tokens.

The NE tag classifications (1) are then treated as definitive information, and all instances of these tokens need no further human input to be correctly tagged. We allow the experts to add an alias or *preferred label* for any of these token classes, thereby automating the creation of a thesaurus that can take local jargon and abbreviations into account.

The redundant tokens (2) do not add additional information, and can be ignored. As an example, consider the recognized tokens from the phrase "bar feeder chain". Extracted tokens from 3-gram TF-IDF might be:

> {bar, feeder, chain, bar feeder, feeder chain, bar feeder chain}

However, in this specific process, a "feeder chain" might not make sense without "bar", and a feeder might not be an object existing outside of "bar feeder". Such a scheme is depicted in Fig. 3. With necessary redundant classifications in place, the resulting tags extracted using expert-classified vocabulary might be

{bar, bar feeder, bar feeder chain}

Similarly, stop words or ambiguous tokens (3) are deemed either non-important filler words or words too ambiguous to give a strict Item/Problem/Solution designation. This process of annotation results in a classification and alias list in order of TF-IDF "importance", as seen in Table II. It is then computationally inexpensive to return to the Bag of Words representation of the original raw text issue descriptions and



Figure 3. An example heuristic rule for determining what extracted tokens might be considered "redundant". Each arrow is a determination if, within the context of a given corpus, some n-gram token makes sense to talk about without its (n+1)-gram parent. If not, it can be ignored, since it will almost always occur within one of the sensible n-gram parents.

Table II EX. TOP TOKENS, CLASSIFICATIONS, AND ALIASES

Token	NE	Alias
replace	S	replace
unit	I	unit
motor	I	motor
spindle	I	spindle
leak	P	leak
valve	I	valve
replaced	S	replace
	.	
:	:	:

flag extracted tokens as useful tags, redundant/ambiguous, or as completely unknown due to incompleteness of the vocabulary list. This automated tagging approach is described in the case study section below using an industry data-set.

Brundage, Michael; Hoffman, Michael; Morris, Katherine; Sexton, Thurston. "Hybrid Datafication of Maintenance Logs from AI-Assisted Human Tags." Paper presented at 2nd Symposium on Data Analytics for Advanced Manufacturing, Boston, MA, United States. December 11, 2017 - December

14. 2017.

B. Human Critique and Completion

Once a first pass has been completed by the automated vocabulary-based tagger, humans may now provide critique of the set of tags for each issue. This can be handled with a combination of approaches.

As a consequence of the vocabulary method above, each maintenance issue will now have a list of **Unknown** tokens. It is then possible to perform a TF-IDF importance ranking *on only the unknown tokens*, then passing this to a human expert for classification. This iterative process is hypothesized to quickly reduce the number of unknown tokens while better completing the vocabulary list than the TF-IDF heuristic ranking alone.

Another possibility, which particularly becomes useful when multiple human agents can tag resources in a dataset, takes a nod from previous taxonomic and crowdsourcing work [8], [9]. If users are to ever manually add or remove tags from individual issues directly, it is important that they are directed in such a way as to

1) maximize the utility of their time investment, and

2) optimize completeness of the issue-to-tag mapping

This can be achieved by maximizing the *expected information gain* over the set of available issues, and queuing issues for the users to tag, ordered in this way.

The human critique may serve another, potentially important, goal: the separation of Items into problem-items and solution-items. These categories are highly ambiguous and the vocabulary list may not be useful to differentiate between them, since the tags for the two categories are nearly mutually-inclusive. Consequently, these must be determined on a per-issue level. The benefits for doing so are a current area of research, especially concerning diagnostic prediction using the tagged data; a minimal example of such diagnostics is further discussed in the *Diagnostics* part of Section III-B below.

III. INDUSTRY CASE STUDY

To test the vocabulary-based tagging procedure outlined above, a manufacturing data-set containing 3,438 raw-text descriptions and resolutions of maintenance issue workorders was analyzed.

First, two domain experts collaboratively tagged as many individual issues as possible, resulting in 1,814 structured data-points via fully-manual tagging. The experts reported being able to tag approximately 100 issues/hour, giving a rough estimate of 18 hrs, 8min to tag that number of issues. They were asked to list the Problem, Solution, and Item tags, further differentiating between problem-items and solutionitems.

Next, they were given a TF-IDF ordered list of extracted tokens, with the task of meta-classification described previously, assigning one of: Item (I), Problem (P), Solution (S), Ambiguous (U), or Stop-word (X), along with any preferred labels to replace lingo, misspellings, etc. The experts

reported finishing 1000 vocabulary items in approximately 1 hour, while ultimately completing 1,362 classifications, giving an estimate of 1hr, 22min spent.

A. Datafication Quality

To estimate the quality of the automated tagging, we assume that any issues with *no* Unknown tags are considered fully "datafied", or tag-complete. Additionally, for issues where no known tags were extracted at all, the datafication process was completely non-useful. The top of Figure 4 reports both of these groups as fractions of the entire dataset, versus the size of the available vocabulary list. The automated tagger datafies nearly all issues, and extracts all available information from over half of the data-set with only 1,000 tokens in its vocabulary.

1) Comparison to Fully-Manual Tagging: Another way to quantify the tag quality is by directly comparing the tagset from the vocabulary tagger to the human-generated fullymanual tags as a test-set. This requires a suitable metric. Typically, accuracy is defined as the ratio of "correct" to "incorrect" label instances, which in this context would be how many times one set of tags perfectly matches the other set. This is an overly-harsh metric for the performance of multi-label classification, since it ignores how close we got to the correct output in each case. The Hamming score (S_H) measures the "distance" between the two sets of tags, averaged over the data-set, which is a more forgiving and useful metric [10]. Given a set of n resources, where the *i*th resource has "true" tags T_i and predicted tags P_i :

$$S_{H} = \frac{1}{n} \sum_{i=1}^{n} \frac{|T_{i} \cap P_{i}|}{|T_{i} \cup P_{i}|}$$
(1)

However, this is somewhat difficult to interpret. Additionally, we can use the more intuitive notions of *precision*, *recall*, and their weighted harmonic mean F_{β} -score.⁴ Simplistically, precision is a classifier's ability to avoid false-positives, and recall is its ability to *not miss* true-positives. F_{β} then combines them, attributing β times more importance to recall than to precision:

$$Pr = \frac{1}{n} \sum_{i=1}^{n} \frac{|T_i \cap P_i|}{|P_i|}$$
(2)

$$Re = \frac{1}{n} \sum_{i=1}^{n} \frac{|T_i \cap P_i|}{|T_i|}$$
(3)

$$F_{\beta} = \frac{1}{n} \sum_{i=1}^{n} (1+\beta^2) \frac{Pr_i Re_i}{\beta^2 Pr_i + Re_i}$$
(4)

Since we do not really trust that the original tags given by humans were all-inclusive, we want to place more importance on recall in our measure — for example, twice as much

⁴There are several ways to aggregate these scores across a data-set; here we have adopted the simple mean over all resources/issues (macro-average).

— leading to our usage of F_2 . The results of calculating these values, assuming the manually-tagged issues define the "true" tag set, are in the bottom of Fig. 4.



Figure 4. The top chart illustrates the return on labor investment for meta-classification of the tag vocabulary. The bottom chart shows the information-retrieval scores for the vocabulary-tagger when calculated against un-assisted human-derived tags.

It is interesting to note that the Precision actually goes down with more vocabulary; this arises from Precisionpunishing false-positives. This implicitly assumes that the human-derived tags are "the whole truth", when in reality these tags only consist of what was deemed important. This, coupled with the way that Precision losses and Hamming scores level out rather quickly with further increases in vocabulary size, indicate that it is likely better to start with the auto-tagging procedure as a "gold standard", since we can be certain extracted tokens *existed* in the original text.

2) NLP Comparative Study: Another common way to assist datafication efforts with NLP is by using Machine Learning models that, for example, mimic the tagging patterns of observed human-generated tags. Generally this requires a training set, i.e. there are resources with existing "true" tags, which are used to train a multi-label classifier. ⁵ This obviously requires some initial time-investment from human taggers to create this training set, but once this is completed, the model has potential to quickly tag remaining



Figure 5. Box-plot comparison of the performance for Vocabulary tagger vs. multi-label SVM classification using Word2Vec semantic embeddings of the raw-text issues. Notches indicate 95% confidence intervals for median score, via 10,000 bootstrap samples.

resources.

For this study, a linear-kernel support-vector machine (SVM) was used in a One-vs-Rest multi-label scheme⁶. Inputs were 1,814 Word2Vec semantic embeddings [12], one for each raw-text maintenance issue tagged by the experts. Target (training) labels were the expert-generated tags themselves. Stochastic Gradient Descent (SGD) was used to minimize the hinge loss, and an elastic-net penalty was used to reduce over-fitting [13]. The model was written in Python using the Sci-kit Learn and SpaCy libraries [14], [15]. Results compared to the vocabulary-based tagger w.r.t. the expert tags are displayed in Fig. 5.

It is apparent that the SVM out-performs the vocabulary tagger in every available metric, especially in Precision (i.e. a much lower false-positive rate). Recall, however, that we do not expect the expert taggers to generate all-inclusive labels, which prompted the usage of an F_2 metric. The F_2 -score for both methods are remarkably close. Considering the drastic (over an order of magnitude) human time-investment discrepancy between implementing the two methods, the 7-point gain for the SVM is not a particularly worth-while return on investment.

Still, the increased precision is a desirable feature of the SVM. Once the initial vocabulary-based tags have been through the Human-Critique phases described above, an SVM using Word2Vec would potentially pick up on contextual tag patterns not existing explicitly in the raw text. This layer of human-machine hybridization should provide a robust and scalable approach to the continued tagging of issues, especially when the corpus size becomes considerably larger.

B. Potential Applications of Datafication by Tags

In this section, we provide an initial foray into the usage of tagged maintenance data for 1) Diagnostic assistance, and 2) Prognostics and health management (PHM). While there

"Hybrid Datafication of Maintenance Logs from Al-Assisted Human Tags." Paper presented at 2nd Symposium on Data Analytics for Advanced Manufacturing, Boston, MA, United States. December 11, 2017 - December

⁵Mutli-label classification is the task of assigning some set of target labels to an input, meaning that the output "classes" may not be mutually exclusive (as would be the case in multiclass classification). This is necessary for tag prediction, since resources must receive multiple tags.

 $^{^{6}\}mbox{Called Binary Relevance method in the multi-label classification literature [11]$

is not space here for a holistic tag-based approach to either field (nor would any single study necessarily suffice), we hope to demonstrate the effectiveness of tagged maintenance issues at capturing some useful information from otherwise unused natural-language data.

1) Diagnostics: The most straight-forward application of tags is to query on the occurrence rate for certain tag combinations, as proxy for occurrence rate of certain issues. This is especially effective for its simplicity. Using the case-study data, the most common raw-text description was an "Accumulator check request", occurring uniquely 14 times. After automated vocabulary tagging, the issue count for occurrence of {accumulator} \cup {check request} was 73. Additionally, the rate for {hydraulic} \cup {leak}, which was not in the top 20 most common (raw-text) issue descriptions, in fact occurred in 159 separate issues. The high occurrence of hydraulic leaks was not known to the maintenance management previously.

Extending this idea, tag occurrence can be shown over time, visually illustrating issue "hot-spots", as demonstrated in the top of Fig. 6. Note the increase of hydraulic leaks in the summer months, information which could potentially be correlated with specific jobs underway in those periods, that might be causing such behavior.

Another possible use-case concerns solving new maintenance work-orders. Given the set of existing tags, along with some new problem and item tags as input from a maintenance tech, an algorithm might predict the set of most likely solution tags, along with other item tags of significance to this issue. If the additional step of classifying Items as "problem-items" and "solution-items" is taken as part of the Human-Critique phase, it is possible to predict the object involved in solving some input problem. This might be done with a type of "maintenance-tag language", using the classified tags as an n-gram language model, such as in [16]. An visual implementation of this style of diagnostic assistance is demonstrated in Fig. 7.

2) Preliminary Prognostics: Prognostics is primarily concerned with predicting the useful life of some system or component before it ceases to function [18]. This remaining useful life (RUL) is often calculated using empirical observations and/or a model of the process failure rates. The expected time between actual failures is called the failure inter-arrival time (δ_f) , and it is generally modeled as a stochastic process from which failures are drawn. [19]

Note that failures are not-in general-the same as maintenance requests, since a machine able to be repaired has not actually failed or reached 0 RUL. With this in mind, even if we make assumptions that failures are a constant-rate stochastic process, it is often quite difficult to calculate RUL values, let alone when there is a lack of sensor or other preexisting data in a CMMS system. However, with the tagging system proposed above, it is possible to approximate failures as occurring when some set of pre-defined tags appear, such

as {broken} \cup {replace}. A calendar view of the tag cooccurrences over time, with this explicit condition, is shown in the bottom of Fig. 6. Using that data, is now possible to approximate δ_f as drawn from, for example, an exponential distribution $\delta_f \sim \text{Exp}(\lambda)$, where λ is the rate parameter in a constant-rate Poisson Process. The approximation, while certainly rough, can be visually verified as reasonable for the "spindle" and "motor" tags in Fig. 8.

As a final item of interest, it is perhaps useful to consider a taxonomy of maintenance concepts. Determining which tags are most discriminatory for use in prognostics is a non-trivial problem, and a taxonomy might help to organize the tags as inheriting information or even classification from other, more generalized tags. However, generating such a taxonomy or ontology is a labor-intensive design task, and could not easily be generalized across multiple manufacturing domains. One possibility is the use of automated taxonomy-generation algorithms built around the sorting of tag co-occurrences according to their graph-theoretic centrality. One such tool, the Heymann algorithm [20], has shown preliminary success at hierarchically ordering the item tags by generality. This will be discussed at length in further research, but initial results suggest that this automatically generated taxonomy is easy for humans to critique for robustness This may prove a much simpler task than designing a process taxonomy from scratch.

IV. CONCLUSION

The work outlined in this paper is part of a larger framework presented in [3]. The framework is the basis for a knowledge base of manufacturing performance problems, to support diagnosis of performance issues. The use of manufacturing work order data is one method for identifying these problems, and others are being researched as well. The tagging procedures described in this research promise to provide a rich source of data for the knowledge base.

The progress shown thus-far in datafication of maintenance logs addresses one of the primary challenges of constructing such a knowledge base-formulating reusable semantics around data being collected by manufacturers today. By most estimates, a significant portion of today's manufacturing data is not being used-or perhaps only analyzed later, when significant problems occur. A barrier to using the data more proactively is the use of inconsistent terminology and lack of context for the data. Our tagging approach can help address these problems.

In addition, the insights gained through the datafication and analysis processes themselves suggest that meaningful insights may be gained from the data independent from its inclusion in a knowledge base. The goal for this work is to provide methods to improve manufacturing performance and help to address common challenges that industries face. We envision publishing a formal methodology and tool that manufacturers could use to learn from their own data sets.

Brundage, Michael; Hoffman, Michael; Morris, Katherine; Sexton, Thurston. "Hybrid Datafication of Maintenance Logs from AI-Assisted Human Tags." Paper presented at 2nd Symposium on Data Analytics for Advanced Manufacturing, Boston, MA, United States. December 11, 2017 - December

14. 2017.



Figure 6. Illustration of the usefulness of tags in quickly analyzing maintenance data as a time-series. Additionally, an approximation to "failure rates" of parts might be approximated by returning only tags co-occurring with both the "broken" and "replace" tags. This rough approximation may be useful for preliminary RUL models.

MAXIMUM LIKELIHOOD TAG PREDICTION

Load a dataset and type the observed problem tags (if any). Select your probability threshold. The plot view can be auto-scaled at the right of the plot (refresh key).

NIST





Figure 7. A simple app interface demonstrating the usefulness of further categorizing Items into "problem items" and "solution items". Simply using ordered tag co-occurrences previously observed in the database, one might calculate the unsmoothed likelihood of observing some solution tags given some input problem tags, as illustrated above. Each node is weighted by its likelihood given the nodes before it (left-to-right), and the edges indicate how much each node is contributing to the likelihood of its successors. In this example, given an issue like "Hydraulic leak; motor is broken with a loose brush unit", the suggested solution is to replace the output module. Note that only nodes above some tunable probability threshold are displayed. Created using interactive widgets via the Bokeh visualization library [17].



Figure 8. Approximate failure inter-arrival time (δ_f) distributions extracted using vocabulary-based tags. Basic models for spindle or motor failure could perhaps assume a constant-rate Poisson Process, allowing one to predict failures, e.g., for a motor once every 18 days, on average.

Future work will also include reproducing our results with data from a broader variety of organizations and sources, as well as understanding the idiosyncrasies of computation with sensor data. Finally, we plan to explore additional interface technologies that help humans in validating and augmenting data sets produced through automated means of processing. These mechanisms, along with the previously discussed human-critique methodologies, are areas for future research.

REFERENCES

- [1] E. Dumbill, "A revolution that will transform how we live, work, and think: An interview with the authors of big data," Big data, vol. 1, no. 2, pp. 73-77, 2013.
- [2] R. L. Kegg, "One-line machine and process diagnostics," CIRP Annals - Manufacturing Technology, vol. 33, no. 2, pp. 469 - 473, 1984. [Online]. Available: http://www. sciencedirect.com/science/article/pii/S0007850616301688
- [3] M. P. Brundage, B. Kulvantunyou, T. Ademujimi, and B. Rakshith, "Smart manufacturing through a framework for a knowledge-based diagnosis system," in Proceedings of the ASME 2017 International Manufacturing Science and Engineering Conference, MSEC2017. American Society of Mechanical Engineers, 2017.
- [4] K. Doyle, J. Saulman, and B. Cary, "IBM solution approach for enterprise asset management," IBM Software Group, Tech. Rep., 1 2014, document Version 1.2.
- [5] M. Strohmaier, C. Körner, and R. Kern, "Understanding why users tag: A survey of tagging motivation literature and results from an empirical study," Web Semantics: Science, Services and Agents on the World Wide Web, vol. 17, pp. 1-11, 2012.
- [6] T. Vander Wal, "Folksonomy," 2007.
- [7] A. Rajaraman and J. D. Ullman, Mining of Massive Datasets. Cambridge University Press, 2011.

- [8] X. Chen, P. N. Bennett, K. Collins-Thompson, and E. Horvitz, "Pairwise ranking aggregation in a crowdsourced setting," in Proceedings of the sixth ACM international conference on Web search and data mining. ACM, 2013, pp. 193-202.
- [9] J. Bragg, D. S. Weld et al., "Crowdsourcing multi-label classification for taxonomy creation," in First AAAI conference on human computation and crowdsourcing, 2013.
- [10] S. Godbole and S. Sarawagi, "Discriminative methods for multi-labeled classification," Advances in knowledge discovery and data mining, pp. 22-30, 2004.
- [11] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," IEEE transactions on knowledge and data engineering, vol. 26, no. 8, pp. 1819-1837, 2014.
- [12] T. Mikolov, K. Chen, G. Corrado, J. Dean, L. Sutskever, and G. Zweig, "word2vec," 2014.
- [13] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 67, no. 2, pp. 301-320, 2005.
- [14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
- [15] M. Honnibal and M. Johnson, "An improved non-monotonic transition system for dependency parsing," in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: Association for Computational Linguistics, September 2015, pp. 1373-1378. [Online]. Available: https://aclweb.org/anthology/D/ D15/D15-1162
- [16] Y. Lv and C. Zhai, "Positional language models for information retrieval," in Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2009, pp. 299-306.
- [17] Bokeh Development Team, Bokeh: Python library for interactive visualization, 2014. [Online]. Available: http: //www.bokeh.pydata.org
- [18] G. J. Vachtsevanos, F. Lewis, A. Hess, and B. Wu, Intelligent fault diagnosis and prognosis for engineering systems. Wiley Online Library, 2006.
- [19] Y. Xie, K. Tsui, M. Xie, and T. Goh, "Monitoring timebetween-events for health management," in Prognostics and Health Management Conference, 2010. PHM'10. IEEE. 2010, pp. 1-8.
- [20] P. Heymann and H. Garcia-Molina, "Collaborative creation of communal hierarchical taxonomies in social tagging systems," Stanford, Tech. Rep., 2006.

Brundage, Michael; Hoffman, Michael; Morris, Katherine; Sexton, Thurston. "Hybrid Datafication of Maintenance Logs from AI-Assisted Human Tags." Paper presented at 2nd Symposium on Data Analytics for Advanced Manufacturing, Boston, MA, United States. December 11, 2017 - December

PERFORMANCE CHARACTERIZATION OF PROCESS MONITORING SENSORS ON THE NIST ADDITIVE MANUFACTURING METROLOGY TESTBED

B. Lane¹, S. Grantham², H. Yeung¹, C. Zarobila², J. Fox¹

¹Engineering Laboratory, ²Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899

Abstract

Researchers and equipment manufacturers are developing in-situ process monitoring techniques with the goal of qualifying additive manufacturing (AM) parts during a build, thereby accelerating the certification process. Co-axial melt pool monitoring (MPM) is one of the primary in-situ process monitoring methods implemented on laser powder bed fusion (LPBF) machines. A co-axial MPM system is incorporated on the Additive Manufacturing Metrology Testbed (AMMT) at the National Institute of Standards and Technology (NIST); a custom LPBF and thermophysical property research platform where one of many research goals is to advance measurement science of AM process monitoring. This paper presents the methods used to calibrate and characterize the spatial resolution of the melt pool monitoring instrumentation on the AMMT. Results from the measurements are compared to real melt pool images, and analysis is provided comparing the effect on spatial resolution limits on image analysis.

Introduction

Process monitoring for additive manufacturing describes a suite of sensor technologies that may be applied in-situ during an additive manufacturing (AM) build, with the goal of providing a record of the build by correlating the sensor signals with various part qualities or defects. This quality record may then be used as substitute for costly and time-intensive ex-situ part or material qualification. This concept is described with multiple names; e.g., *rapid qualification* [1], *certify-as-you-build* [2], and *in-process quality assurance* [3], and the types of sensor systems and applications used are widely reviewed [4–8]. For laser powder bed fusion (LPBF) process, one of the most promising technologies is co-axial melt pool monitoring (MPM), which is already appearing on commercial LPBF systems.

Co-axial MPM uses an imaging detector (camera) and/or photodetectors aligned co-axially along the optical path of the laser using a beam splitter, such that optical emission from the melt pool are captured stationary within the field of view of the sensors while the laser scans over the powder layer. These signals or images are then processed and parameterized. For example, length, width, area, or intensity may be calculated from melt pool images [9], then mapped to the spatial coordinates within the volume of a part, and correlated to part quality or defects measured ex-situ. Additionally, these in-situ coordinates may be correlated to X-ray computed tomography (XCT) measurements of part porosity [10,11]. Of course, the resolution of this spatial mapping is set by the temporal resolution of the co-axial sensors combined with the scanning speed of the co-located laser spot and imager field of view. Several works have looked at the temporal resolution or sampling rate requirements for MPM systems [10,12,13].

Similarly, there are spatial resolution requirements for evaluation of melt pool images themselves as opposed to the resolution for spatially mapping the melt pool data within the 3D part space. Most AM process monitoring literature using imaging techniques provides only a statement of the measurement system's instantaneous field of view (iFoV), which is the equivalent size of the detector pixel projected onto the object plane (µm per pixel). The ratio of the iFoV to the detector pixel size is equal to the

magnification of the system. The pixel pitch is the distance between centers of adjacent pixels on the detector. Grasso et al. provide a good review of the various iFoV sizes for multiple co-axial (image coordinates move with the laser) and staring (image coordinates are fixed) AM process monitoring systems [5]. However, the iFoV is not the technically correct definition of image resolution, which can be limited by inherent blurring from the optical system. This blurring, and the contributing image degradation stemming from components within the measurement system, can be described by modulation transfer functions (MTFs), which are further described later in this paper. Lott et al. provided one of the few examples of characterization of the optical system performance of a co-axial MPM system [14].

This spatial resolution is ultimately set by 1) the performance of the optics, 2) characteristics of the imaging detector, and 3) the image processing algorithms used to parameterize the melt pool image. The scope of this paper primarily deals with the first two, while providing an example of the third applied to co-axial melt pool images. Measured spatial resolution of the co-axial MPM system is measured and compared against the modeled optimal performance. The goals of this paper are to 1) evaluate performance of the melt pool monitoring system, primarily regarding image resolution, and 2) give example measurements and examine the effect of limiting resolution on melt pool image analysis. Ultimately, for co-axial MPM and process monitoring to be useful in rapid qualification, standard characterization of the system performance, including spatial resolution, is necessary, as is the understanding of its effect on melt pool image processing.

Co-axial Melt Pool Monitoring System Spatial Resolution

Description of the laser injection and process monitoring optical paths on the AMMT, including system modeling and design, was provided in [15]. The system was designed using Zemax¹ modeling software to enable 1x magnification imaging of the melt pool optimized at 850 nm \pm 25 nm waveband, while being constrained to enable a 100 µm full-width half max (FWHM) focused laser spot at 1070 nm wavelength. To avoid chromatic aberrations inherent with an f-theta lens, it utilizes a linear translating z-lens (LTZ) to maintain a planar focus of the laser and imager.

Two imagers are tested in this paper; a high speed, low resolution camera (pixel pitch = $20 \mu m$), and a low speed, high resolution camera (pixel pitch = $1.67 \mu m$). Given the 1:1 magnification design, the iFoV is equal to each detector's pixel pitch (measured to be within 1%). The high speed camera, capable of frame rates in excess of 100 000 frames per second, is utilized to research high speed melt pool dynamics [13]. The low speed camera is a general purpose, gigabit Ethernet (GigE) camera, not suitable for the high temporal bandwidth requirements for MPM. The fine detector pitch of the GigE camera can be used to measure the optical performance of the system, without being limited by detector size.

Knife-Edge Measurement of Modulation Transfer function (MTF)

The modulation transfer function (MTF) provides a more comprehensive definition of spatial resolution for an imaging system. For a measured object with spatially modulated intensity, the MTF is the ratio of the optical system output over the modulated intensity input, at a specific spatial frequency [16,17]. It is essentially the Fourier transform of the impulse response (point spread function, PSF), or a measure of the 'blurriness' of an image, but in frequency space. When the MTF = 0, this means the relative intensities of adjacent bright and dark parts of a measured object are the same, resulting in zero difference or contrast

Fox, Jason; Grantham, Steven; Lane, Brandon; Yeung, Ho; Zarobila, Clarence. "Performance Characterization of Process Monitoring Sensors on the NIST Additive Manufacturing Metrology Testbed." Paper presented at Solid Freeform Fabrication Symposium, Austin, TX, United States. August 7, 2017 - August 9, 2017.

¹ Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

in intensity in the corresponding image. A commonly cited definition of 'spatial resolution' is given as the 10 % of the MTF curve [17], or the spatial frequency at which the contrast is 90 % attenuated.

An optical system MTF can be considered to be composed of multiple components that describe the system, such as the optics (MTF_{optics}), the detector ($MTF_{detector}$), or even aberrations, defocusing, motion blur, etc [17]. By comparing the modeled system MTFs vs. measured, much can be learned about the real system performance vs. optimal design. For the co-axial MPM system analyzed here, we consider the system composed of the optics and the detector:

$$MTF_{system} = MTF_{detector} \cdot MTF_{optics}.$$
 (1)

The MTF of the AMMT co-axial MPM system was determined from Zemax optical modeling software during the system design [15]. The design enables diffraction limited performance, which provides the maximum throughput (or minimum attenuation) physically limited by the wavelength and optics geometry. $MTF_{optics}(f) = 0$ when $f_0=1/\lambda N$, where N is the f-number of the system, λ is the light wavelength, and *f* is the spatial frequency (cycles/mm).

MTFs for a detector with square pixels can be computed as $MTF_{detector}(f) = |\operatorname{sinc}(w f)|$, where w is the pixel pitch (mm/pixel). This means the 0% cutoff ($MTF_{detector}(f) = 0$) occurs at $f_0 = 1/w$. It should be noted that this calculation assumes square pixels, and provides the MTF along the primary detector axes, whereas the MTF along diagonals or for non-square detectors are more complicated [16]. However, the simpler approximation suffices for the analysis made in this paper.

The MTF_{system} can be measured by performing a knife edge (KE) measurement, as shown in Figure 1 for the co-axial MPM system and two detector types. A KE measurement is made by imaging a well-focused straight edge that splits uniform light and dark portions of an image. In Figure 1, this was accomplished by setting a razor blade on top of a fiber illuminator. The co-axial MPM imager was focused on the KE, while the fiber end was at a defocused distance below the KE, and covered with Teflon tape to act as a diffuser.



Figure 1: Knife-edge image results for both detectors, showing the finer detail in the $1.67 \mu m$ detector measurement. Image axes are in pixels, and are scaled to approximately the same geometric dimensions.

To compute the measured MTFs from the KE images, a custom code was written that performs the methodology defined in ISO 12333 [18,19]. In short, multiple line profiles orthogonal to the KE (x vs. digital level) are sampled, resulting in an edge spread function (ESF). The MTF is then computed as the Fourier transform of the spatial derivative (d/dx) of the ESF. Further description of this algorithm is given in [20]. Figure 2 provides the results of the measured MTF_{system} utilizing the 20 µm and 1.67 µm detectors, and comparison to the MTF_{optics}, MTF_{detector}, and MTF_{system} calculated using the aforementioned methods.



Figure 2: Comparison of modeled vs. measured MTF results for the 20 µm and 1.67 µm coaxial monitoring systems.

Multiple observations can be made about the optical system by comparing the curves in Figure 2. First, the measured MTF_{system} curves are below the modeled results, which is physically consistent. The calculated and measured MTF_{system} for the 20 µm detector cuts off at \approx 50 cycle/mm, which corresponds to the measured curve. The 1.67 µm $MTF_{detector}$ is zero at approximately 598 cycle/mm (green curve), the 20 µm $MTF_{detector}$ cutoff is at \approx 50 cycle/mm, and diffraction limited (Zemax) optics cutoff is \approx 80 cycle/mm. Therefore the 20 µm detector limits the maximum resolvable spatial frequency for that detector, while the optics limit the maximum resolvable frequency for the 1.67 µm detector.

Since the calculated 1.67 μ m MTF_{detector} shows minimal attenuation, it essentially does not limit the optical system, so that calculated 1.67 μ m MTF_{system} \approx MTF_{optics} (maximum physical resolution), and the measured 1.67 μ m MTF_{system} describes the real optical performance of the system without detector size being a limitation. Since the measured MTF_{system} for the 1.67 μ m detector is below this optimal curve, this indicates there are other contributing factors such as obscurations, aberrations, or misalignment that need to be investigated. Similarly, the measured MTF_{system} for the 20 μ m detector is below the calculated curve. Overall, the measured resolution, defined as 10% of the MTF, is 33.1 cycle/mm for the 20 μ m detector, and 44.2 cycle/mm for the 1.67 μ m detector.

While MTFs provide a more appropriate and encompassing definition of spatial resolution as opposed to the iFoV, they do not fully describe the performance of an optical system for its intended purpose. For example, a 10 % MTF resolution of 44.18 cycle/mm does not mean that only $(44.18 \text{ cycle/mm})^{-1} = 22.6 \,\mu\text{m}$ or larger objects can be measured. To more closely analyze the real effect of spatial resolution on co-axial MPM measurements, we can test objects that more nearly resemble melt pool images.

Aperture Measurements as Surrogate Melt Pool

Fox et al. measured melt pool track widths ex-situ made in a prototype version of the AMMT, and compared these measurements to various in-situ melt pool image widths calculated after thresholding and binarizing the images [13]. Ex-situ measured track widths were relatively consistent at 118 μ m ± 4.3 μ m (± one standard deviation). However, melt pool widths measured in-situ with the MPM camera depend on the grayscale level chosen. In Fox et al., this could vary by the diagonal distance between two pixels, or ± 28 μ m. However, the additional effect of limited optical resolution could not be investigated solely with in-situ melt pool images. To imitate a melt pool image under controlled conditions and known geometry, precision apertures were placed above the fiber illuminator, and were subsequently imaged with the co-axial MPM and both detector sizes. The apertures ensure a uniform illumination, whereas real melt pool images are typically saturated at the hottest part of the melt pool, and have very high intensity gradients stemming from the high temperature gradients on the melt pool surface. The uniform illumination in the aperture is set below the saturation level of the imager, which ensures any signal attenuation or effect on size measurements from the images can be measured.

Apertures at 100 μ m, 200 μ m, 300 μ m, and 500 μ m were tested, with the manufacturer's stated diameter tolerance of ±5 % for each. The centroid of the aperture images is calculated, and X and Y profiles through the centroid are shown in Figure 3 for the 1.67 μ m detector, and Figure 4 for the 20 μ m detector. There is visible ghosting in the aperture images, which is an unfocused duplicate of the aperture slightly offset from the main image. This may be due to imperfect anti-reflective coating of one of the dozen or so mirrors, lenses, etc. in the optical path, and likely contributes to the sub-optimal MTF_{system} measured in the previous section. However, this ghosting cannot be easily discerned from a KE image or MTF, which demonstrates the value in using a variety of characterization tests.



Figure 3: Aperture measurement results for 1.67 µm detector.



Figure 4: Aperture measurement results for 20 µm detector.

One can see attenuation of maximum signal value for the 100 μ m aperture for both the 20 μ m and 1.67 μ m detectors. A 100 μ m aperture image may be considered ½ a cycle, or $w/2 = 100 \ \mu$ m, which corresponds to $f \approx 5$ cycle per mm. At f = 5 cycle per mm, the MTF_{detector} = 89 % for the 20 μ m detector and 92.3 % for the 1.67 μ m detector. The measured attenuation is roughly 80 % for the 100 μ m aperture measured by both detector sizes. It can be expected that smaller aperture sizes would result in further attenuation.

To further study the effect on a real melt pool measurement through image processing, geometric parameters from the melt pool images are calculated using the image processing toolbox in Matlab:

- 1.) Convert to 8-bit grayscale.
- 2.) Resize using sub-pixel interpolation (Matlab function 'imresize').
- 3.) Convert to binary images at three user-selected levels. The central binary image is used for further image processing and calculation.
- 4.) Calculate the centroid, major, and minor axes of the binarized melt pool image, and relative angle of these axes.
- 5.) Calculate the length and width of the melt pool along the major and minor axes, respectively.

It should be emphasized that the resizing function in step 2 interpolates between pixels, which results in sub-pixel resolution geometric measurements. Two example processed images of the 100 μ m aperture are shown in Figure 5. The central, target gray level is set such that the 100 μ m aperture image results in an accurate measure of the aperture width based on the detector iFoV as shown in Table 1. Two other gray levels contours are plotted at approximately ½ and 2x the central contour to show the shape of higher and lower levels. Width, length, and major axis angle are calculated from the image analysis for the four aperture sizes and two detectors.



Figure 5: Example results from melt pool image analysis performed on back-illuminated 100 µm precision aperture. Note that image scales are not equal.

Table 1: Results from melt pool image analysis performed on precision apertures. Manufacturing
tolerance on the apertures is ±5%.

	1.67 μm Pixel Camera			20 µ	m Pixel C	amera
Aperture	Width	Length	Angle	Width	Length	Angle
μm	μm	μm	degrees	μm	μm	degrees
100	100	115	78	100	105	-26
200	204	224	78	202	208	-20
300	300	322	75	294	301	-27
500	494	527	71	483	494	-33

The ghosting that was apparent in the raw aperture images is exemplified in in the lower grey level contour, creating an asymmetric shape, which causes the unequal length and width measurements. The angular difference is due to the fact the imagers are held in different orientations. In a real melt pool measurement, asymmetry in the optical performance may bias angle or MP direction calculation, particularly if lower grey levels, or lower temperature ranges in the melt pool, are the target measurement.

Analysis of Real Melt Pool Images

Due to exceedingly high temperature gradients in and around the melt pool, the limited dynamic range of an imaging sensor results in saturated pixels at the center of the melt pool image, and pixels at the detector noise floor surrounding it. The image gradient, or contrast, in between is also very high, which corresponds to high spatial frequencies on the MTF curve. In this way real melt pool images are similar to the previously analyzed aperture images, although the center of a melt pool very likely is not uniform intensity nor were the aperture images set to saturate the camera. While some effects of the real spatial resolution can be observed with the aperture images, analysis of real melt pool images elucidate further complexities.

In the following example, a single line on a bare nickel alloy 625 plate is scanned at 3.4 m/s and 400 W laser power. A second scan was performed at the same parameters, but with an approximately 100 μ m layer of alloy 625 powder spread on top of the nickel alloy plate. Laser spot is approximately 100 μ m full-width half max (FWHM). Frame rate of the co-axial MPM camera with 20 μ m detector size is set to 30 000 frames per second, integration time $t_{int} = 33 \,\mu$ s, and filtered to 850 nm \pm 20 nm wavelengths. Figure 6 shows the raw melt pool images, process images, and processing results. Only one frame taken from the bare metal surface scan is presented since there was minimal variation between frames.

				Calc	Calculated Results		
		Raw Image	Processed Image	Width (µm)	Length (µm)	Angle (deg)	
Bare meta surface	(1)	٠		197	216	18	
Single Powde	e (1) r	110		252	658	-10	
Layer (2	r (2)	ALC:		278	727	-12	
	(3)	14.1		274	571	-11	
	(4)			304	638	-8	
	(5)		-	305	435	3	

Figure 6: Example MPM images and image processing results for the 20 µm imager. Top row is a single melt pool image when scanning bare metal, subsequent rows are high speed video frames while scanning a single-layer of powder.

The inclusion of powder adds complexity to the image analysis. Particle images still connected to main melt pool body will give erroneous size or angle measures, which is evident in Figure 6. Particles counting or frequency measurement may warrant additional image processing, as there is evidence to suggest spatter quantification may be correlated to AM track formation quality or process stability [21–23]. Since particles are at a lower temperature than the center of the MP; reducing integration time would effectively 'remove' particles from the MP images, but limit the measurement to only the highest apparent temperature regions of the MP, which may not be as sensitive the overall MP geometry changes.

Other potential factors limiting imager performance:

Blooming occurs when a region of pixels saturates due to excessive light flux, and the excess charge in the saturated pixels affects the signal output of adjacent pixels. Since it is nearly inevitable that melt pool images will have some portion saturated, blooming may have a significant effect on resultant apparent melt pool size measured via image processing. The light source used to create the aperture images in Figure 3 and Figure 4 was not bright enough. To evaluate the effect of blooming, a source intensity approaching that of a real melt pool is necessary². Anti-blooming technologies exist; however, this changes the linear behavior of the detector (signal being proportional to incident light flux), which may introduce additional complexities in processing the images. In addition, detector linearity, or at least consistent flux to signal mapping, is essential for a radiant temperature calibration.

² Using Planck's law integrated over the 850 nm \pm 20 nm waveband and evaluated at temperatures > 2500 K, and assuming the f-number of the system of 14.7 with opening aperture of 20 mm, this source flux would need to be on the order of 1 to 100 W/mm².

Motion blur is another resolution limiting factor, which is not widely discussed in co-axial MPM literature. The blur distance may be estimated as $v t_{int}$, with v being the linear motion velocity relative to the field of view frame of reference, and t_{int} being the camera integration time. Since the melt-pool image is stationary within the field of view for co-axial MPM systems, only fluctuations of the melt pool boundary or spatter particle motion may be of concern. Still, the exceedingly high radiant flux stemming from a laser-induced melt pool means relatively short t_{int} is required (e.g., $t_{int} = 33 \ \mu s$ for the melt pool images in Figure 6). For melt pool geometry measurements, this may not be a concern, however, for spatter analysis it may.

Discussion and Future Work

Many AM process monitoring define 'resolution' as iFoV, whereas true image resolution should be defined in more comprehensive terms such as the MTF. Measuring MTF, and comparing against calculated design, can illuminate what are the limiting factors in the optical system. However, MTF in and of itself does not fully define the performance of a co-axial MPM system; the final image processing, which may achieve sub-pixel resolution, will determine the true precision. By modeling then measuring the MTF, non-optimized performance of the imaging system was discovered. Therefore, future work will entail sequential examination of each component in the optical system to determine the culprit.

Based on measurement of the apertures, and the sub-pixel averaging nature of the melt-pool image processing, both a 1.67 µm/pixel and 20 µm/pixel imagers performed approximately the same in calculating aperture diameter, despite one detector being over 10x in size and limiting the MTF. This is mostly due to the interpolating nature of the MP image processing. However, the aperture image analysis demonstrated how intensity values may still be attenuated, despite a less sensitive effect on geometric boundary measurement. This could have significant consequence if parameters such as melt pool intensity or average intensity levels, or calibrated temperature values, are the measurand output from image analysis. Beyond MTF, spatial resolution, and choice in image processing algorithm, further physical complexities in MPM of powder surfaces stem from the ejected particles or spatter. This can create errors in simple length, width, area, etc. parameterizations of the melt pool image, and may warrant additional image processing to quantify the particles, or remove their influence.

Ultimately, co-axial MPM system performance evaluation depends on what objective measurand is to be extracted from the images. A suite of spatial, temporal, and intensity characterization tests, and a thorough understanding of their consequence on the melt pool image analysis is necessary to enable this process monitoring tool to become the key technique for in-situ part qualification.

References

- A.D. Peralta, M. Enright, M. Megahed, J. Gong, M. Roybal, J. Craig, Towards rapid qualification of [1] powder-bed laser additively manufactured parts, Integrating Mater. Manuf. Innov. (2016) 1–23. doi:10.1186/s40192-016-0052-5.
- J. Mazumder, Design for Metallic Additive Manufacturing Machine with Capability for "Certify as [2] You Build," Procedia CIRP. 36 (2015) 187–192. doi:10.1016/j.procir.2015.01.009.
- Sigma Labs, In-Process Quality Assurance[™] (IPQA®) Solutions | Sigma Labs, (2017). [3] https://www.sigmalabsinc.com/node/5 (accessed July 27, 2017).
- S.K. Everton, M. Hirsch, P. Stravroulakis, R.K. Leach, A.T. C lare, Review of in-situ process [4] monitoring and in-situ metrology for metal additive manufacturing, Mater. Des. (2016). doi:10.1016/j.matdes.2016.01.099.
- [5] M. Grasso, B.M. Colosimo, Process defects and in situ monitoring methods in metal powder bed fusion: a review, Meas. Sci. Technol. 28 (2017) 044005. doi:10.1088/1361-6501/aa5c4f.

Fox, Jason; Grantham, Steven; Lane, Brandon; Yeung, Ho; Zarobila, Clarence. "Performance Characterization of Process Monitoring Sensors on the NIST Additive Manufacturing Metrology Testbed." Paper presented at Solid Freeform Fabrication Symposium, Austin, TX, United States. August 7, 2017 - August 9, 2017.

- [6] E. Reutzel, A. Nassar, A survey of sensing and control systems for machine and process monitoring of directed-energy, metal-based additive manufacturing, Rapid Prototyp. J. 21 (2015) 159-167. doi:10.1108/RPJ-12-2014-0177.
- [7] T.G. Spears, S.A. Gold, In-process sensing in selective laser melting (SLM) additive manufacturing, Integrating Mater. Manuf. Innov. 5 (2016). doi:10.1186/s40192-016-0045-4.
- [8] G. Tapia, A. Elwany, A Review on Process Monitoring and Control in Metal-Based Additive Manufacturing, J. Manuf. Sci. Eng. 136 (2014) 060801–060801. doi:10.1115/1.4028540.
- [9] T. Craeghs, S. Clijsters, J.-P. Kruth, F. Bechmann, M.-C. Ebert, Detection of process failures in Layerwise Laser Melting with optical process monitoring, Phys. Procedia. 39 (2012) 753–759.
- [10] S. Clijsters, T. Craeghs, S. Buls, K. Kempen, J.-P. Kruth, In situ quality control of the selective laser melting process using a high-speed, real-time melt pool monitoring system, Int. J. Adv. Manuf. Technol. 75 (2014) 1089-1101. doi:10.1007/s00170-014-6214-8.
- [11] Concept Laser, Sintavia uses QM Meltpool to Ensure Part Quality in Metal AM, Concept Laser. (2017). http://www.conceptlaserinc.com/sintavia-uses-qm-meltpool-ensure-part-quality-metal/ (accessed June 19, 2017).
- [12] S. Berumen, F. Bechmann, S. Lindner, J.-P. Kruth, T. Craeghs, Quality control of laser- and powder bed-based Additive Manufacturing (AM) technologies, Phys. Procedia. 5, Part B (2010) 617-622. doi:10.1016/j.phpro.2010.08.089.
- [13] J.C. Fox, B.M. Lane, H. Yeung, Measurement of process dynamics through coaxially aligned high speed near-infrared imaging in laser powder bed fusion additive manufacturing, in: Proc SPIE 10214 Thermosense Therm. Infrared Appl. XXXIX, Anaheim, CA, 2017: pp. 1021407-1021407-17. doi:10.1117/12.2263863.
- [14] P. Lott, H. Schleifenbaum, W. Meiners, K. Wissenbach, C. Hinke, J. Bültmann, Design of an Optical System for the In-Situ Process Monitoring of Selective Laser Melting (SLM), Phys. Procedia. 12 (2011) 683-690. doi:10.1016/j.phpro.2011.03.085.
- [15] B. Lane, S. Mekhontsev, S. Grantham, M. Vlasea, J. Whiting, H. Yeung, J. Fox, C. Zarobila, J. Neira, M. McGlauflin, L. Hanssen, S. Moylan, M.A. Donmez, J. Rice, Design, developments, and results from the nist additive manufacturing metrology testbed (ammt), in: Proc. 26th Annu. Int. Solid Free. Fabr. Symp., Austin, TX, 2016: pp. 1145–1160.
- [16] K. Fliegel, Modeling and measurement of image sensor characteristics, RADIOENGINEERING-PRAGUE-. 13 (2004) 27-34.
- [17] G.C. Holst, Holst's Practical Guide to Electro-optical Systems, JCD Publishing, Winter Park, FL, 2003.
- [18] ISO 12233:2014, Photography Electronic still-picture cameras Resolution measurements, ISO, Geneva, Switzerland, n.d.
- [19] M. Estribeau, P. Magnan, Fast MTF measurement of CMOS imagers using ISO 12333 slanted-edge methodology, in: Proc. SPIE, St. Etienne, France, 2003: pp. 243-252.
- [20] B. Lane, E. Whitenton, Calibration and measurement procedures for a high magnification thermal camera, NIST Internal Report 8089, National Institute of Standards and Technology, Gaithersburg, MD. 2015.
- [21] L.E. Criales, Y.M. Arisoy, B. Lane, S. Moylan, A. Donmez, T. Özel, Laser powder bed fusion of nickel alloy 625: Experimental investigations of effects of process parameters on melt pool size and shape with spatter analysis, Int. J. Mach. Tools Manuf. 121 (2017) 22 - 36.doi:10.1016/j.ijmachtools.2017.03.004.
- [22] S.A. Khairallah, A.T. Anderson, A. Rubenchik, W.E. King, Laser powder-bed fusion additive manufacturing: Physics of complex melt flow and formation mechanisms of pores, spatter, and denudation zones, Acta Mater. 108 (2016) 36-45. doi:10.1016/j.actamat.2016.02.014.
- [23] G. Repossini, V. Laguzza, M. Grasso, B.M. Colosimo, On the use of spatter signature for in-situ of Laser Powder Bed Fusion, Addit. Manuf. 16 monitoring (2017)35-48. doi:10.1016/j.addma.2017.05.004.

Fox, Jason; Grantham, Steven; Lane, Brandon; Yeung, Ho; Zarobila, Clarence. "Performance Characterization of Process Monitoring Sensors on the NIST Additive Manufacturing Metrology Testbed." Paper presented at Solid Freeform Fabrication Symposium, Austin, TX, United States. August 7, 2017 - August 9, 2017.

An Investigation of Different Strategies for Solving Coupled Thermal Airflows by Multi- Zone Network Method

> ¹Ali Katal ¹Liangzhu Wang ²W. Stuart Dols ²Brian J. Polidoro

¹Centre for Zero Energy Building Studies, Department of Building, Civil and Environmental Engineering Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8, Canada ²Engineering Laboratory, National Institute of Standards and Technology 100 Bureau Drive Gaithersburg, MD 20899

> Content submitted to and published by: **COBEE 2018** 4th Conference on Building Energy and Environment

> > U.S. Department of Commerce Wilbur Ross, Secretary of Commerce



National Institute of Standards and Technology Walter Copan, Director



DISCLAIMERS

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

Any link(s) to website(s) in this document have been provided because they may have information of interest to our readers. NIST does not necessarily endorse the views expressed or the facts presented on these sites. Further, NIST does not endorse any commercial products that may be advertised or available on these sites.

AN INVESTIGATION OF DIFFERENT STRATEGIES FOR SOLVING COUPLED THERMAL AIRFLOWS BY MULTI-ZONE NETWORK **METHOD**

A. Katal¹, L. Wang¹, W. Stuart Dols², B.J. Polidoro²

¹Centre for Zero Energy Building Studies, Department of Building, Civil and Environmental Engineering Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8, Canada

²Indoor Air Quality and Ventilation Group. Engineering Lab. US National Institute of Standards and Technology. 100 Bureau Drive, Stop 8633, Gaithersburg, MD, USA.

SUMMARY

Natural ventilation (NV) is a method of allowing outdoor air into indoor living spaces by natural means without the use of air conditioning units and other types of mechanically-driven devices. Therefore, NV of buildings is a cooling strategy that requires less energy use and has received more attention in recent decades. Multi-zone airflow network models provide a relatively simple method for simulation of NV in buildings. In NV problems airflow and energy are highly coupled and building engineers and architects need simulation programs that are capable of simultaneous simulation of building airflows and thermal performance. Three different strategies are available for coupling thermal and airflow models, fullysimultaneous, semi-simultaneous and segregated strategies. All three methods are demonstrated for different steady state problems including an analytical case, a three-zone case and a light well case. It is shown that the fully and semisimultaneous methods can present non-convergent solutions with fluctuations due to the ill-conditioned matrix equations. While, for the segregated solver, matrix equations are wellconditioned and this solver provides convergent solutions in all cases tested. Study of parametric equations and corresponding curves of semi- and fully-simultaneous methods show that these curves are nearly parallel in all cases, which correspond to ill-conditioned systems of equations.

INTRODUCTION

NV of buildings has received more attention in recent decades because it is a cooling strategy that requires no energy use. However, since it is generally difficult for NV to maintain indoor comfort under all conditions, it is frequently supplemented with mechanical cooling or fan operation in what has now become known as "hybrid ventilation." It is important for building designers to accurately predict the performance of natural and hybrid ventilation. The simulation of these systems needs to consider the combined impact of thermal and airflow transport. Multi-zone methods can calculate the airflow and indoor air temperature for the whole building, in which a building is treated as a collection of nodes representing zones, i.e., rooms, plenums, etc., with inter-nodal connections representing the flow paths associated with cracks, doors and the like. Multi-zone models have advantages over more detailed models, e.g., computational fluid dynamics, with its ease of use, straightforward application and fast convergence. Several programs based on the multi-zone airflow network model are available to predict building airflows, such as COMIS (Pelletret and Keilholz 1997) and CONTAM (Dols and Polidoro 2015). In these programs node temperatures are

defined by the user and a set of mass balance equations is solved to predict node pressures and resultant inter-node airflows. Requiring temperatures as inputs may be reasonable for mechanically conditioned buildings, assuming thermostat set-points are maintained. However, in the case of naturally ventilated buildings, requiring indoor air temperature as inputs is less reasonable due to the strong interdependence between temperature and airflow. On the other hand, building energy simulation tools such as EnergyPlus, which are capable of performing whole-building heat transfer calculations, require building airflows as input values (Crawley et al. 2001). Therefore, current multi-zone programs cannot directly address these highly coupled thermal airflow problems. In recent years, efforts have been pursued to integrate multizone airflow network models with whole-building energy simulation tools. Three different strategies have been used for coupling the airflow and energy balance equations: fullysimultaneous, semi-simultaneous and segregated solvers (Wang et al. 2012). For a problem having N number of zones, the fully-simultaneous method solves a 2N by 2N matrix equation for air pressure and temperature of all zones. The semi-simultaneous method solves one 2 by 2 matrix equation for pressure and temperature of each zone, and the segregated method solves one N by N matrix equation for pressure of all zones and another N by N matrix equation for the temperature of all zones. Compared to previous work, this study investigates these different strategies from a mathematical point of view to compare the performance of these different methods in solving steady-state problems.

GOVERNING EQUATIONS

In multi-zone network models, air density and temperature are assumed to be uniform within each zone. These network models solve two sets of governing conservation equations mass and energy - for each zone to predict the building temperatures and airflow rates for pre-defined boundary conditions, e.g., outdoor temperature and wind conditions. This section presents the governing conservation equations.

Mass balance equation

The steady-state mass balance equation applied to each zone is

$$\sum_{j} F_{ji} + M s_i = 0 \tag{1}$$

Where F_{ji} is airflow rate [kg/s] between zones j and i; a positive value indicates flow from *i* to *i* and a negative value

indicates flow from i to j. Ms_i [kg/s] defines sources and sinks that could add or remove air to or from the zone.

Energy balance equation

The steady-state energy balance equation for zone *i* is

$$\sum_{j:F_{ji}>0} F_{ji}C_{pa}T_j - \sum_{j:F_{ij}>0} F_{ij}C_{pa}T_i = \sum_k h_k A_k(T_i - T_{mk}) - S_i$$
(2)

where T_i is air temperature of the zone j [K], S_i is heat gain in the zone i[W]. T_{mk} is the surface temperature of the thermal mass (ceiling, floor, or wall) that is connected to the zone i [K]. h_k is convective heat transfer coefficient between the air and the thermal mass $[W/m^2, K]$, A_k is the corresponding surface area of the thermal mass $[m^2]$, and C_{pa} is the specific heat of air [J/kg.K].

Inter-nodal connection equations

The airflow through connections are driven by the pressure difference between connected zones. For the pressure driven flow, the airflow rate from zone *i* to zone *j*, F_{ij} , is a function of the pressure difference across the opening, ΔP_{ij} :

$$F_{ij} = C \left(\Delta P_{ij} \right)^n \tag{3}$$

where F_{ij} is the mass flow rate through the connection [kg/s], C is the flow coefficient, and n is the pressure exponent. The pressure difference across connection ij is calculated by

$$\Delta P_{ij} = (P_i - \rho_i g H_i) - (P_j - \rho_j g H_j) + \frac{\rho_i + \rho_j}{2} g(z_i + H_i - z_j - H_j) + P_{ij,w}$$
(4)

where H_i is the height of airflow opening relative to the floor of the zone $i [m], z_i$ is the floor elevation of the zone $i [m], P_{ii,w}$ is the wind pressure at opening ij [Pa], ρ_i is the density of zone *i*, and *g* is the gravity acceleration $[m/s^2]$.

NUMERICAL SOLUTION PROCEDURE

Three different strategies can be implemented for coupling the airflow and thermal modules in a multi-zone simulation program. The first one is the "fully-simultaneous" coupling strategy. In this strategy, the mass and energy balance equations for all zones are solved simultaneously. The second strategy is the "semi-simultaneous" coupling strategy in which the mass and energy balance equations for one zone are solved simultaneously in an internal loop and this procedure is repeated for all zones in sequence. This sequence of steps is repeated in an external loop until overall convergence is achieved. The third is the "segregated" coupling strategy, in which the airflow and thermal modules are run in a sequence. In other words, the segregate strategy solves the mass balance equations for all zones, then the airflows are provided to the energy balance equations and temperatures are calculated for all zones. Finally, the updated temperatures of all zones are substituted back into the mass balance equation and this procedure is repeated until overall convergence of the problem is attained. The matrix formulations of these three solvers are provided in the following sections.

Fully-simultaneous method

As mentioned before, in this method the continuity and energy equations for all zones are solved simultaneously

$$\begin{cases} fm_1(P_1, T_1, P_2, T_2, \dots, P_n, T_n) = 0 \\ fe_1(P_1, T_1, P_2, T_2, \dots, P_n, T_n) = 0 \\ fm_2(P_1, T_1, P_2, T_2, \dots, P_n, T_n) = 0 \\ fe_2(P_1, T_1, P_2, T_2, \dots, P_n, T_n) = 0 \\ \vdots \\ fm_n(P_1, T_1, P_2, T_2, \dots, P_n, T_n) = 0 \\ fe_n(P_1, T_1, P_2, T_2, \dots, P_n, T_n) = 0 \end{cases}$$
(5)

where n is the number of zones. The system of equations 5 can be combined into a matrix formation

$$f(X) = 0 \tag{6}$$

where

$$\boldsymbol{f} = (fm_1, fe_1, fm_2, fe_2, \dots, fm_n, fe_n)^T,$$
(7)

$$\boldsymbol{X} = (P_1, T_1, P_2, T_2, \dots, P_n, T_n)^T$$
(8)

Semi-simultaneous method

In the semi-simultaneous method, the continuity and energy equations of each zone are solved separately. The corresponding system of equations of zone *i* is as follows

$$\begin{cases} fm_i(P_i, T_i) = 0\\ fe_i(P_i, T_i) = 0 \end{cases}$$
(9)

Segregated method

System of equations for solving mass balance and energy equations using the segregated method are as follows

$$\begin{cases} fm_1(P_1, P_2, \dots, P_n) = 0\\ fm_2(P_1, P_2, \dots, P_n) = 0\\ \vdots\\ fm_n(P_1, P_2, \dots, P_n) = 0 \end{cases}$$
(10)

$$\begin{cases} fe_1(T_1, T_2, \dots, T_n) = 0\\ fe_2(T_1, T_2, \dots, T_n) = 0\\ \vdots\\ fe_n(T_1, T_2, \dots, T_n) = 0 \end{cases}$$
(11)

Applying the Newton-Raphson method to non-linear systems of equations 5, 9, 10, and 11 yields linear systems of Jacobian-based equations, which will be solved by a linear solver. This process is considered as the internal loop.

$$J(X^{(n)})(\Delta X^{(n)}) = f(X^{(n)})$$
(12)

After equation 12 is solved, the zones state variables can be corrected by equation 13

$$\boldsymbol{X}_i^{n+1} = \boldsymbol{X}_i^n - \Delta \boldsymbol{X}_i^n \tag{13}$$

 L_{∞} norm of error is used to measure the convergence of nonlinear system of equations, which is defined as follows

$$Error = \max_{i} \left| f_i^{n+1} \right| \tag{14}$$

(15)

Matrix condition number

Condition number of square non-singular matrix A defined by

$$ond(A) = ||A^{-1}|| \times ||A||$$

For matrices with a large condition number, solving an associated linear system might pose numerical difficulties. For a matrix A with a large condition number the problem Ax = b is referred to as ill-conditioned. A large condition number indicates a nearly singular matrix, whereas a matrix with a condition number close to 1 is far from being singular and indicative of a well-posed and solvable system of equations.

RESULTS

In the current study, we investigated several steady state problems and found that the segregated solver reaches converge for all the cases studied, but the semi-simultaneous and fully-simultaneous solvers show convergence problems in most of the cases. The convergence problem is found due to the large condition number of the A matrix which means that the non-linear and linearized system of equations in these two solvers are ill-conditioned. In the following sections, three selected cases are discussed. First, a case with an analytical solution is solved using all methods to verify the developed programs. Then, two other cases including a three-zone horizontal case and a light well are simulated using all solvers.

Analytical test case

In order to verify the coupling strategies, a three-zone steady case, which has an analytical solution from (Dols et al. 2016), was developed and implemented. The case illustrated in Figure 1 is a three-zone model having one zone cooled to a constant temperature below that of the outdoor temperature. There is an analytical solution for this problem by solving the heat balance equations for the three zones, A, B, and C, which can be solved for the three unknowns T_A , T_B , and \dot{Q}_C . Solving the equations analytically yields values for T_A, T_B and \dot{Q}_C of 30.31 °C, 23.95 °C, and 1600 W, respectively.



Figure 1. Plan view of analytical case (Dols et al. 2016)

The resultant simulation values of different strategies and simulation values of Dols et al. are exactly the same as those of the analytical values.

3-zone horizontal test case

The second test case consists of 3-zones arranged horizontally (Figure 2). Zone 3 is at a constant temperature 30 °C higher than the outdoor temperature. Zones are connected to neighbouring zones and/or ambient by airflow paths located at various elevations. Implementation of the semi-simultaneous method shows fluctuations in solving the system of equations of zone 1, which occurs after 28 iterations of the external loop. Figure 3 shows fluctuations of the semisimultaneous internal loop. The condition number of the coefficient matrix is shown in Figure 4. As can be seen, the coefficient matrix has a very large condition number which means that the linearized, and therefore the corresponding non-linear system of equations, are ill-conditioned causing convergence problems. The reason for the large condition number in the semi-simultaneous method will be discussed in the next section.







Figure 3. Fluctuations of semi-simultaneous method, (a) Error history, (b) airflow rate through path af 2.

Dols, William; Katal, Ali; Polidoro, Brian; Wang, Liangzhu. "An Investigation of Different Strategies for Solving Coupled Thermal Airflows by Multi-Zone Network Method." Paper presented at COBEE 2018 Conference, Melbourne, Australia. February 5, 2018 - February 9, 2018.



Figure 4. Condition number of matrix equation in semisimultaneous method

The fully-simultaneous method shows a similar fluctuation problem (Figure 5). The coefficient matrix of the linearized system of equations also has a very large condition number for all iterations (Figure 6). Therefore, the non-linear system of equation is ill-conditioned, so the fully-simultaneous method performs poorly and cannot reach convergence.







Figure 6. Condition number of matrix equation in fullysimultaneous method

Unlike the two other methods, the segregated solver shows a convergent solution for this problem as shown in Figure 7. The condition number for this case using this solver is always smaller than 10, which means that the system of equation is well-conditioned and solvable.



Figure 7. Convergence history of segregated solver, (a) Error history, (b) airflow rate through path af 2.

Light well natural ventilation case

The third case is a light well with airflows driven by both wind force and thermal buoyancy. Wind tunnel data are available for the validation of the simulation results (Kotani et al. 2003). The schematic of the light well is shown in Figure 8, which corresponds to a 1:250 scaled model of a 41-storey high-rise apartment building. The parameters of the case are listed in Table 2. An orifice airflow equation is used for all the openings. Walls are considered to be adiabatic because heat transfer through them was neglected in the experiment.



Figure 8. Schematic of the light well (Kotani et al. 2003)

Table 2. Simulation parameters used in the light well problem

Heat generation rate	40		
Wind velocity (m/s)	0		
Outside Temperatur	12		
Number of the zone	8		
Height of the zones (m)		0.0585	
	Top opening	2.187*	
Discharge coefficient	Bottom opening	0.855*	
	Inter-zonal openings	1*	
* adopted from the original study (source)			

Simulation of the light well problem using the semisimultaneous and fully-simultaneous methods show the same fluctuations in the solution, which is due to the large condition number of the coefficient matrices. Unlike the semisimultaneous and fully-simultaneous solvers, the segregated solver provides a convergent result for this case. The condition number of the segregated system for this case is always smaller than 10 indicating a well-conditioned problem. The simulated airflow rate through the light well obtained by the segregated solver is compared to measured data and simulation studies of Kotani et al. in Figure 9. Temperature distribution in the light well is compared to measured data in Figure 10. The overall trend of the simulation results agrees reasonably well with the measured values.



Figure 9. Airflow rate through the light well



Figure 10. Temperature change with height inside the light well

Condition number investigation

For the light well case, the condition numbers of the semisimultaneous and fully-simultaneous methods are very large which means that the system of equations is ill-conditioned and in most of the problems we observe fluctuations in the solution. While, in the segregated solver the condition number of matrix equations is very small (less than 10). Therefore, matrix equations of the segregated solver are wellconditioned, and this solver shows convergent results in all simulated cases. In this section the reason for ill-conditioned and well-conditioned systems of equations is explained by drawing the corresponding curves of the equations.

Figure 11 shows a general case for investigation of condition numbers for the semi-simultaneous method. This is a single zone surrounded by 4 zones, and is meant to be a typical zone in any problem, for example the three zone test case or light well problem. The system of equations in the fullysimultaneous method is an extended form of the semisimultaneous method. In the one-zone problem, the system of equations of the fully- and semi-simultaneous solvers are the same.



Figure 11. The general case for investigating condition number of semi-simultaneous and fully-simultaneous methods

Mass and energy balance equations of this single zone as a function of $(P_i = x, T_i = y)$ are as follows

$$f(x,y) = \sum_{i} sign(\Delta P_{ij})C |\Delta P_{ij}|^{n} + Ms_{i} = 0$$
(16)

$$g(x,y) = \sum_{j:F_{ji}>0} C(\Delta P_{ji})^n C_{pa} T_j - \sum_{j:F_{ij}>0} C(\Delta P_{ij})^n C_{pa} y + S_i + Dy + E = 0$$
(17)

$$\Delta P_{ij} = x \left(1 + \frac{a_j}{y} \right) + b_j \tag{18}$$

where

$$\begin{cases} a_{j} = -\frac{gH_{i}}{R} + \frac{g}{2R}(z_{l} + H_{i} - z_{j} - H_{j}) = f(H, z) \\ b_{j} = -(P_{j} - \rho_{j}gH_{j}) + \frac{gH_{j}}{2}(z_{i} + H_{i} - z_{j} - H_{j}) + P_{lj,w} = f(P, T, H, z, P_{w}) \\ D = -\sum_{k} h_{k}A_{k} \\ E = \sum_{k} h_{k}A_{k}T_{mk} \end{cases}$$
(19)

The system of equations are exponential functions of pressure and temperature. According to equations 16 - 19, when different cases are simulated using semi-simultaneous and fully-simultaneous methods, changes in the parameters, i.e., $P, T, H, z, and P_w$ will lead to changes in the coefficients $(a_i, b_i, D, and E)$, but these coefficients do not change the general shape of the functions. Therefore, we can conclude that in all cases equations 16 and 17 have similar curves.

Dols, William; Katal, Ali; Polidoro, Brian; Wang, Liangzhu. "An Investigation of Different Strategies for Solving Coupled Thermal Airflows by Multi-Zone Network Method." Paper presented at COBEE 2018 Conference, Melbourne, Australia. February 5, 2018 - February 9, 2018.

Figure 12 shows the shape of the functions for the 3-zone and light well cases. As seen in both cases, corresponding curves of mass and energy balance equations are two nearly parallel curves. Parallel curves similar to that of Figure 12 correspond to an ill-conditioned system of equations (Dennis et al. 1996). Therefore, linearized systems of equations in the semisimultaneous and fully-simultaneous methods are illconditioned, and the coefficient matrix A is almost singular.



Figure 12. Corresponding curves of mass and energy balance equations (a) 3-zone test case, (b) light well test case

CONCLUSIONS

This study investigated three different coupling strategies for the simulation of steady-state coupled thermal airflow problems using the multi-zone network method: fullysimultaneous, semi-simultaneous and segregated solvers. Compared to the previous work, this study investigates these different strategies from a mathematical point of view to compare the performance of these different methods in solving steady-state problems. Solving different problems using all three solvers shows that for most of the cases the fully-simultaneous and semi-simultaneous solvers show fluctuations in the solution, while the segregated solver shows good convergent results for all cases. Studying the condition number of the matrix equations shows that fully and semisimultaneous solvers have very large condition numbers which means that the system of equations of these two solvers are ill-conditioned, while the matrix equations corresponding to the mass and energy balance equations of segregated solver have smaller condition numbers. Therefore, the matrix equations of the segregated solver are well-conditioned for the cases tested. To show that the matrix equation of fully and semi-simultaneous methods are always ill-conditioned, the parametric mass and energy balance equations of a typical one-zone case are extended, and it is shown that corresponding curves of mass and energy balance equations are nearly parallel which results in an ill-conditioned system of equations.

ACKNOWLEDGEMENT

Drs. Katal and Wang acknowledge the financial support from the Engineering Laboratory of the US National Institute of Standards and Technology (Grant #60NANB15D286).

REFERENCES

Crawley DB, Lawrie LK, Winkelmann FC, Buhl WF, Huang YJ, Pedersen CO, Strand RK, Liesen RJ, Fisher DE, Witte MJ, Glazer J (2001). EnergyPlus: Creating a new-generation building energy simulation program. Energy and Building, 33: 319-331.

J. E. Dennis, Jr., and R. B. Schnabel. 1996. Numerical Methods for Unconstrained Optimization and Nonlinear Equations. SIAM, ISBN 0-89871-364-1.

Kotani, H., R. Satoh, and T. Yamanaka. 2003. Natural ventilation of light well in high-rise apartment building. Energy and Buildings 35(4):427-34.

L. Wang, W.S. Dols and S.J. Emmerich. Simultaneous solutions of coupled thermal airflow problem for natural ventilation in buildings. HVAC&R Research. 18(1-2):11. 2012.

R.Y. Pelletret, W.P. Keilholz. 1997. COMIS 3.0 - a new simulation environment for multi-zone air flow and pollutant transport modelling. Building Simulation 1997. Prague. **IBPSA**

W.S. Dols and B.J. Polidoro. 2015. CONTAM User Guide and Program Documentation, Version 3.2. NIST Technical Note 1887. Gaithersburg, MD. National Institute of Standards and Technology.

W.S. Dols, S.J. Emmerich, B.J. Polidoro. 2016. Coupling the Multizone Airflow and Contaminant Transport Software CONTAM with EnergyPlus using Co-simulation. Building Simulation 9 (4) 469-479, http://dx.doi.org/10.1007/s12273-016-0279-2.

Characterization for an Industrial Wireless Network in a Gas Sensing Scenario

Mohamed Kashef and Richard Candell

Abstract

Employing wireless communications in gas sensing and air quality monitoring is essential in many industrial scenarios where wired networks cannot perform the task safely and effectively. In industrial environments, deploying wireless gas sensing networks becomes a major safety requirement. Hence, we consider the use of wireless networking in safety gas sensing applications. At the National Institute of Standards and Technology (NIST), we have developed a wireless network characterization method to measure latency and reliability of the deployed network without accessing network-level metrics. In this work, we use our characterization method to study the performance of a gas sensing wireless network operating over industrial wireless channels. We have built the gas sensing scenario using the NIST industrial wireless testbed which includes ISA100.11a wireless devices, a channel emulator, and a high performance programmable logic controller (PLC), where the physical process is simulated. We use the channel emulator to replicate the path loss and multipath of the industrial environment while the signal injected into the emulator comes from ISA100.11a wireless devices. Moreover, we inject 4 mA-20 mA gas sensing signals into the wireless devices. In this work, we test various network parameters over the described setup.

Index Terms

Wireless gas sensing, Industrial wireless, ISA100.11a, Confined spaces, Cyber-physical systems, Safety, Channel models.

I. INTRODUCTION

Gas sensing has become a typical component in many industrial systems because of its broad usage in many industrial areas, such as manufacturing, automotive, and medical industries. Specifically, gas leaks monitoring is employed for safety purposes in scenarios when leaks can result in human fatalities. Moreover, distributed gas sensing networks are widely deployed because of different gas densities at different locations and heights. Hence, wiring of these distributed networks represents a major concern in implementation because of the cost and complexity of wiring in some environments. As a result, wireless gas sensing networks provide a more flexible and suitable solution for continuous distributed gas sensing applications.

M. Kashef is an associate researcher with Advanced Networking Division, Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA 20899 (e-mail: mohamed.hany@nist.gov). R. Candell is with Intelligent Systems Division, Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA 20899 (e-mail: richard.candell@nist.gov).

> Distributed with permission of author(s) by ISA 2017 Presented at 2017 Process Control and Safety Symposium and Exhibition; http://www.isa.org

Candell, Richard; Hany, Mohamed. "Characterization for an Industrial Wireless Network in a Gas Sensing Scenario." Paper presented at 2017 ISA International Instrumentation Symposium, Houston, TX, United States. November 6, 2017 - November 9, 2017.

Moreover, confined space monitoring plays an important role in reducing injury that may happen due to gas accumulations. Entering a confined space for performing a maintenance task requires following a proper safety protocol. The major requirements of safety protocols typically include checking air quality before entering and having a portable gas sensor for continuous gas level monitoring while inside. Hence, technologies of gas monitoring equipment are being advanced to achieve these requirements. Wireless communication is used to connect distributed gas monitoring sensors to various control and alarming units allowing continuous monitoring even for closed confined spaces and more reliable measurements from portable devices [1].

A. RESEARCH PURPOSE AND APPROACH

In this work, we test the applicability of using wireless sensing in gas level detection in confined spaces. We focus on the use of ISA100.11a nodes to transmit a typical accumulating carbon monoxide (CO) gas density signal. We characterize the performance of the deployed wireless system through calculating the end-to-end delay and error in the transmitted signals. We also use a channel emulator to generate confined space wireless channel effects on the signals. We study the effects of various system parameters on the performance where we evaluate the use of wireless communications in confined spaces for gas monitoring.

B. RELATED WORK

In [2]–[6], the use of wireless gas sensing is introduced in various scenarios. In [2], the notion of smart sensors is introduced and their requirements and characteristics are discussed. A wireless gas sensor is introduced as an example of the smart sensing applications where the sensor is able to sense various types of gases, store data, and generate alarms through the integration of a microprocessor. In [3], a long-term air quality monitoring testbed is built and evaluated. A wireless IEEE 802.15.4 based node was continuously being operated and it is compared to a standard non-wireless air quality monitoring system. In [4], an experimental setup of a sensor-actuator system for gas detection and control is introduced. The system uses wireless sensors and actuators that deploy the ZigBee standard with the BACnet building automation protocol. In [5], the use of wireless gas sensing is discussed in industrial environments. The authors have justified the importance of wireless gas sensing networks due to the ease of installation and the low cost of maintenance. Finally, in [6], the use of wireless gas sensing is proposed for underground gold and platinum mines. The existence of methane gas in these mines can cause fires and toxicity. Hence, wireless communications can be used in order to improve gas detection capabilities underground.

C. PAPER ORGANIZATION

The remainder of the paper is organized as follows. The gas sensing application and the test setup are discussed in Section II. Numerical results are shown and discussed in Section III. Finally, the paper is concluded in Section IV.

> Distributed with permission of author(s) by ISA 2017 Presented at 2017 Process Control and Safety Symposium and Exhibition; http://www.isa.org

Candell, Richard; Hany, Mohamed. "Characterization for an Industrial Wireless Network in a Gas Sensing Scenario." Paper presented at 2017 ISA International Instrumentation Symposium, Houston, TX, United States. November 6, 2017 - November 9, 2017.

II. GAS SENSING MONITORING SYSTEM

A. APPLICATION

In this paper, we consider the case of welding a tank where the welding process produces gases in rates which are generally non-hazardous. In confined spaces, gas accumulations may lead to an increased density level of these gases which may lead to high risk situations [7], [8]. Specifically, the welding smoke can be extremely toxic as it contains many substances, such as chromium, nickel, arsenic, asbestos, manganese, silica, beryllium, cadmium, nitrogen oxides, phosgene, acrolein, fluorine compounds, carbon monoxide, cobalt, copper, lead, ozone, selenium, and zinc. An example of a case study in which CO became dangerous and life-threatening was discussed in [9].

The existence of CO at high density can lead to poisoning of humans and increased risk of fire and explosion. For a gas to lead to a fire or an explosion, it has to be in the range between the lower explosive limit (LEL) and the upper explosive limit (UEL) which are 12.5 parts per million (ppm) and 74 ppm, respectively, for CO. Moreover, the permissible exposure limit (PEL) of CO is 25 ppm which defines a threshold for alarming a worker to vacate from the confined space.

The considered scenario is shown in Fig. 1. Two ISA100.11a wireless nodes exist inside the tank and the wireless access point is located directly at the opening of the tank. One of the sensors is close to the welding location while the other sensor is on the other side of the tank.



Fig. 1. Wireless System Model.

B. TEST SETUP

The testbed includes a high-performance programmable logic controller (PLC) equipped with 16-bit digital-to-analog (D/A) conversion modules which convert digital stimuli to 0V-10V analog signals. The wireless network is composed of the wireless sensors and a wireless gateway for infrastructure connectivity between nodes. Moreover, the testbed includes a radio frequency channel emulator capable of replicating the multi-path and

> Distributed with permission of author(s) by ISA 2017 Presented at 2017 Process Control and Safety Symposium and Exhibition; http://www.isa.org

path loss environments for a mesh network of up to 8 physical nodes and 56 half-duplex emulated links between those nodes. The channel emulator supports an instantaneous bandwidth of 250 MHz for each emulated channel with an effective dynamic range of 73 dB. The emulator is controlled by a nearby computer which loads the path loss and channel impulse response models for each link.

In the testbed, we have employed the channel emulator in order to include industrial confined space environment effects on the wireless transmissions. We consider the confined space channel model in [10]. This channel model is a generic simulated channel model in a confined space with different numbers of absorbers to represent the existence of workers in the space and/or approximate the effect of the confined space wall material reflectivity on the performance.

The primary objective of this testing is to evaluate the performance of a wireless network deployed in a confined industrial space. We employ a two-stage method for assessing the performance of an industrial wireless network. The method calls for estimation of network delay using maximal length pseudo-random binary sequence correlation followed by estimation of the signal errors through the network adjusted by the delay estimates.

We exploit the conventional general cross correlation (GCC) method for delay estimation [11]. In order to obtain a good delay estimation using the GCC method, we choose an input signal with no repeating patterns within the window of delay estimation. Thus, we use a maximal length binary sequence generated by a linear feedback shift register (LFSR).

In order to quantify the error effect on the performance of an industrial network, an arbitrary input signal is injected at the transmitting node. The injected signal is selected to be similar to practical signals of the sensing node, for example, being a linear signal with certain slope, or a signal with a charging behavior to represent gas accumulations. The received signal is distorted and delayed.

III. **RESULTS**

We employ the wireless system model shown in Figure 1 where the distance between sensor 1, near the welding point, and the access point is fixed at 17.5 meters. Sensor 2 takes two locations at 10 meters and 20 meters away from sensor 1. The distance between sensor 2 to the access point is 17.5 meters when it is 10 meters from sensor 1. The wireless nodes used follow the ISA100.11a standard, but their topology is not enforced. The channel power delay profile is obtained from [10] for four different cases with 0, 1, 3, and 7 absorbers. The increase in number of absorbers may reflect an increase in number of workers inside the confined space which reduces the reverberation within the confined space. An increase in the number of absorbers may also reflect a decrease in reflectivity of the walls. The test of each case has been run for 15 minutes.

First, we measure the average delay for each of the sensors by inserting a maximal length sequence and computing the delay using the GCC method. The time resolution of the obtained delay is 0.5 second. The results are shown in Table I. The delays are

Candell, Richard; Hany, Mohamed. "Characterization for an Industrial Wireless Network in a Gas Sensing Scenario." Paper presented at 2017 ISA International Instrumentation Symposium, Houston, TX, United States. November 6, 2017 - November 9, 2017.

not much affected by the sensors locations and hence the major source of this delay is not packet losses and retransmissions. Instead, the major source of delay is the polling by the PLC of the Modbus transmission control protocol (TCP) server which is located within the ISA100 gateway. The Modbus is a serial communications protocol widely used in industrial applications. Using this testing method, we ascertain the total system delay which includes all sources of delay from data acquisition to PLC computation. The obtained values are generally satisfactory for gas sensing applications due to the slow rate of accumulations.

TABLE I TABLE OF THE DELAY VALUE IN SECONDS FOR THE TWO GAS SENSORS AGAINST THE DISTANCE FOR VARIOUS CHANNEL MODELS.

Channel/Sensor(Distance)	Sensor 1(10m)	Sensor 1(20m)	Sensor 2(10m)	Sensor 2(20m)
No Absorbers	2.5	2.5	3.0	3.0
1 Absorber	2.5	2.5	4.0	3.5
3 Absorbers	2.5	2.0	4.0	3.5
7 Absorbers	2.5	2.5	4.0	4.0

Then, we evaluate the average root mean square (RMS) error in the received signal compared to the transmitted signal after being aligned using the calculated delay. We first consider a ramp signal where the density of the CO increases with a fixed rate over time. The slope of the ramp signal at sensor 1 is 5 ppm/minute and at sensor 2 is 1 ppm/minute because sensor 2 is located away from the welding point. The obtained errors are shown in Table II. The obtained RMS errors compared to signal magnitude are very low, and hence the sensors have a satisfactory performance. The errors are smaller at sensor 2 because of the smaller magnitude of the transmitted signal. Moreover, the error is not highly affected by the nodes locations, and hence, the major source of error is signal acquisition which may include quantization errors and biases.

TABLE II TABLE OF ERROR VALUES FOR A RAMP SIGNAL FOR THE TWO GAS SENSORS AGAINST THE DISTANCE FOR VARIOUS CHANNEL MODELS.

Channel/Sensor(Distance)	Sensor 1(10m)	Sensor 1(20m)	Sensor 2(10m)	Sensor 2(20m)
No Absorbers	0.0633	0.0654	0.0258	0.0300
1 Absorber	0.0593	0.0599	0.0193	0.0210
3 Absorbers	0.0743	0.0614	0.0167	0.0182
7 Absorbers	0.0669	0.0699	0.0279	0.0259

Then, we consider a charge signal to represent the effect of the gas accumulations in a confined space where the density of the CO increases with a time constant of 1.67 minutes for sensor 1 and 8.33 minutes for sensor 2. The time constant is smaller, and hence, the accumulation of the gas is faster at sensor 1, because sensor 1 is located close to the welding point. The obtained errors are shown in Table III. The charge signal errors have a similar trend to the ramp signal. The average RMS is relatively higher in this case because quantization errors are larger in the charge signal compared to the ramp signal.

> Distributed with permission of author(s) by ISA 2017 Presented at 2017 Process Control and Safety Symposium and Exhibition; http://www.isa.org

Candell, Richard; Hany, Mohamed. "Characterization for an Industrial Wireless Network in a Gas Sensing Scenario." Paper presented at 2017 ISA International Instrumentation Symposium, Houston, TX, United States. November 6, 2017 - November 9, 2017.
TABLE III TABLE OF ERROR VALUES FOR A CHARGE SIGNAL FOR THE TWO GAS SENSORS AGAINST THE DISTANCE FOR VARIOUS CHANNEL MODELS.

Channel/Sensor(Distance)	Sensor 1(10m)	Sensor 1(20m)	Sensor 2(10m)	Sensor 2(20m)
No Absorbers	0.1741	0.1634	0.0709	0.0681
1 Absorber	0.1588	0.1739	0.1211	0.0878
3 Absorbers	0.1612	0.1840	0.0582	0.0868
7 Absorbers	0.1599	0.1827	0.1253	0.1257

IV. CONCLUSIONS

In this work, we have tested the performance of two ISA100.11a wireless gas sensors operating in a simulated welding process within a confined space. The sensors transmit the CO level produced by a welding process close to one of the sensors. We have employed the NIST wireless testbed to replicate the confined space wireless channel effects on the performance of the wireless nodes. In the current setup with the existence of a line of sight (LOS) and having the typically slow gas accumulations, the measured performance is considered satisfactory for gas safety applications. The measured delays and errors are mainly produced by Modbus polling and data acquisition errors. Hence, more improvement still can be obtained while deploying wireless for confined space communications by reducing Modbus publication and polling delays. Indeed, many ISA100.11a gateways have Modbus servers that refresh at a much slower rate than the fastest allowable sensor update rates. This can mask or exacerbate delay issues of the wireless network. Improvements to the standards to address storage and publication rates of sensor data by the gateway are needed. However, the current ISA100.11a standard supports gas sensing and alarming for single-chamber confined spaces as demonstrated by this study. Confined spaces with multiple chambers will be addressed in a later study.

DISCLAIMER

Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

REFERENCES

- [2] G. W. Hunter, J. R. Stetter, P. Hesketh, and C.-C. Liu, "Smart sensor systems," The Electrochemical Society Interface, vol. 19, no. 4, pp. 29-34, 2010. [Online]. Available: http://interface.ecsdl.org/content/19/4/29.abstract
- [3] M. Leccardi, M. Decarli, L. Lorenzelli, P. Milani, P. Mettala, R. Orava, and E. Barborini, "Long-term outdoor reliability assessment of a wireless unit for air-quality monitoring based on nanostructured films integrated on micromachined platforms," Sensors, vol. 12, no. 12, p. 81768192, Jun 2012. [Online]. Available: http://dx.doi.org/10.3390/s120608176

Distributed with permission of author(s) by ISA 2017 Presented at 2017 Process Control and Safety Symposium and Exhibition; http://www.isa.org

Candell, Richard; Hany, Mohamed. "Characterization for an Industrial Wireless Network in a Gas Sensing Scenario." Paper presented at 2017 ISA International Instrumentation Symposium, Houston, TX, United States. November 6, 2017 - November 9, 2017.

^{[1] &}quot;Technical advisory on working safely in confined spaces," https://www.wshc.sg/files/wshc/upload/cms/file/2014/ cs2.pdf, accessed: 2017-07-26.

- [4] A. Somov, A. Baranov, and D. Spirjakin, "A wireless sensoractuator system for hazardous gases detection and control," Sensors and Actuators A: Physical, vol. 210, pp. 157 - 164, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924424714000892
- [5] P. C. Jain and R. Kushwaha, "Wireless gas sensor network for detection and monitoring of harmful gases in utility areas and industries," in 2012 Sixth International Conference on Sensing Technology (ICST), Dec 2012, pp. 642-646.
- [6] A. M. Abu-Mahfouz, S. J. Isaac, C. P. Kruger, N. Aakvaag, and B. Fismen, "Wireless gas sensing in south african underground platinum mines," in 2014 IEEE Wireless Communications and Networking Conference (WCNC), April 2014, pp. 3432-3437.
- [7] "Afscme. health and safety factsheet," http://www.afscme.org/news/publications/workplace-health-and-safety/factsheets/pdf/Welding-Hazards-AFSCME-fact-sheet.pdf, accessed: 2017-07-26.
- "What are the hazards from gases during welding and cutting?" http://www.twi-global.com/technical-[8] knowledge/faqs/faq-what-are-the-hazards-from-gases-during-welding-and-cutting/, accessed: 2017-07-26.
- [9] A.-B. Antonsson, B. Christensson, J. Berge, and B. Sjgren, "Fatal carbon monoxide intoxication after acetylene gas welding of pipes," The Annals of Occupational Hygiene, vol. 57, no. 5, pp. 662-666, 2013. [Online]. Available: + http://dx.doi.org/10.1093/annhyg/mes104
- [10] E. Genender, C. L. Holloway, K. A. Remley, J. M. Ladbury, G. Koepke, and H. Garbe, "Simulating the multipath channel with a reverberation chamber: Application to bit error rate measurements," IEEE Transactions on Electromagnetic Compatibility, vol. 52, no. 4, pp. 766-777, Nov 2010.
- [11] Algorithms for Radiolocation. Boston, MA: Springer US, 2000, pp. 41-66. [Online]. Available: http://dx.doi.org/10.1007/0-306-47329-1 4

Candell, Richard; Hany, Mohamed. "Characterization for an Industrial Wireless Network in a Gas Sensing Scenario." Paper presented at 2017 ISA International Instrumentation Symposium, Houston, TX, United States. November 6, 2017 - November 9, 2017.

Diagnostics of machine tool linear axes via separation of geometric error sources

Gregory W. Vogl¹ and Michael E. Sharp¹

¹National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, 20899, USA gregory.vogl@nist.gov michael.sharp@nist.gov

ABSTRACT

Manufacturers need automated, efficient, and robust methods to diagnose the condition of their machine tool linear axes with minimal disruptions to production. Recently, a method was developed to use data from an inertial measurement unit (IMU) to measure changes in geometric error motions. A linear axis testbed, established for verification and validation purposes, revealed that the IMU-based method was capable of measuring translational and angular deviations with acceptable test uncertainty ratios. In this study, a rail of the linear axis testbed was mechanically degraded to simulate spalling, a common degradation mechanism that can occur during machine tool operations. The rail was degraded in discrete steps from its nominal state (no degradation) to its final state (a failure state of the rail), and IMU and laser-based reference data was collected at each test stage. The contribution of geometric errors from the rail-based degradation was then separated with a technique that utilizes the various data for each run. Diagnostic metrics can then be defined for use with the IMU to facilitate industrial applications by informing the user of the magnitude and location of wear and any violations of performance tolerances.

1. INTRODUCTION

Over a machine tool's lifetime, various faults lead to performance degradation, lowering accuracy and reliability (Li, Wang, Lin & Shi, 2014). Typical sources of errors within linear axes are due to pitting, wear, corrosion, and cracks of the system components such as guideways and recirculating balls (Zhou, Mei, Zhang, Jiang & Sun, 2009). A typical machine tool has multiple linear axes, and their accuracies directly impact the quality of manufactured parts. As degradation increases, tool-to-workpiece errors increase which may eventually result in a loss of production quality

and/or a performance failure (Uhlmann, Geisert & Hohwieler, 2008). Yet knowledge of degradation is elusive; proper assessment of axis degradation is often a manual, time-consuming, and potentially cost-prohibitive process.

While direct methods for machine tool performance evaluation are well-established (International Organization for Standardization, 2012) and reliable for positiondependent error quantification, such measurements typically interrupt production (Khan & Chen, 2009). An online condition monitoring system for linear axes is needed to help reduce machine downtime, increase productivity and product quality, and improve knowledge about manufacturing processes (Teti, Jemielniak, O'Donnell & Dornfeld, 2010). Previous attempts at condition monitoring of linear axes had limited success, partly because of the lack of robustness and defined relationships of signals to axis degradation composed of a wide range of spatial frequencies. Consequently, efficient quantitative measures are needed to monitor the degradation of linear axes.

Recently, accelerometers have been used for dynamic metrology of machine tools (Sato & Nagaoka, 2011; Sato, Nagaoka & Sato, 2015; Smith & Hocken, 2013) and sixdegree-of-freedom motion sensors exist within integrated circuit (IC) components (InvenSense Incorporated, 2016). Thus, the use of an inertial measurement unit (IMU) is attractive for on-machine condition monitoring.

Previous work has shown that one potential solution for online monitoring of linear axis degradation is the use of an IMU (Vogl, Weiss & Donmez, 2015; Vogl, Donmez & Archenti, 2016) that processes accelerometer and rate gyroscope data to detect changes in the translational and angular error motions due to axis degradation (Vogl et al., 2016). For industrial application, the IMU should be physically small and economical while satisfying measurement needs. As seen in Figure 1, a custom industrial IMU was created that is about $9.0 \text{ cm} \times 6.5 \text{ cm} \times 4.6 \text{ cm}$ in size and contains a triaxial accelerometer and a triaxial rate gyroscope to satisfy the necessary design constraints of cost,

Sharp, Michael; Vogl, Gregory. "Diagnostics of machine tool linear axes via separation of geometric error sources." Paper presented at 2017 Annual Conference of the Prognostics and Health Management Society, St. Petersburg, FL, United States. October 2, 2017 - October 5, 2017.

Gregory Vogl et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



size, and accuracy. The bandwidths and noise properties of these sensors are shown in Table 1.

Figure 1. (a) Isometric view of industrial IMU and (b) top view of industrial IMU without its lid.

Table 1. Specified properties o	of sensors used in the IMU
---------------------------------	----------------------------

Sensor	Bandwidth ^a	Noise					
Accelerometer	0 Hz to 500 Hz	20 (µm/s²)/√Hz					
Rate Gyroscope	0 Hz to 200 Hz	35 (µrad/s)/√Hz					
^a frequencies correspond to half-nower points also known as 3 dB							

points

In this study, a rail of a linear axis was mechanically degraded from its nominal state (no degradation) to its final state (a failure state of the rail) to simulate spalling. IMU data was collected at each test stage, and an analysis technique was used to isolate the contribution of geometric errors due to the rail-based degradation. This paper outlines the experimental setup, analysis techniques, and preliminary metrics for diagnostics of rail-based degradation of linear axes. The results support the future use of IMUs for on-machine condition monitoring of linear axes within manufacturing assets.

2. EXPERIMENTAL SETUP

During operation in a machine tool, the components of a linear axis, such as rails and recirculating balls, degrade and cause the geometric performance of the axis to change. To study the ability of the IMU and accompanying diagnostics

to assess the degradation, a linear axis testbed is utilized. As seen in Figure 2, the testbed includes a linear axis system, the IMU on the carriage of the linear axis system, and a commercial laser-based system for measuring the geometric errors of the axis. While the laser-based metrology system measures the motion of the carriage with respect to the base of the linear axis, the IMU measures the changes in the inertial motion of the carriage. The laser-based system is used for verification and validation (V&V) of the IMU-based results and measures straightness and angular error motions over the travel length of 0.32 m with standard uncertainties of 0.7 µm and 3.0 µrad, respectively.



Figure 2. Linear axis testbed containing a linear axis, laserbased reference system, and IMU.

In this study, one rail of the linear axis (seen in Figure 2) was mechanically degraded to simulate spalling that can occur while a linear axis operates within machine tools. As seen in Figure 3(a), the rail has raceway grooves, upon which the recirculating balls of the trucks roll, constraining the motion of the carriage to be nominally linear. The interaction of the balls and the raceway grooves causes geometric error motions, and as the balls or the raceway grooves degrade, the geometric performance of the carriage changes.

The experiment performed involved degradation of a raceway groove, seen in Figure 3(a) prior to any mechanical degradation. To mechanically simulate spalling, a handheld grinder was used to wear the surface of the raceway groove. The length of the degradation zone was increased over the course of the experiment. Figure 3(b) to Figure 3(d) shows the progression of the 'spalling' on the raceway groove of the rail up to a length of 75 mm. The degradation zone length was increased incrementally by about 5.4 mm from its nominal state (no degradation, seen in Figure 3(a)) to its final state (significant degradation, seen in Figure 3(d)). Figure 4 shows the relative location and length of the degradation zone in the raceway groove for each of the fifteen (15) test stages.

Of course, the degradation is three-dimensional and not just represented by a nominal length; each degradation zone has a depth experienced by each ball of a truck. Figure 5 shows the degradation profiles measured with a 3-mm-diameter probe, typically used for coordinate measuring machines,

Sharp, Michael; Vogl, Gregory. "Diagnostics of machine tool linear axes via separation of geometric error sources." Paper presented at 2017 Annual Conference of the Prognostics and Health Management Society, St. Petersburg, FL, United States. October 2,

which was attached to a digital micrometer (with an uncertainty of 1 µm) fixed to the carriage. The probe diameter of 3 mm is similar, but different from, the bearing ball nominal diameter of 3.969 mm. These measurements are used to corroborate and correlate the relative degradation estimations of the IMU-based results. Also, the micrometer measurements have uncertainties due to fixturing, thermal differences, etc., that are beyond the scope of this paper. One sample (Test stage 5) appears to be an outlier, likely caused by human error in fixturing. Regardless of this outlier, it appears that 'spalling' exists with depths greater than 50 µm and cumulative (integral) damage linearly increasing from 0 to $\approx 58 \ \mu m^2$ (depth \times length).



Figure 3. (a) Rail with no degradation, (b) rail with about 14 mm of 'spalling', (c) rail with about 37 mm of 'spalling', and (d) rail with about 75 mm of 'spalling'.

Finally, for each test stage, fifty (50) runs of IMU data were collected bidirectionally at slow, moderate, and fast axis speeds (Vogl et al., 2016). Afterwards, ten (10) runs of reference data were collected bidirectionally at finite positions of travel, specifically every 1 mm between travel positions 1 mm and 321 mm. The IMU data (accelerometer and rate gyroscope data) for each run was processed according to the IMU-based method to yield the 6-degree-offreedom error motions (Vogl et al., 2016): one positioning error motion, two straightness error motions, and three angular error motions. Thus, six error motions are associated with each of fifty runs for all fifteen stages. These error motions were used for physics-based diagnostics described in the following sections.



Figure 4. Schematic of relative location and length of degradation zone in raceway groove for each stage.



Figure 5. Degradation profiles of the rail raceway groove for certain test stages.

3. CONTRIBUTORS TO MEASURED ERROR MOTIONS

The IMU-based method should ultimately be used for diagnostics, prognostics, and health management of linear axes within production machines. Achieving these goals relies on using the IMU data to detect changes in error motions and determine the causes of these changes through diagnostic methods. However, the development of a root-

Sharp, Michael; Vogl, Gregory. "Diagnostics of machine tool linear axes via separation of geometric error sources." Paper presented at 2017 Annual Conference of the Prognostics and Health Management Society, St. Petersburg, FL, United States. October 2,

cause analysis first requires an understanding of the general types of error sources and their effects on the error motions.

Table 2 shows the physical sources of error motions in linear axes and the main errors affected by these physical sources. As seen in Table 2, the trucks and rails affect mainly the two straightness error motions and the three angular error motions, while the other components (lead screw, ball nut, and controlled motor) move the carriage and affect mainly the one positioning error motion. Technically, all components affect all error motions, but Table 2 focuses on the significant relationships between sources and error motions for root-cause diagnostics.

Та	b	le 1	2.	Error	sources	and	error	moti	ons	in	typi	cal	li	near	ax	is

	Main Error Motions Affected by Source							
Error Source	Positioning	Straightness	Angular					
Truck		Yes	Yes					
Rail		Yes	Yes					
Lead Screw	Yes							
Ball Nut	Yes		_					
Controlled Motor	Yes							

Each error motion of Table 2 can be decomposed into its constituent parts, as outlined in Table 3, which shows the types of error motions caused by each physical error source. In general, each error motion is composed of repeatable and non-repeatable terms. The non-repeatable terms simply do not repeat from run to run among collected IMU data. The repeatable components, on the other hand, are repeatable from run to run but are of one of two natures: 'phase-fixed' or 'phase-shifted'. The 'phase-fixed repeatable' terms are fixed functions of carriage position, but the 'phase-shifted repeatable' terms shift their phases as functions of carriage position from run to run. For example, as seen in Table 3, a rail yields phase-fixed repeatable error motions because the rail is fixed in space, while a truck yields phase-shifted repeatable error motions because the balls in the truck recirculate in a way that causes the balls to 'shift' in phase relative to the rail from run to run.

Tał	ole 3	3. E	Error	motion	types	in	typical	linear	axis
-----	-------	------	-------	--------	-------	----	---------	--------	------

	Error Motion Type							
Error Source	Phase-Fixed	Phase-Shifted	Non-					
	Repeatable	Repeatable	Repeatable					
Truck		Yes	Yes					
Rail	Yes	—	Yes					
Lead Screw	Yes	—	Yes					
Ball Nut		Yes	Yes					
Controlled Motor	Yes	_	Yes					

The question remains: How do we separate the error terms outlined in Table 3 for each of the error motions outlined in Table 2? Answering this question relies on the development of a comprehensive root-cause analysis, which is beyond the scope of this paper. In contrast, the work described herein focuses only on determining and diagnosing the rail-based errors. The 'phase-fixed repeatable' rail-based error motions are first separated from all other errors via the technique outlined in the next section, then diagnostics are performed on these error motion components to track rail degradation.

4. METHOD TO DIAGNOSE RAIL-BASED CHANGES IN ERROR MOTIONS

After collecting a total of 50 individual runs for the IMU data and 10 runs for the laser-based reference data, a converged value for each point along the rail path needed to be created from these individual passes. This operation serves to remove noise (part of the non-repeatable errors referenced in Table 3) and decrease the uncertainty associated with the individual runs by creating a higher fidelity map of the displacement at each point along the rail. The simplest method for combining the 50 individual runs would be to take a truncated mean of the concurrent position data, excluding the extrema values as outliers. However, for this work, it has been found that the high variation of the IMU signal low-frequency components between collection instances obfuscates small-scale degradation isolation. This variance is likely due to sensor noise and natural variations in the absolute position and orientation of the rail and IMU due to temperature and other external effects.

To help reduce the effects of the low-frequency noise on the final resulting metric, a high-pass median filter is applied to the data prior to combining. The width of this filter is determined as half the length of the rail (≈ 0.15 m). Mandating the largest window allows the capture of the lowest frequency data possible, while still removing the base deflection of the rail.

By using this form of high-pass filtering, environmental effects (such as slight warping from temperature change) and benign changes in setup can be efficiently filtered out of the resulting analysis. The basic assumption allowing this is that detrimental degradation will occur in localized areas rather than to the overall rail. Pitting or spalling are examples of degradation attuned to this form of degradation monitoring. Full beam shifting or relative torqueing respective to the opposing rail could also be monitored for via static baseline monitoring or concurrent change monitoring between the rails. These are not a focus of this paper, but may be addressed in future work.

Once an accurate estimation of the true underlying deflection path of the rail for each test stage is created, the next step is to map the change of the path deflection over time. The deflection path of the earliest test path (zero degradation) is established as the prototype baseline to quantify these changes by comparing it to all subsequent tests paths.

After the point-by-point residuals are determined for each query path, the final step is to calculate the analytic amplitude

4

Sharp, Michael; Vogl, Gregory. "Diagnostics of machine tool linear axes via separation of geometric error sources." Paper presented at 2017 Annual Conference of the Prognostics and Health Management Society, St. Petersburg, FL, United States. October 2,

of these values via the Hilbert transform. This allows for better trending with algorithms set to monitor the integral of the total degradation across the rail.

5. RESULTS

Both the IMU data and data taken from the laser-based reference system are processed and evaluated in the exact same way to allow for direct comparisons between the results. An example of the raw data for the vertical (Z) direction collected during a single testing cycle is shown in Figure 6. This is a direction expected to be affected by the induced spalling of this experiment.



Figure 6. Example uncleaned data.

Each line in the plots represents one of the redundant runs that were performed sequentially, with the assumption that no appreciable change occurs to the condition of the rails during data collection. Each testing cycle collects 10 runs from the laser-based sensor, and 50 from the IMU. The difference in magnitude from these two sets of data occurs due to the IMU having a worse signal to noise ratio, and incidentally, is why the IMU needs more runs to characterize the underlying characteristics.

In Figure 6, the variation between runs for the IMU data is significantly greater than the laser-based data. This is expected as the laser measurement device is a higher fidelity piece of equipment. Nonetheless, the IMU-based results can be cleaned and merged to find the underlying form of the rail deflection path. Figure 7 shows the average deflection values taken during the first test cycle for the translational data in each of the three directions.

Once the data for each test day has been cleaned and consolidated, the next step is to compare each day with the baseline fault-free data collected on the first testing cycle (Stage 1). The analytic amplitude of these residual values can be tracked over time to produce metrics indicating the overall health of the rail.



Figure 7. Cleaned and merged data.

Figure 8 shows these error metrics based on rail position for both the high-fidelity laser-based data and IMU-based data for selected test stages. Much can be learned about the quality of the IMU versus that of the laser from these plots. First, notice that the progressive peak centered near the 0.13 m position appears clearly in both sets of data. This corresponds to the front truck encountering the induced degradation. A secondary peak near 0.23 m corresponding to the second truck encountering the degradation can also be seen appearing to a lesser degree in both data sets. This echoing of form from the high-fidelity laser-based data to the IMUbased data confirms that the degradation being captured by the IMU is accurate in form and position.



Figure 8. Select position-based straightness error values.

Another important conclusion that can be drawn from these plots is that, although the form and shape of the captured degradation is similar between the two sensors, the magnitudes are significantly different. The amplitudes of the captured degradation from the IMU are roughly two to three times smaller than those from the laser-based system. This

Sharp, Michael; Vogl, Gregory. "Diagnostics of machine tool linear axes via separation of geometric error sources." Paper presented at 2017 Annual Conference of the Prognostics and Health Management Society, St. Petersburg, FL, United States. October 2,

indicates that the sensitivity of the IMU-based method is lower than that of the laser-based system. This result is further corroborated by the difference in the signal-to-noise ratio between the IMU and reference data. Despite this higher noise level and lower sensitivity, the IMU is still able to capture and accurately characterize the induced degradation with a traceable progression similar to that found in the laser data

Figure 9 shows the progression of the total damage indicated across the rail at each test stage for both the IMU- and reference-based datasets. These charts show the integral of the measured damage in both the vertical and horizontal directions, which are the directions most affected by the induced 'spalling'. In both instances, Stage 1 is taken as the baseline and therefore by definition has the trivial value of zero.



Figure 9. Straightness integral damage progression in directions of interest.

An important point of clarification is that the metrics developed in this work do not directly translate to the physically measurable units of degradation. The layers of processing to produce the metrics obfuscate the direct mapping. In other words, a metric value of 20 m² does not necessarily imply that a hand-made, direct measure of the degradation would also yield a result of 20 m². For example, if both the IMU derived metric and the hand measurements show values of 40 when the laser sensor derived metric shows a value closer to 100, this does not instantly imply that the IMU is more accurate than the laser-based results. Such discrepancies could result because the signal-to-noise ratio is higher in this bandwidth, or because of any number of other artifacts of processing. Nonetheless, and more importantly, the relative progressions and their correlations to the values of degradation manually measured with a handheld micrometer are very good.

The correlation between cumulative damage captured by the IMU- and the laser-based methods is approximately 0.98 for both the Y- and Z- directions, respectively. This confirms statistically that though there are differences in scale, the captured trends are nearly identical. Also important to note, is that the Y- and Z- directions for both the laser- and IMUbased degradation show a correlation above 0.95, both to each other and the handmade direct measurements of the degradation, lending further credibility to the merit and quality of the captured values and trends. A full correlation analysis can be seen in Figure 10.



Figure 10. Straightness correlation analysis.

A final interesting note on this analysis is that, even though as expected, both the laser- and IMU-based X-directional errors (positioning error motions) show little correlation to the handmade measurements of cumulative degradation, they do exhibit moderately strong correlations to each other (0.72). This seems to indicate that although the positioning error motion does not capture the induced damage in a significant manner (per Table 2), the positioning error motion is still affected by the subtle changes in the physical system; in this case, specifically in the direction of travel.

Similar to the translational errors captured by the accelerometer, angular errors were captured via the accelerometer and rate gyroscope in the IMU. Shown in Figure 11, the progressive stages of the induced degradation are captured by both the IMU and the laser-based reference system. In Figure 11, it is easy to distinguish the two peaks associated with the two separate rail trucks encountering the degradation as they pass. The angular data is more consistent than the straightness data for lower frequencies, making it easier to analyze for fine details.

The angular integral damage progressions, as shown in Figure 12, are all also highly similar to those found in both the straightness analysis and handmade measurements. Again, the scales and seeming sensitivities are significantly lower in the IMU-based results than those of the high fidelity

Sharp, Michael; Vogl, Gregory. "Diagnostics of machine tool linear axes via separation of geometric error sources." Paper presented at 2017 Annual Conference of the Prognostics and Health Management Society, St. Petersburg, FL, United States. October 2,

laser-based results, but the overall trends are nearly identical. It is important to note here that the relative values of the metrics in each direction are sensitive both to the level of noise in that signal and the size of the filtering window selected for denoising, as described in an earlier section. Steps to mitigate this effect and/or more precisely prescribe the selection of the denoising window are currently under investigation. The more important aspect for this paper is that the resulting trends show strong correlations both between themselves, and to the direct hand-held micrometer measurements, despite the differences in scales.



Figure 11. Select position-based angular error values.



Figure 12. Angular integral damage progression in directions of interest.

Figure 13 shows the correlation analysis of the angular errors with the manual hand-micrometer along depth measurements. From this figure, we can see that the lowest correlation to the independent hand measurements is 0.93, which is a strong correlation. In fact, all the correlations

presented in this analysis are strong enough to suggest that the metrics are all capturing the same underlying cause, in this case, the progression of induced degradation. Such high correlation between independent metrics inferring degradation supports a high confidence in any calculated degradation values reported by an IMU utilizing this diagnostic approach.



Figure 13. Angular correlation analysis.

6. CONCLUSIONS

An experiment was designed and implemented on a linear axis rail to simulate 'spalling' on a common production machine. The purpose of this experiment was to verify the viability of a novel IMU-based sensor system and to develop metrics suitable for capturing and quantifying degradation signatures for diagnostic and prognostic monitoring. Highfidelity laser-based data and handmade measurements of the known degradation were collected and used for verification and validation of the IMU-based results and associated metrics.

Both the IMU-based and laser-based detectable degradation metrics showed strong correlations to one another as well as to the hand-taken values. These strong correlations confirm that the IMU is a viable, accurate, and efficient sensing system that can provide comparable results to the more expensive laser-based system. The analysis further shows that although the accelerometer data from the IMU exhibits a significant low-frequency variance, steps can be taken to mitigate this effect.

Future work will focus on the development of comprehensive methods for identifying and isolating both truck-based defects and rail-based defects. Additional data will be generated in an ongoing effort to verify and validate the IMUbased method and associated metrics as well as to produce standard datasets on which novel diagnostic algorithms can be developed. Ultimately, such algorithms could enable the use of IMU-based methods for diagnostics, prognostics, and

Sharp, Michael; Vogl, Gregory. "Diagnostics of machine tool linear axes via separation of geometric error sources." Paper presented at 2017 Annual Conference of the Prognostics and Health Management Society, St. Petersburg, FL, United States. October 2,

2017 - October 5, 2017

health management of linear axes within production machines

ACKNOWLEDGEMENTS

The authors thank Brian Pries, Travis Shatzley, Dan Falvey, and Jay Brandenburg of the Fabrication Technology Group at the National Institute of Standards and Technology (NIST) for their outstanding contributions with the experimental setup.

NIST DISCLAIMER

Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

REFERENCES

- International Organization for Standardization (2012). ISO 230-1 - test code for machine tools – part 1: Geometric accuracy of machines operating under no-load or quasistatic conditions.
- InvenSense Incorporated (2016). MPU-6050 six-axis (gyro + accelerometer) MEMS MotionTracking[™] device: https://www.invensense.com/products/motiontracking/6-axis/mpu-6050/
- Khan, A. W. & Chen, W. (2009). Calibration of CNC milling machine by direct method. 2008 International Conference on Optical Instruments and Technology: *Optoelectronic* Measurement Technology and Applications (p. 716010), November 16-19, 2008, Beijing, China. doi: 10.1117/12.807066
- Li, Y., Wang, X., Lin, J., & Shi, S. (2014). A wavelet bicoherence-based quadratic nonlinearity feature for translational axis condition monitoring. Sensors, vol. 14(2), pp. 2071-2088.
- Sato, R. & Nagaoka, K. (2011). Motion trajectory measurement of NC machine tools using accelerometers. International Journal of Automation Technology, vol. 5(3), pp. 387-394.
- Sato, R., Nagaoka, K., & Sato, T., "Machine motion trajectory measuring device, numerically controlled machine tool, and machine motion trajectory measuring method," USA Patent US9144869 B2, Sep. 29, 2015.
- Smith, K. S. & Hocken, R. J., "Dynamic metrology methods and systems," USA Patent US8401691 B2, Mar. 19, 2013.
- Teti, R., Jemielniak, K., O'Donnell, G., & Dornfeld, D. (2010). Advanced monitoring of machining operations. CIRP Annals - Manufacturing Technology, vol. 59(2), pp. 717-739. doi: 10.1016/j.cirp.2010.05.010
- Uhlmann, E., Geisert, C., & Hohwieler, E. (2008). Monitoring of slowly progressing deterioration of

computer numerical control machine axes. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, vol. 222(10), pp. 1213-1219.

- Vogl, G. W., Weiss, B. A., & Donmez, M. A., "A sensorbased method for diagnostics of machine tool linear axes," presented at the Annual Conference of the Prognostics and Health Management Society 2015, Coronado, CA, 2015.
- Vogl, G. W., Donmez, M. A., & Archenti, A. (2016). Diagnostics for geometric performance of machine tool linear axes. CIRP Annals - Manufacturing Technology, vol. 65(1), pp. 377-380.
- Zhou, Y., Mei, X., Zhang, Y., Jiang, G., & Sun, N. (2009). Current-based feed axis condition monitoring and fault diagnosis. 4th IEEE Conference on Industrial Electronics and Applications, ICIEA 2009 (pp. 1191-1195), May 25-27, 2009, Xi'an, China. doi: 10.1109/ICIEA.2009.5138390

BIOGRAPHIES



Dr. Gregory W. Vogl is a Mechanical Engineer at the National Institute of Standards and Technology (NIST) located in Gaithersburg, Maryland. He received his B.S. (2000), M.S. (2003), and Ph.D. (2006) degrees in Engineering Mechanics from Virginia Tech, Virginia, USA. Currently,

Greg is a member of the Prognostics, Health Management, and Controls (PHMC) project, which seeks to help enable robust real-time diagnostics and prognostics for smart manufacturing systems. Previously, he designed, fabricated, and experimented on microelectromechanical systems as a National Research Council Postdoctoral Researcher at NIST. He then joined the Production Systems Group, in which he worked on machine tool metrology and standards development. His interests include machine tool spindle health, diagnostic and prognostic methods, nonlinear dynamics, engineering mechanics, and metrology.



Dr. Michael E. Sharp is a Reliability Engineer at the National Institute of Standards and Technology (NIST) located in Gaithersburg, MD. He received a B.S (2007), M.S. (2009), and Ph.D. (2012) in Nuclear Engineering from the University of Tennessee, Knoxville, TN, USA. His

research interests include signal analytics, machine learning, artificial intelligence, optimization, and natural language processing. Michael has worked on a wide array of projects including image processing for elemental material recognition, prognostics for electric motors undergoing transient operations, and control systems for deep space reactors. He currently works with the NIST Systems Integration Division for Smart Manufacturing.

Sharp, Michael; Vogl, Gregory. "Diagnostics of machine tool linear axes via separation of geometric error sources." Paper presented at 2017 Annual Conference of the Prognostics and Health Management Society, St. Petersburg, FL, United States. October 2,

Propagation of orientation uncertainty of 3D rigid object to its points

Marek Franaszek National Institute of Standards and Technology Gaithersburg, MD 20899, USA marek@nist.gov Geraldine S. Cheok National Institute of Standards and Technology Gaithersburg, MD 20899, USA cheok@nist.gov

Abstract

If a CAD (Computer-Aided Design) model of a rigid object is available, the location of any point on an object can be derived from the measured 6DOF (degree-offreedom) pose of the object. However, the uncertainty of the measured pose propagates to the uncertainty of the point in an anisotropic way. We investigate this propagation for a class of systems that determine an object pose by using point-based rigid body registration. For such systems, the uncertainty in the location of the points used for registration propagates to the pose uncertainty. We find that for different poses of the object, the direction corresponding to the smallest propagated uncertainty remains relatively unchanged in the object's local frame, regardless of object pose. We show that this direction may be closely approximated by the moment of inertia axis which is based on the configuration of the fiducials. We use existing theory of rigid-body registration to explain the experimental results, discuss the limitations of the theory, and practical implications of our findings.

1. Introduction

The pose of a rigid object is determined by six degrees of freedom (6DOF): three angles parametrizing the object orientation matrix **R** and three components of vector $\boldsymbol{\tau}$ defining the object's location. When the Computer Aided Design (CAD) model of an object is available, any Point of Interest (POI) on the object can be calculated using the 6DOF data obtained from pose measuring systems. It is then important to understand how uncertainty in the measured pose propagates to the uncertainty of a POI. In manufacturing, when rigid parts are assembled by autonomous robots, uncertainty in pose has to be propagated to the grasp points or points of contact between parts [1-4]. In peg-in-hole testing (routinely used to benchmark a robot's performance [5-8]) uncertainty in the hole location directly impacts the test outcome [9-13].

In [14], it was shown that object orientation uncertainty propagates to a POI in an anisotropic way (i.e., it depends on the direction in space) and the direction along which the propagated uncertainty is the smallest is determined by the eigenvector corresponding to the largest eigenvalue of the covariance matrix of orientation noise. Prior knowledge of this direction would be beneficial in planning assembly tasks but three issues need to be addressed. 1) It is not known if the direction of the relevant eigenvector is fixed in the local object coordinate frame, i.e. whether it depends on object orientation (large, unknown variations of the direction would negate any potential benefits of exploiting anisotropic propagation of object orientation the uncertainty). 2) Evaluating the covariance matrix of orientation noise is a time-consuming task which requires collection of many repeated measurements. Therefore, it would be helpful if the direction of the smallest propagated uncertainty could be estimated in a more efficient way. 3) In practical applications, the need for very accurate locations of different POIs may be required in an assembly process. To optimize process planning, users may need to know if the direction of the smallest propagated uncertainty can be controlled and changed based on the location of the POI. The objective of this study is to address these issues for a certain class of pose measuring systems.

Orientation components of pose are rarely measured directly, and in this paper, we investigate the class of pose measuring systems that acquire points (fiducial points) for rigid body registration to determine pose. Some systems use rigidly attached markers to the object or can automatically identify some characteristic points on the object. For these systems, the transformation $\{R, \tau\}$ is found by minimizing the Fiducial Registration Error (*FRE*)

$$FRE(\boldsymbol{R},\boldsymbol{\tau}) = \sqrt{\frac{1}{J}\sum_{j=1}^{J} \left\| \boldsymbol{R}\boldsymbol{X}_{j} + \boldsymbol{\tau} - \boldsymbol{Y}_{j} \right\|^{2}}, \qquad (1)$$

where $\{X\}_J$ is a set of J fiducials measured in one coordinate frame (working frame, for example CAD frame or other predefined local frame) and $\{Y\}_J$ is a set of corresponding fiducials measured in the second frame (destination frame, for example global or instrument's frame). Pose measuring systems track the movement of the object's frame, and the transformation $\{R, \tau\}$ defines the pose of the object in the instrument's frame. Once the transformation is known, it can be applied to any POI which was not used in the process of registration. Such a point (also called a target point) located at **U** may be used to calculate the Target

[&]quot;Propagation of orientation uncertainty of 3D rigid object to its points." Paper presented at ICCV 2017 International Conference on Computer Vision, Venice, Italy. October 22, 2017 - October 29, 2017.

Registration Error TRE(U) which is defined as the distance between the transformed target and its location in the instrument's frame.

Uncertainty in the locations of fiducials propagates to the transformation $\{\mathbf{R}, \boldsymbol{\tau}\}$ and, further, to the location of the transformed target. Thus, noise perturbing the locations of fiducials affects TRE(U). Finding an analytic expression which relates TRE(U) with the statistical properties of noise perturbing fiducials has been the subject of extensive research for many years [15-19]. The main conclusions from these studies can be summarized as follow: 1) TRE(U) depends on the location of target U relative to the three main axes of the moment of inertia determined from the spatial configuration of fiducials $\{X\}_{I}$; 2) TRE(U) can be expressed as the sum of two components: one related to uncertainty in position τ and the second related to uncertainty in the orientation data R; 3) both components are proportional to the magnitude of noise; 4) the orientation component is anisotropic while the positional component is isotropic.

For the class of pose measuring systems which use pointbased rigid body registration to determine the pose of a rigid object, uncertainty of any POI on that object will exhibit the properties of TRE described above. Thus, the propagation of the object orientation uncertainty to a POI should depend on the location of the fiducials. We verify this by acquiring different poses using a two-camera system and small spherical targets to get 6DOF data.

2. Background

Let column vector **U** define the location of a POI in the CAD coordinate frame and \boldsymbol{u} be a unit vector parallel to \boldsymbol{U} such that $\boldsymbol{U} = \|\boldsymbol{U}\| \boldsymbol{u}$. If \boldsymbol{R}_n is the orientation matrix of a rigid object and t_n its location obtained from the *n*-th noisy measurement, then U_n is the location of the POI on the rotated object in the coordinate frame of the pose measuring instrument

$$\boldsymbol{U}_n = \|\boldsymbol{U}\| \, \boldsymbol{w}_n + \boldsymbol{t}_n \,, \tag{2}$$

where \boldsymbol{w}_n is a unit vector pointing to the rotated POI in the coordinate frame of the instrument

$$\boldsymbol{w}_n = \boldsymbol{R}_n \, \boldsymbol{u} \,. \tag{3}$$

In the rest of this paper, any unit column vector \boldsymbol{a} will be parametrized by two spherical angles as

 $\boldsymbol{a}(\vartheta,\varphi) = [\cos\vartheta\cos\varphi,\cos\vartheta\sin\varphi,\sin\vartheta]^T. \ (4)$ We investigate the propagation of uncertainty of R_n to the uncertainty of w_n . We assume that

$$\boldsymbol{R}_n = \boldsymbol{R}_{avg} \,\Delta \boldsymbol{R}_n \,, \tag{5}$$

where \mathbf{R}_{avg} is the averaged orientation obtained from N repeated measurements, $\Delta \mathbf{R}_n$ is a small random rotation (noise), and n = 1, ..., N. In axis-angle representation (a_n, ρ_n) , the smallness of the random rotation is gauged by small values of angle ρ_n and this leads to the following expression for $\Delta \mathbf{R}_n$ in linear approximation

$$\Delta \mathbf{R}_{n}(\mathbf{a}_{n},\rho_{n}) \approx \mathbf{I} + \begin{bmatrix} 0 & -q_{z,n} & q_{y,n} \\ q_{z,n} & 0 & -q_{x,n} \\ -q_{y,n} & q_{x,n} & 0 \end{bmatrix},$$
(6)

where $I_{3\times 3}$ is the identity matrix, a_n is a unit vector defining the axis of rotation, and

$$q_n = \rho_n a_n$$
. (7)
e covariance matrix $C(a)$ of the orientation data is then

The given by

$$\boldsymbol{C}(\boldsymbol{q}) = \frac{1}{N} [\boldsymbol{q}_1, \dots, \boldsymbol{q}_N] [\boldsymbol{q}_1, \dots, \boldsymbol{q}_N]^T .$$
(8)

Repeated measurements of the orientation matrix \mathbf{R}_n in (3) yield a corresponding set of unit vectors $\{w\}_N$ which are tightly distributed around the average direction w_{avg} . If μ denotes the angle between w_n and w_{ava} , then its distribution can be described by the Fisher-Bingham-Kent (FBK) distribution $G_{\sigma,\beta}(\mu)$ [20-22]. FBK has two parameters: angular uncertainty σ and eccentricity parameter β . Larger values of σ indicate larger deviations of \boldsymbol{w}_n from \boldsymbol{w}_{ava} (i.e., more noisy data). Larger values of β correspond to a more elongated elliptical contour of constant probability on the (ϑ, φ) plane (as $\beta \to 0$, the contour approaches a circle).

If the orientation matrices R_n are obtained by pose measuring systems which derive orientation from a set of measured fiducial markers and use the rigid-body, pointbased registration, then there are equations which provide σ for certain types of noise disturbing the locations of fiducials. For example, for Gaussian, homogenous and anisotropic noise (covariance matrix Ψ is the same at each point in space but the three variances, i.e. eigenvalues of Ψ , substantially differ from each other) an expression for TRE(U) is provided in [23]. To study the component of TRE which depends on direction only, the isotropic part of TRE and the linear dependence on the distance of the POI to the origin are ignored (i.e., vector \boldsymbol{U} is replaced by the unit vector $\boldsymbol{u} = [u_1, u_2, u_3]^T$). Then, the variance σ^2 of the deviation angles μ between instantaneous vectors \boldsymbol{w}_n in (3) and their mean \boldsymbol{w}_{avg} is given by

$$\sigma^{2}(\boldsymbol{u}) = \sum_{i=1}^{3} \sum_{j\neq i}^{3} \left[\frac{u_{j}^{2} (\Lambda_{i}^{2} \Psi_{i,i} + \Lambda_{i}^{2} \Psi_{jj})}{(\Lambda_{i}^{2} + \Lambda_{j}^{2})^{2}} + \sum_{k\neq i,j}^{3} S_{i,j,k} \right], (9)$$

where

$$s_{i,j,k} = u_j u_k \Psi_{j,k} \frac{\Lambda_i^2}{\left(\Lambda_i^2 + \Lambda_j^2\right) \left(\Lambda_i^2 + \Lambda_k^2\right)} \,. \tag{10}$$

Parameters Λ_i can be calculated from the Singular Value Decomposition (SVD) of the matrix $[X]_I = [X_1, ..., X_I]$ where column vector X_i is the location of the *i*-th fiducial in the CAD coordinate frame. Thus,

$$[\boldsymbol{X}]_J = \boldsymbol{U}_X \boldsymbol{\Lambda} \boldsymbol{V}^T , \qquad (11)$$

where the diagonal matrix $\mathbf{\Lambda} = diag(\Lambda_1, \Lambda_2, \Lambda_3)$ and U_X is the rotation matrix which aligns axes of $[X]_{I}$ with the CAD frame axes. The three columns of U_X also define the eigenvectors $[\boldsymbol{m}_{min}, \boldsymbol{m}_{mid}, \boldsymbol{m}_{max}]$ of the moment of inertia

Cheok, Geraldine; Franaszek, Marek. "Propagation of orientation uncertainty of 3D rigid object to its points." Paper presented at ICCV 2017 International Conference on Computer Vision, Venice, Italy. October 22, 2017 - October 29, 2017.

matrix **M** where

$$\boldsymbol{M} = trace(\boldsymbol{covX}) \, \boldsymbol{I} - \boldsymbol{covX} \tag{12}$$

and the covariance matrix $covX_{3\times3} = cov([X]_J^T)$. Eigenvalues $[h_{min}, h_{mid}, h_{max}]$ of M are equal

 $(J-1)[\Lambda_2^2 + \Lambda_3^2, \Lambda_3^2 + \Lambda_1^2, \Lambda_1^2 + \Lambda_2^2]$, (13) where h_{min} corresponds to the largest while h_{max} to the smallest element of Λ , as follows from (12).

Note that in (9) and (10) only elements of the diagonal matrix Λ are used while covariance of noise Ψ may have non-zero off-diagonal elements. This is a consequence of the fact that the orientation of the noise matrix is independent of the orientation of M matrix (i.e., spatial configuration of fiducials). Thus, for any arbitrary orientation of a rigid object, three major axes of the moment of inertia may be aligned with the rotated CAD frame axes (by using rotation U_X) but it will not lead to alignment of noise matrix Ψ with CAD frame.

3. Experiment Description

The 6DOF pose measurements were obtained using an OptiTrack Duo system which is a two-camera, motion capture system [24]. The system locates small spherical markers covered with reflective material in its work volume and finds the centers of these spheres. A group of markers attached to a rigid object forms a configuration of fiducials. When the new configuration is created, the measured locations of the markers at that time define the initial object pose. Thus, the system tracks the motion of the object by tracking these markers. It uses a rigid-body, point-based registration (i.e., it minimizes FRE in (1)) to calculate the current pose of an object relative to its initial pose. As the object moves, some markers are occluded, but the system tracks the visible markers and use them to determine the object pose. The system outputs four components of a unit quaternion and three components of the positional part of the 6DOF pose data. In addition to object pose, xyzlocations of each visible marker are also provided.

In the experiment, spherical markers were glued to a rigid object (a piece of aluminum rail). The total of M = 15 randomly selected poses were measured using two different configurations of markers. For poses m = 1 to 5, the same configuration consisting of J = 5 markers was used. For poses m = 6 to 15, the configuration consisting of J = 7 markers was used. At each *m*-th pose, a large number $N \ge 50,000$ of repeated measurements were obtained.

Figure 1 shows fifteen average measured poses $\mathbf{R}_{avg}(m)$ together with locations of fiducials. For poses m = 1 and m = 6, the new configurations of fiducials are defined and therefore the corresponding orientations $\mathbf{R}_{avg}(m) = \mathbf{I}$. Black dots represent fiducials which were not visible at a given pose and the number of undetected fiducials is provided for each pose. The same fiducials were undetected for pairs of poses: m = (4,5), (9,10), (8,14).

4. Data processing

For each data file containing N repeated measurements, the average rotation \mathbf{R}_{avg} in (5) was calculated first. There are different ways of calculating the average rotation and not all of them are correct. In this study, we used the mean rotation in the Euclidean sense [25]. Once the average rotation was determined, small random rotations $\Delta \mathbf{R}_n$ were calculated from (5), i.e.

$$\Delta \boldsymbol{R}_n = \boldsymbol{R}_{avg}^T \boldsymbol{R}_n \,, \tag{14}$$

where we used the property that the inverse of rotation matrix is equal to the transposed matrix, $\mathbf{R}^{-1} = \mathbf{R}^{T}$. Calculating small rotations from experimental data requires extra caution. The data file contained four components of unit quaternion \boldsymbol{g}_n which can be converted to the rotation matrix $\mathbf{R}(\mathbf{g}_n)$. However, we noticed that due to truncation error in the stored values of \boldsymbol{g}_n (six decimal digits), the resulting matrix was not orthonormal, i.e. $R(g_n) R^T(g_n) \neq I$. This caused the resulting matrix on the left side of (14) to also violate orthonormality and linear approximation in (6) yielding incorrect orientation noise q. To overcome this problem, we determined the rotation matrix \mathbf{R}_n directly by minimizing *FRE* in (1). Specifically, for pose m = 1 when the configuration of five fiducials was defined, we calculated the mean locations of fiducials $\{\overline{X}\}_{I}$ from N repeated measurements. Then, for poses m =2, ..., 5, for each repeated *n*-th measurement of fiducials, the registration between $\{\overline{X}\}_{I(m)}$ and $\{Y\}_{I(m),n}$ was performed and the corresponding rotation matrix \mathbf{R}_n was calculated as in [26]. The matrix determined in this manner satisfied the condition of orthonormality, ensuring that linear approximation in (6) yielded the correct orientation noise **q**. The same procedure was repeated for pose m = 6 when the new configuration with seven fiducials was defined and registration between new $\{\overline{X}\}_{J(m)}$ and $\{Y\}_{J(m),n}$ was performed for poses m = 7, ..., 15. Due to marker occlusions, different poses used different subsets of the fiducials, and the registrations were performed using varying number of fiducials J(m).

Once rotations \mathbf{R}_n , n = 1, ..., N were calculated from the experimental data, a set of unit vectors $\{\mathbf{w}\}_N$ in (3) was formed for a unit vector $\mathbf{u}(\vartheta, \varphi)$ parametrized by two angles as in (4). For each set $\{\mathbf{w}\}_N$, two parameters of FBK distribution, $\sigma(\vartheta, \varphi)$ and $\beta(\vartheta, \varphi)$ were determined as in [21]. The calculations were repeated on a grid of angles $[-90^\circ 90^\circ] \times [0^\circ 360^\circ]$ so that the angular uncertainty σ could be displayed on a unit sphere. In addition, for some selected unit vectors \mathbf{u} , the histograms of deviation angles μ_n between \mathbf{w}_n and \mathbf{w}_{avg} were constructed.

For each *m*-th pose, the average rotation $\mathbf{R}_{avg}(m)$ was evaluated and small random rotations $\Delta \mathbf{R}_n$ were determined as in (14). From each $\Delta \mathbf{R}_n$, the corresponding noisy orientation perturbation \mathbf{q}_n was extracted using (6) and from (7) the axis of small random rotation \mathbf{a}_n was obtained.



Figure 1. Fifteen measured poses $R_{avg}(m)$ where dots represent the fiducial locations. Gray dots are visible fiducials, and black dots are occluded fiducial(s); numbers in parenthesis provide the total number of occluded fiducials. Configurations of fiducials for rigid-body registration were defined in poses m = (1,6) and, therefore, the corresponding rotations $R_{avg} = I$.

The axis is a unit vector which can be parametrized as in (4) so a histogram of axes $\boldsymbol{a}_n(\vartheta_n, \varphi_n)$ could be displayed on a unit sphere using a grid of angles, similar to uncertainty $\sigma(\vartheta, \varphi)$. Once all N random orientation perturbations $\{q\}_N$ were calculated for the *m*-th pose, their covariance matrix C(q) was determined as in (8). Then, its eigenvalues $[l_{min}, l_{mid}, l_{max}]$ and the corresponding eigenvectors $[\boldsymbol{c}_{min}, \boldsymbol{c}_{mid}, \boldsymbol{c}_{max}]$ were calculated.

For pose m = 1, SVD of the full configuration of J = 5average locations of fiducials $[\overline{X}]_I$ was performed as in (11). Then, for poses m = 2, ..., 5, the corresponding subset of fiducials detected by cameras was identified and SVD of that subset $[\overline{X}]_{I(m)}$ was performed. The same procedure was repeated for pose m = 6 (where the full configuration consisted of I = 7 fiducials) and the remaining poses m =7, ..., 15. Thus, for each *m*-th pose, the corresponding diagonal matrix $\Lambda(m)$ and orthonormal matrix $U_{\chi}(m)$ in (11) was determined. Similar calculations could be done for the moment of inertia matrix M(m), yielding equivalent results as follows from (12) and (13). We note that eigenvectors are defined in the object's local frame and in the instrument's coordinate frame they follow exactly the *m*-th orientation of the object $\mathbf{R}_{avg}(m)$ if the same set of fiducials is detected in each m-th pose. Thus, for the instruments which cannot detect all fiducials for every pose, the moment of inertia and its eigenvectors will be varying.

To study the consequences of misalignment between the moment of inertia matrix and the covariance matrix of noisy fiducials, equations (9) and (10) were used for a fixed configuration of fiducials (fixed Λ in (11)) and different noise covariance matrices Ψ . Specifically, noise matrix was

$$\Psi(\omega) = \mathbf{R}(\mathbf{a}, \omega) \Psi_0 \mathbf{R}(\mathbf{a}, -\omega)$$
(15)

where Ψ_0 was a diagonal matrix and $R(a, \omega)$ was a rotation matrix in the (axis, angle) representation (we used the property that $\mathbf{R}^{-1}(\mathbf{a}, \omega) = \mathbf{R}(\mathbf{a}, -\omega)$). For $\omega = 0$, matrix $\Psi = \Psi_0$ and all off-diagonal elements $\Psi_{i,k} = 0$ in (10). This case corresponds to rather exceptional configuration in which the object is rotated in such manner that the axes of moment of inertia are aligned with the eigenvectors of the covariance matrix of the noise. As angle ω is gradually increased (i.e., the object is gradually turned away from the exceptional configuration) absolute values of the offdiagonal elements $|\Psi_{i,k}|$ also increase. Then, depending on the values of the elements of the diagonal matrices Λ and Ψ_0 , the locations of the extreme points of angular uncertainty $\sigma(\vartheta, \varphi)$ may move on a unit sphere (recall that a unit vector \boldsymbol{u} in (9) can be parametrized by two angles as in (4)). Note that for perfectly isotropic noise, i.e. $\Psi_0 = \psi I$, off-diagonal elements $\Psi_{j,k}$ will always be zero, regardless of object orientation.

5. Results

Figure 2 shows the directions associated with the largest and the smallest principal moment of inertia M. For poses m = 6, ..., 15, directions associated with the smallest moment of inertia (shown as solid, black lines) form four groups. For five of the poses, the directions overlap almost

exactly; for the remaining five poses, four have three occluded fiducials (poses m = 8,11,14,15). All lines are plotted in the coordinate frame associated with the first pose m = 1.



Figure 2. Directions associated with the smallest principal moment of inertia M (eigenvector m_{min} , solid lines) and the largest one (eigenvector m_{max} , dashed lines). Blue color marks poses m = 1, ..., 5 while black marks poses m = 6, ..., 15.

Figure 3 shows the histograms of the orientation noise qdefined in (7) obtained from data acquired for pose m = 5. Similar distributions were observed for the other poses. Figure 4 shows the histograms of deviation angles μ_n between the mean w_{avg} and noisy rotated unit vectors w_n defined in (3). The two histograms correspond to two different choices of vector \boldsymbol{u} in (3) which is rotated by noisy rotations R_n : one is aligned with the direction in which propagated uncertainty σ is the largest (i.e., c_{min}), the other is aligned with the direction where the propagated uncertainty is the smallest (c_{max}). In addition, for each set of rotated noisy vectors $\{w\}_N$, two parameters of FBK distribution (i.e., orientation uncertainty σ and eccentricity β) are calculated. The corresponding values are $\sigma =$ (0.161, 0.019) [mrad] and β = (15.4, 1334.2) [mrad⁻²] for the direction of the largest and the smallest propagated uncertainty, respectively. Once both parameters σ and β are known, the theoretical distribution FBK of the deviation angles μ can be plotted. The experimental histograms and theoretical FBK distributions displayed in Figure 4 were obtained from data acquired in pose m = 10, corresponding plots obtained for other poses are similar.

Figure 5 shows the histograms of the spherical angles (ϑ_n, φ_n) parametrizing (as in (4)) the rotated noisy unit vectors \boldsymbol{w}_n defined in (3) which were used to create the plots in Figure 4. The histograms are centered at angles $(\vartheta_{avg}, \varphi_{avg})$ which parametrize the mean unit vector \boldsymbol{w}_{ava} , so $\Delta \vartheta_n = \vartheta_n - \vartheta_{avg}$ and $\Delta \varphi_n = \varphi_n - \varphi_{avg}$.

In Figure 6 histograms of axes a_n of small random rotations $\Delta \mathbf{R}_n(\mathbf{a}_n, \rho_n)$ defined in (6) are shown. In addition, the eigenvectors corresponding to the largest and the smallest eigenvalue of the covariance matrix of orientation noise C(q) given by (8) and the moment of inertia matrix

M defined in (12) are plotted. Histograms were obtained from data acquired in poses m = (8, 14). For that poses, the same subset of J = 5 fiducials was detected and used to perform rigid-body registration. Both graphs are plotted from the same view angle in the object's coordinate frame associated with the full configuration of seven fiducials (pose m = 6).



Figure 3. Histogram of orientation noise $q = [q_{x_1} q_{y_2} q_z]$ calculated from small random rotations $\Delta R_n(q)$ in (6) and (14).



Figure 4. Histograms of the angles of deviation μ (markers) and the corresponding theoretical FBK distributions (solid lines) for two different directions of vector \boldsymbol{u} in (3): a) \boldsymbol{u} aligned with the eigenvector \boldsymbol{c}_{min} of the covariance matrix $\boldsymbol{C}(\boldsymbol{q})$; b) \boldsymbol{u} aligned with c_{max} .

In Figure 7, the angular dependence of the orientation uncertainty $\sigma(\vartheta, \varphi)$ is shown for the same poses as in Figure 6, i.e. m = (8, 14). Eigenvectors corresponding to the largest and the smallest eigenvalue of covariance matrix C(q) are plotted for each data set (black solid and dashed lines, respectively) and eigenvectors corresponding to the smallest and the largest eigenvalue of the moment of inertia matrix **M** are plotted (grey solid and dashed lines, respectively). The plots in Figure 7 are displayed in the

Cheok, Geraldine; Franaszek, Marek. "Propagation of orientation uncertainty of 3D rigid object to its points." Paper presented at ICCV 2017 International Conference on Computer Vision, Venice, Italy. October 22, 2017 - October 29, 2017.

same coordinate frame and from the same view angle as the plots in Figure 6, so the directions of the axes in both figures are the same.



Figure 5. Histograms (plotted in log scale) of the spherical angles $(\boldsymbol{\vartheta}_n, \boldsymbol{\varphi}_n)$ parametrizing the same unit vectors \boldsymbol{w}_n which were used to create the plots in Figure 4: a) \boldsymbol{u} aligned with the eigenvector c_{min} ; b) u aligned with c_{max} . -inf indicates empty bins.



Figure 6. Histograms of axes $a_n(\vartheta_n, \varphi_n)$ of small random rotations (plotted in log scale) for pose: a) m = 8; b) m = 14. Black lines show directions of eigenvectors of covariance matrix C(q): solid lines correspond to c_{max} and dashed line to c_{min} . Blue lines indicate eigenvectors of the moment of inertia M: solid lines correspond to m_{min} while dashed lines correspond to m_{max} . -inf indicates empty bins, bin size [1° x 2°].

Plots similar to Figure 7 are shown in Figure 8 for data acquired in poses m = (4, 9). However, the graph presented in a) is displayed in the coordinate frame associated with the full configuration of I = 5 fiducials defined in pose m = 1 while the graph in b) in coordinate frame associated with the full configuration of J = 7fiducials defined in pose m = 6. Also, different view angles are used to ensure a clear view of relevant features (i.e., misalignment of solid black and grey lines).



Figure 7. Angular distribution of orientation uncertainty $\sigma(\vartheta, \varphi)$ in [mrad] for data acquired in pose: a) m = 8; b) m = 14. Black lines show directions of eigenvectors of C(q): solid lines correspond to c_{max} and dashed lines to c_{min} . Grey lines indicate eigenvectors of M: solid lines correspond to m_{min} while dashed lines correspond to m_{max} .



Figure 8. Angular distribution of orientation uncertainty $\sigma(\vartheta, \varphi)$ in [mrad] for data acquired in pose: a) m = 4; b) m = 9. Black lines show directions of eigenvectors of C(q): solid lines correspond to c_{max} and dashed lines to c_{min} . Grey lines indicate eigenvectors of M: solid lines correspond to m_{min} while dashed lines correspond to m_{max} .

In Figure 9, misalignment angles between eigenvectors of the covariance matrix C(q) and the corresponding eigenvectors of the moment of inertia matrix \boldsymbol{M} are plotted for all poses excluding poses for m = (1, 6) which were used as reference poses. In Figure 9a, α_{max} is the angle between the eigenvector c_{max} of C(q) and the eigenvector m_{min} of **M**. In Figure 9b, α_{min} is the angle between the eigenvector c_{min} of C(q) and one of the eigenvectors (m_{mid}, m_{max}) whichever is closer to c_{min} . In Figure 9c, α_{mid} is the angle between the eigenvector c_{mid} of C(q) and the remaining third eigenvector of the matrix M.

Noise characteristics of the measured fiducial locations were obtained by calculating the covariance matrix for each

Cheok, Geraldine; Franaszek, Marek. "Propagation of orientation uncertainty of 3D rigid object to its points." Paper presented at ICCV 2017 International Conference on Computer Vision, Venice, Italy. October 22, 2017 - October 29, 2017.

visible fiducial in each pose (total of 72 locations). Eigenvalues $\lambda = [\lambda_{min}^2, \lambda_{mid}^2, \lambda_{max}^2]$ of each matrix were calculated and the magnitude of the noise $\kappa = \sqrt{trace(\lambda)}$ and the measure of noise anisotropy $v = \lambda_{max}/\lambda_{min}$ were determined. The median and the largest magnitude were $\kappa_{med} = 0.0375$ [mm] and $\kappa_{max} = 0.1726$ [mm]. For anisotropy, the corresponding characteristics were v_{med} = 10.4 and $v_{max} = 18.7$.



Figure 9. Misalignment angles between the eigenvectors of the covariance matrix C(q) and the eigenvectors of the moment of inertia matrix M for different poses; see text for definitions of three displayed angles (shown here in degrees).

6. Discussion

Repeated N measurements of object pose yield bell shaped, Gaussian like distribution of orientation noise q, as shown in Figure 3. The resulting uncertainty in the object orientation propagates to a POI on the object. This propagation appears to be anisotropic, dependent on the unit vector $\boldsymbol{u}(\vartheta, \varphi)$ pointing in the direction of the POI in the object's local coordinate frame (CAD frame). In extreme cases, when \boldsymbol{u} is aligned with the eigenvectors corresponding to the largest or the smallest eigenvalue of the covariance matrix of orientation noise C(q), the resulting uncertainties of the POI may be vastly different, as seen in Figure 4 and Figure 5. Such characteristics of pose measuring systems may impact processes which rely on 6DOF data acquired by these systems, for example part assembly in manufacturing. For systems which derive pose from raw measurements of fiducial markers attached to a rigid object, existing theoretical results provide a link between the properties of the covariance matrix C(q), propagated orientation uncertainty $\sigma(\vartheta, \varphi)$ and the placement of markers on or around the measured rigid object. The spatial configuration of fiducials is

characterized by its moment of inertia M and the eigenvectors of C(q) and M are parallel for isotropic noise, i.e., $\Psi_{j,k} = 0$ in (10).

In our experiments, two different configurations of fiducials were defined: in pose m = 1 and pose m = 6. These two configurations yielded major axes of the moment of inertia M rotated relative to each other by roughly 90° around Y axis, see blue and black solid lines shown in Figure 2 (this can be understood by looking at actual placement of fiducials in Figure 1 for m = 1 and m = 6). However, different definitions of the full configuration of fiducials are not the only reason for the observed variability in the orientation of the major axes of M. The instrument which we used in our experiment cannot detect all fiducials in every measured pose (see Figure 1). This implies that even for the same nominal configuration of fiducials different moments of inertia could be observed in different poses as different subsets of fiducials are used for rigidbody registration. Thus, the location of a POI relative to the major axes of **M** depends indirectly on object orientation as does the uncertainty of that POI propagated from the object pose uncertainty. We recall that covariance matrix C(q) is calculated from the set of vectors $\{q\}_N$ and each q_n is parallel to its respective axis a_n of small random rotation $\Delta \mathbf{R}_n(\mathbf{a}_n, \rho_n)$, as follows from (7). Thus, the eigenvector c_{max} of C(q) defines the direction with the highest concentration of axes a_n , i.e. the spot on a unit sphere where histogram of axes a_n reaches its maximum, see Figure 6. Similarly, c_{min} defines the direction in which the concentration of axes a_n is minimum. Furthermore, any rotation has the largest impact on vectors perpendicular to its axis of rotation and has no impact on vectors parallel to the axis. Therefore, the uncertainty σ (which accounts for the spread of vectors \boldsymbol{w}_n around their mean \boldsymbol{w}_{avg} in (3)) is the smallest in the direction defined by c_{max} and the largest in the direction aligned with \boldsymbol{c}_{min} , see Figure 7 and Figure 8. Thus, the eigenvectors of C(q) depend on the object orientation and so do $\sigma_{min}(\boldsymbol{c}_{max})$ and $\sigma_{max}(\boldsymbol{c}_{min})$. If the configuration of the fiducials does not change, the moment of inertia M is fixed and variations of c_{max} in the local object's frame would be small.

However, the variability in the directions of the major axes of **M** only partially explains why eigenvectors of C(q)depend on the object orientation. For example, the misalignment angle α_{max} between c_{max} and m_{min} is small for pose m = 8 in Figure 7a and large for pose m = 14 in Figure 7b (solid black and grey lines) although the same subset of fiducials was used to perform registration for both poses. Furthermore, for pose m = 4 in Figure 8a and pose m = 9 in Figure 8b both misalignments between c_{max} and m_{min} are small, although different subsets of fiducials were used for registration, see Figure 1. As shown in Figure 9, for most poses (except for poses m = 11, 13, 14) angle $\alpha_{max} < 15^{\circ}$ while the other two misalignment angles α_{mid}

and α_{min} are larger. Thus, another factor must be responsible for the observed dependence of C(q) on object pose. We note that noise perturbing fiducials is strongly anisotropic, i.e., $\Psi_{i,k} \neq 0$ in (10). This means that for different object orientations in the instrument frame, different degrees of misalignment between the covariance of fiducial noise Ψ and matrix **M** are expected. This impacts uncertainty σ as shown in Figure 10 where theoretical model (9) of σ for homogenous and anisotropic noise affecting fiducials is plotted for increasing misalignment angle ω in (15). (Arbitrary axis of rotation **a** in (15) was used and diagonal noise matrix $\Psi_0 = diag(13, 17, 1080)$ $[\mu m^2]$, diagonal matrix Λ in (9) equal to diag(160, 81, 17)[mm].) Experimental noise in fiducials is strongly nonhomogeneous and therefore the theoretical σ does not exactly replicate experimental σ (compare Figure 10 with Figure 7 and Figure 8). Nevertheless, the angle α_{min} between the dashed black and grey lines in Figure 10 clearly increases with the increasing angle ω (i.e., increasing absolute values of off-diagonal elements of noise matrix $|\Psi_{j,k}|$ in (10)).



Figure 10. Theoretical uncertainty $\sigma(\vartheta, \varphi)$ in [mrad] from (9) for different noise covariance matrix $\Psi(\omega)$ from (15): a) $\omega = 5^{\circ}$; b) $\omega = 40^{\circ}$; c) $\omega = 50^{\circ}$; d) $\omega = 110^{\circ}$. Black lines show directions where $\boldsymbol{\sigma}$ is the smallest (solid lines) and the largest (dashed lines). Grey solid lines are aligned with m_{min} and grey dashed lines with m_{max} .

In summary, the direction along which the propagated uncertainty is the smallest corresponds to the eigenvector associated with the largest eigenvalue of the covariance matrix of orientation data. For the class of instruments investigated in this study, this direction is closely aligned to

the direction of the axis corresponding to the smallest moment of inertia, for most of the measured object orientations. Even for poses m = (11, 13, 14) where the misalignment angle α_{max} is large, as shown in Figure 9a, m_{min} points in the direction where the propagated to POI uncertainty σ is still small, see Figure 7b. This feature has a practical consequence: if a given POI is important, it is possible to attach fiducial markers to the rigid object so that the resulting axis of the smallest moment of inertia is aligned with the direction of the vector pointing to the POI. This will ensure that the POI will be in the area of the smallest possible uncertainty. If the propagated uncertainty needs to be minimized for different POIs on the same object, the system may dynamically select different subsets of fiducials to ensure that each POI is paired with the appropriately oriented m_{min} . The system in this study does dynamically select different subsets of fiducials for different object poses because of occlusions of some of the markers. However, the operator cannot control this selection and optimal choice of fiducials for multiple POIs is not possible. Our study also shows that the direction along which the propagated uncertainty is the largest, corresponds to the eigenvector associated with the smallest eigenvalue of the covariance matrix of orientation data. However, this eigenvector is not aligned with the axis of the largest moment of inertia, i.e. the angle between the eigenvector and the axis of moment of inertia varies greatly for different object orientations. The large variations of angle can be attributed to the misalignment between the moment of inertia axes and the eigenvectors of the covariance matrix of fiducial noise.

7. Conclusions

For most manufacturing applications, not all points on an object are of equal interest. If a given point is important, then the uncertainty propagated to that point should be minimized. For pose measuring systems that determine pose by measuring fiducials, this can be achieved by strategic placement of fiducials so that the moment of inertia axis, m_{min} , is aligned with a vector pointing to that point. We expect systems that are trained to extract points on object will show similar results. Other systems which use different methods for pose determination (e.g., fitting point cloud to CAD model) will require further investigation.

References

- [1] X. Jiang, A. Konno, and M. Uchiyama, "A Vision-Based Endpoint Trajectory and Vibration Control for Flexible Manipulators," Int. Conference on Robotics and Automation ICRA, Roma, 2007.
- [2] L. M. Brignone and M. Howarth, "A Geometrically Validated Approach to Autonomous Robotic Assembly,"

Cheok, Geraldine; Franaszek, Marek. "Propagation of orientation uncertainty of 3D rigid object to its points." Paper presented at ICCV 2017 International Conference on Computer Vision, Venice, Italy. October 22, 2017 - October 29, 2017.

IEEE/RSJ Int. Conference on Intelligent Robots and Systems IROS, 2002.

- [3] M. Prats, P. Martinet, A. P. Pobil, and S. Lee, "Robotic execution of everyday tasks by means of external vision/force control," *Inteligent Service Robotics*, vol. 1, no. 3, pp. 253-266, 2008.
- [4] S. Huang, Y. Yamakawa, T. Senoo, and M. Ishikawa, "A Direct Visual Servo Scheme Based on Simplified Interaction Matrix for High-Speed Manipulation," Int. Conference on Robotics and Biomimetics, Guangzhou, China, 2012.
- [5] H. Park, J. H. Bae, J. H. Park, M. H. Baeg, and J. Park, "Intuitive Peg-in-Hole Assembly Strategy with a Compliant Manipulator," 44th International Symposium on Robotics, 2013.
- [6] H. Qiao and S. K. Tso, "Three-step precise robotic peg-hole insertion operation with symmetric regular polyhedral objects," *Int. J. Production Reserach*, vol. 37, no. 15, pp. 3541-3563, 1999.
- [7] Z. Liu, Y. Xie, J. Xu, and K. Chen, "Laser tracker based robotic assembly system for large scale peg-hole parts," Int. Conference on Cyber Technology in Automation, Control and Intelligent Systems, Hong Kong, China, 2014.
- [8] D. I. Park, C.Park, H. Do, T. Choi, and J. H. Kyung, "Assembly phase estimation in the square peg assembly process," Int. Conference on Control, Automation and Systems, Jeju, Korea, 2012.
- [9] S. Chhatpar and M. S. Branicky, "Localization for Robotic Assemblies with Position Uncertainty," IEEE/RSJ Int. Conference on Intelligent Robots and Systems IROS, 2003.
- [10] S. Chhatpar and M. S. Branicky, "Search Strategies for Pegin-Hole Assemblies with Position Uncertainty," IEEE/RSJ Int. Conference on Intelligent Robots and Systems IROS, 2001.
- [11] J. Y. Kim, W. S. Kim, and H. S. Cho, "Misalignment estimation and compensation for robotic assembly with uncertainty," *Robotica*, vol. 23, no. 03, pp. 355-364, 2005.
- [12] R. Usubamatov and K. W. Leong, "Analyses of peg-hole jamming in automatic assembly machines," *Assembly Automation*, vol. 31, no. 4, pp. 358-362, 2011.
- [13] R. Usubamatov, S. A. Adam, and A. Harun, "Analyzing the jamming of parts on the shaft in assembly processes," *Assembly Automation*, vol. 32, no. 4, pp. 340-346, 2012.
- [14] M. Franaszek, M. Shah, G. S. Cheok, and K. S. Saidi, "The Axes of Random Infinitesimal Rotations and the Propagation of Orientation Uncertainty," *Measurement*, vol. 72, pp. 68-76, 2015.
- [15] J. M. Fitzpatrick, J. B. West, and C. R. Maurer, "Predicting Error in Rigid-Body Point-Based Registration," *IEEE Trans. Medical Imaging*, vol. 17, no. 5, pp. 694-702, 1998.
- [16] R. R. Shamir, L. Joskowicz, and Y. Shoshan, "Fiducial Optimization for Minimal Target Registration Error in Image-Guided Neurosurgery," *IEEE Trans. Medical Imaging*, vol. 31, no. 3, pp. 725-737, 2012.
- [17] J. B. West, J. M. Fitzpatrick, S. A. Toms, C. R. Maurer, and R. J. Maciunas, "Fiducial Point Placement and the Accuracy of Point-based, Rigid Body Registration," *Neurosurgery*, vol. 48, no. 4, pp. 810-817, 2001.
- [18] A. Seginer, "Rigid-body point-based registration: The distribution of the target registration error when the fiducial registration errors are given," *Medical Image Analysis*, vol. 15, pp. 397-413, 2011.

- [19] M. H. Moghari and P. Abolmaesumi, "Distribution of Target Registration Error for Anisotropic and Inhomogeneous Fiducial Localization Error," *IEEE Trans. Medical Imaging*, vol. 28, no. 6, pp. 799-813, 2009.
- [20] G. S. Watson, "Distributions on the Circle and Sphere," J. Applied Probability, vol. 19, pp. 265-280, 1982.
- [21] J. T. Kent, "The Fisher-Bingham Distribution on the Sphere," J. Royal Statist. Soc. B, vol. 44, no. 1, pp. 71-80, 1982.
- [22] C. Bingham, "An antipodally symmetric distribution on the sphere," *The Annals of Statistics*, vol. 2, no. 6, pp. 1201-1225, 1974.
- [23] A. D. Wiles, A. Likholyot, D. D. Frantz, and T. M. Peters, "A Statistical Model for Point-Based Target Registration Error With Anisotropic Fiducial Localizer Error," *IEEE Trans. Medical Imaging*, vol. 27, no. 3, pp. 378-390, 2008.
- [24] Disclaimer: certain trade names and company products are mentioned in the text or identified in an illustration in order to adequately specify the experimental procedure and equipment used. In no case does such an identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products are necessarily the best available for the purpose.
- [25] M. Moakher, "Means and Averaging in the Group of Rotations," SIAM J.Matrix Analysis and Applications, vol. 24, no. 1, pp. 1-16, 2002.
- [26] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets," *IEEE Trans. PAMI*, vol. 9, no. 5, pp. 698-700, 1987.

Sensors for Safe, Collaborative Robots in Smart Manufacturing

Jeremy A. Marvel, Member IEEE Engineering Laboratory, Intelligent Systems Division U.S. National Institute of Standards and Technology Gaithersburg, MD, USA jeremy.marvel@nist.gov

Abstract—The U.S. National Institute of Standards and Technology (NIST) is developing performance metrics for collaborative robotic systems for smart manufacturing applications. Using a suite of sensor platforms, feedback mechanisms, and novel test artifacts, NIST seeks to provide industry with the means to characterize the performance of robots working collaboratively with humans and other robots. This report serves to outline the applications, requirements, and developmental opportunities for the sensors being used toward this effort. Sensing for safety, human-robot interaction, human-machine interfaces, and multirobot coordination performance are discussed.

Keywords: human-robot interaction, collaborative robots, robot safety, multi-robot coordination, human-robot interaction

I. INTRODUCTION

The U.S. National Institute of Standards and Technology (NIST) is developing new measurement science for the assessment and assurance of the performance of manufacturing robotics by means of its Robotic Systems for Smart Manufacturing (RSSM) program [1]. One focus of the RSSM program is the evaluation of robot systems working with people and other robots [2]. The measurement science being produced includes test methods, metrics, and artifacts for the technology-agnostic verification and validation (V&V) of collaborative robots, their applications, and their behaviors.

For any form of collaboration to be possible, robots must be aware of their surroundings, including the people, tools, and equipment with which they are working. To support this, our work necessitates we must develop and employ novel sensing platforms and techniques to measure the performance of safe, collaborative human-robot interaction (HRI). These sensing platforms are integrated into the collaborative robot testbeds at NIST. The NIST testbeds consist of several collaborative robot platforms including:

- Five six degrees of freedom (6DoF) robotic arms,
- One 7DoF robotic arm,
- Two dual-arm 7DoF (14DoF total) robots, and
- Two mobile robot platforms.

The testbeds are complemented by a suite of sensor platforms used to monitor, measure, and provide feedback to these robots and their human operators. These sensors include multicamera motion capture systems, stereoscopic cameras, commercial red-green-blue-depth (RGB-D) cameras, force/torque (F/T) transducers, and small-scale inertial measurement units (IMUs). These platforms are focused on the identification, localization, and tracking of humans, robots, and parts moving throughout a simulated flexible factory environment. This paper outlines the sensor-enabled areas of research in collaborative robotics at NIST. Section II describes efforts in collaborative robot safety. Section III describes efforts using sensors integrated into the coordination and control of heterogeneous configurations of multiple robots. And Section IV discusses research on sensor-based, intuitive HRI. Throughout this paper, challenges and opportunities for continued sensor development are outlined, as are recommendations for standardized test methods for sensing V&V.

II. COLLABORATIVE ROBOT SAFETY

In 2016, the International Organization of Standardization (ISO) published Technical Specification (TS) 15066, which outlines the safety requirements for collaborative industrial robot systems [3]. ISO/TS 15066 established four safety-related functions of collaborative robots, the two most notable being "speed and separation monitoring" (SSM; maintaining a safe separation distance between a human and an active robot), and "power and force limiting" (PFL; limiting the robot's transfer of pressures and forces onto the human body). Assuring the functionality of both is a challenge, primarily because ISO/TS 15066 currently lacks test methods for the V&V of these functions. In this section, the efforts at NIST to address this shortcoming are discussed.

A. SSM: Presence Detection and Localization

The use of SSM for collaborative operations is intended to maintain a safe distance between the human and the robot. When triggered, the robot must come to a controlled stop prior to making contact with a person by taking into account the total distance traveled by both. The instantaneous separation distance, S, at time t_0 is dictated by the equation

$$S(t_{0}) \geq \int_{\tau=t_{0}+T_{R}+T_{S}}^{\tau=t_{0}+T_{R}+T_{S}} v_{H}(\tau) d\tau + \int_{\tau=t_{0}+T_{R}}^{\tau=t_{0}+T_{R}} v_{R}(\tau) d\tau + , \qquad (1)$$

$$\int_{\tau=t_{0}+T_{R}}^{\tau=t_{0}+T_{R}+T_{S}} v_{S}(\tau) d\tau + (C+Z_{S}+Z_{R})$$

where v_H and v_R are the velocities of the human and robot, respectively, and v_S is the directed velocity of the robot along the stopping path. The variables T_R and T_S capture the time for the robots to respond to the presence of a human, and the time required to stop. System uncertainty is captured by 1) C,



Fig. 1. The full-body model of pressure and force limits from [3].

an intrusion distance safety margin based on statistically expected human reach [4], 2) Z_S , the sensing uncertainty of the safety system, and 3) Z_R , the robot positioning uncertainty. A detailed analysis of the SSM algorithm and recommendations regarding its application are presented in [5].

NIST has focused on the modeling of uncertainty of sensor systems that measure human and robot positions [6], [7], and their integration into the robots' controllers [8], [9]. In a related study on the gaps in the standards for mobile, industrial robot safety [10], a general concern was raised with regards to the state of safety-rated sensors for human detection. Sensors must adhere to specific performance criteria, including mechanisms for the V&V of sensing performance. However, the artifacts specified in the associated standards are not biomimetic. Also, the sensors detect only the presence of objects that have similar sizes of parts of the human body, but not human-specific.

NIST is developing sensor systems designed specifically to identify and localize humans in flexible factory environments. The sensor platform combines cameras that operate in multiple spectra (specifically, RGB and near-infrared), calibrated together for stereoscopic sensing. The cameras are segmented separately to compensate for the limitations of one another (e.g., using thermal to provide information in low-contrast/light color scenes, and using color to correct for thermally noisy scenes). A key challenge in configuring this system, however, is the calibration of the thermal-RGB pair. A traceable-calibration and performance-verification methodology based on [11] is being developed.

B. PFL: Dynamic Force Metrology

With the expectation that physical contact will be made between a human and an active robot, PFL is intended to ensure that the transfer of pressures and forces do not result in injury. ISO/TS 15066 provides specifications for the shape and surface area of robots and tooling to minimize risk, and includes a full body model (Fig. 1) of pressure and force limits based on the onset of pain. In early drafts of the TS, specifications for a sensor system for the V&V of PFL were based on an earlier technical report [12]. NIST provided an analysis of these specifications [13], and called into question their feasibility, in particular with regards to sensor calibration for the measurement of dynamic forces. This issue was later addressed in an unrelated effort at NIST [14].



Fig. 2. The poses of robots and parts are measured directly using physical platforms (left) and markers tracked via motion-capture systems (right)

Today, collaborative robot systems do not provide sufficient pressure/force sensing at the points of contact. Most robot systems rely on joint-level current monitoring paired with gravity models to infer contact forces. As the mass of the robot increases, so too does the inherent noise in the force inference. Moreover, most tooling and onboard support equipment (e.g., cables and hoses) provide no sensing at all. As such, sensors that are reliable, lightweight, easily-integrated, and low-noise are expected to dramatically improve PFL performance. On these lines, efforts in NIST's Performance Assessment Framework for Robotic Systems project [15] seek to produce new test methods for assuring PFL performance. And development of bio-simulant artifacts [16] are expected to positively impact the accessibility and quality of *in situ* PFL V&V.

III. MULTI-ROBOT COORDINATION AND CONTROL

Being industry-focused, the robot-robot collaboration efforts at NIST are targeted toward enabling heterogeneous, multirobot configurations to complete complex assembly tasks. Some robot systems take a negative performance hit in terms of accuracy and repeatability due to mechanical designs [17]. Nominal versus actual positions and orientations may reflect significant errors, so it is necessary to track the motions of the robots to accommodate these errors. Outside of high-cost laser systems used for one-off evaluations of position and path accuracy and repeatability [18], external sensor systems that monitor robot pose are ad hoc solutions with specific applications. Such systems not generally intended for coordinating robots, so multi-robot cells are reliant on direct communication of potentially noisy information for synchronization.

To compensate for issues preventing direct inter-robot communication in heterogeneous cells, NIST leverages external observer systems to 1) identify and track parts and robots, and 2) measure position and orientation uncertainty. These test methods directly influenced the work on evaluating and providing guidance toward improving multi-robot registration [17], [19] using a variety of sensor platforms, including motion capture systems (Fig. 2) and F/T transducers.

NIST's ongoing work is focused on developing new, active, metrology artifacts and sensor platforms for tracking robots and shared workpieces alike (Fig. 2). Moreover, NIST is investigating mobile industrial robot (robotic arms mounted on mobile platforms) coordination, and its impact on safety [10] and process performance. A significant component of this includes issues regarding robot and sensor calibration and registration [20], [21].

IV. INTUITIVE HRI

Emerging work at NIST is directed toward the advancement of HRI and human-machine interfaces. Using sensing to inform the robot systems of human-in-the-loop process performance, human attention, and human intention, NIST aims to develop new test methods to assess and assure the effectiveness of user interfaces and user experiences. Toward these efforts, both intrinsic and extrinsic robot and human sensing are leveraged, borrowing from the systems and results discussed in Sections II and III. These measurements are then processed and fed back to operators and robots via nonstandard means to optimize situation awareness and reduce negative impacts on the team and process.

Small-scale and wearable technologies, in particular, are targeted as potential inputs to the human-robot-team observers. Wireless 9DoF IMUs, simple tilt sensors, light detectors, electromyographic inputs, and time-of-flight distance sensors are integrated into wearable sensor platforms (e.g., protective equipment), shared workpieces, and robot-mounted tools. Consumer products such as smart watches, head-mounted displays, and haptic gloves are also used as alternative interfaces to the robot systems. With the assumption of noisy or missing measurements, these ancillary platforms are then networked together, and the data fused to improve both the quality and the quantity of information to the collaborative team.

Since the provision and maintenance of situation awareness is a motivating concern for these efforts, the presentation of reliable and understandable information is the primary focus. Non-standard interfaces such as augmented and virtual reality displays paired with real-time digital models and 360° video naturally fit into this effort. These interfaces provide intuitive mechanisms for both feedback and control, enabling two-way communications regarding robot and operator performance, attention, and intent. This then propagates naturally into state representations of the process and the human-robot team, in turn improving performance and safety [22].

V. DISCUSSION

This report briefly discusses the ongoing research at NIST involving the novel application of sensors for safe, collaborative HRI. Many challenges and opportunities for the ongoing development of applied sensor systems were briefly described. Worth reiterating here is the need for providing thorough V&V methodologies to such sensor systems. "Hardening" systems to ensure robustness in harsh manufacturing environments is important, but ensuring that the uncertainties and limitations of sensor systems are known is critical when integrating prototype or one-off systems into human-occupied workcells.

DISCLAIMER

Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

REFERENCES

- E. Messina. (2017) Robotic systems for smart manufacturing program. [Online]. Available: "https://www.nist.gov/programs-projects/roboticsystems-smart-manufacturing-program"
- [2] J. Marvel. (2017) Performance of collaborative robot systems. [Online]. Available: "https://www.nist.gov/programs-projects/performancecollaborative-robot-systems"
- [3] "ISO/TS 15066 Robotics and robotic devices Collaborative robots," International Organization for Standardization (ISO), Standard, 2016.
- [4] "ISO 13855 Safety of Machinery Position of Safeguards with Resepcet to the Approach Speeds of Parts of the Human Body," ISO, Standard, 2010.
- [5] J. A. Marvel and R. Norcross, "Implementing speed and separation monitoring in collaborative robot workcells," *Rob. Comput.-Integ. Manuf.*, vol. 44, pp. 144–155, 2017.
- [6] W. Shackleford, R. Norcross, J. Marvel, and S. Szabo, "Integrating occlusion monitoring into human tracking for robot speed and separation monitoring," *Proc. Worksh. Perf. Metrics Intel. Syst.*, pp. 168–173, 2012.
- [7] J. A. Marvel and R. Bostelman, "A cross-domain survey of metrics for modelling and evaluating collisions," *Int. J. Adv. Rob. Syst.*, vol. 11, no. 9, p. 142, 2014.
- [8] J. A. Marvel, "Performance metrics of speed and separation monitoring in shared workspaces," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 2, pp. 405–414, 2013.
- [9] W. Shackleford, G. Cheok, T. Hong, K. Saidi, and M. Shneier, "Performance evaluation of human detection systems for robot safety," *J. Intell. Rob. Syst.*, vol. 83, no. 1, pp. 85–103, 2016.
- [10] J. Marvel and R. Bostelman, "Towards mobile manipulator safety standards," Proc. IEEE Int. Symp. Rob. Sens. Env., pp. 31–36, 2013.
- [11] A. K. Krishnan and S. Saripalli, "Cross-calibration of rgb and thermal cameras with a lidar for rgb-depth-thermal mapping," *Unmanned Syst.*, pp. 1–20, 2017.
- [12] D. G. Unfallversicherung, "Bg/bgia risk assessment recommendations according to machinery directive, design of workplaces with collaborative robots," Report, Tech. Rep., 2009.
- [13] J. Falco, J. A. Marvel, and R. Norcross, "Collaborative robotics: Measuring blunt force impacts on humans," *Proc. Int. Conf. Safety Indust. Autom. Syst.*, 2012.
- [14] N. Vlajic and A. Chijioke, "Traceable dynamic calibration of force transducers by primary means," *Metrologia*, vol. 53, no. 4, p. S136, 2016.
- [15] E. Messina. (2017) Performance assessment framework for robotic systems. [Online]. Available: "https://www.nist.gov/programsprojects/performance-assessment-framework-robotic-systems"
- [16] N. G. Dagalakis, J.-M. Yoo, J.-M. Yoo, T. Oeste, and T. Oeste, "Humanrobot collaboration dynamic impact testing and calibration instrument for disposable robot safety artifacts," *Indust. Rob.: An Int. J.*, vol. 43, no. 3, pp. 328–337, 2016.
- [17] J. A. Marvel and K. Van Wyk, "Simplified framework for robot coordinate registration for manufacturing applications," *Proc. IEEE Int. Symp. Assembly Manuf.*, pp. 56–63, 2016.
- [18] "ISO 9283 Manipulating industrial robots Performance criteria and related test methods," ISO, Standard, 1998.
- [19] K. Van Wyk and J. A. Marvel, "Strategies for improving robot registration performance," *IEEE Trans. Autom. Sci. Eng.*, In Press.
- [20] R. Bostelman, R. Eastman, T. Hong, O. Enein, S. Legowik, and S. Foufou, "Comparison of registration methods for mobile manipulators," *Proc. Int. Conf. Climbing Walking Rob.*, pp. 205–213, 2016.
- [21] S. Legowik, R. Bostelman, and T. Hong, "Sensor calibration and registration for mobile manipulators," *Proc. 5th Int. Conf. Adv. Vehic. Syst., Techn. Appl.*, 2016.
- [22] J. A. Marvel, J. Falco, and I. Marstio, "Characterizing task-based human-robot collaboration safety in manufacturing," *IEEE Trans. Syst.*, *Man, and Cybern.: Syst.*, vol. 45, no. 2, pp. 260–275, 2015.

Experiments to Quantify Firebrand Production from Roofing Assembly Combustion

MANZELLO, Samuel L.1*, SUZUKI, Sayaka2, and NARUSE, Tomohiro3

¹Fire Research Division, National Institute of Standards and Technology (NIST), Gaithersburg, MD USA ²Large Fire Laboratory, National Research Institute of Fire and Disaster (NRIFD), Chofu, Tokyo, JAPAN 3Department of Fire Engineering, Building Research Institute (BRI), Tsukuba, Ibaraki, JAPAN

Abstract: Large outdoor fires present a risk to the built environment. One example often in the international media reports are wildfires that spread into communities, referred to as Wildland-Urban Interface (WUI) fires. Another example are large urban fires including those that have occurred after earthquakes. Firebrands are a key mechanism on how rapidly fires spread in these types of fires, as was observed in the 2016 Itoigawa City Fire in Niigata, Japan. To this end, an experimental protocol has been developed to ignite full-scale roofing assemblies and quantify the degree of firebrand production during these combustion processes. As wind is an important factor in firebrand generation, the experiments were conducted under a range of wind speeds at the Building Research Institute's (BRI) Fire Research Wind Tunnel Facility (FRWTF). Results of these experiments will be presented and discussed.

Keywords: Firebrands, Large Outdoor Fires, Full-Scale Roofing Assemblies

1. Introduction

There have been significant number of wildfires that spread into communities all over the world. Such fires are referred to as Wildland-Urban Interface (WUI) fires. In 2017, more than 60 people perished in these fires in Portugal. Japan is another country that has been subjected to large scale outdoor fires, the most recent of these occurred in Itoigawa City, Japan on December, 2016. In this fire, 147 structures were damaged by fire, with 120 of 147 destroyed. A governing mechanism of structure ignition in both WUI and urban fires under wind is the production of firebrands, or embers.

Firebrands have been known to be a significant cause of structure ignition in WUI fires as well as large urban fires. Limited data exist with regard to firebrand size distributions from actual structures or large outdoor fires, such as WUI and urban fires [1-3]. In WUI fires, the structures themselves may be a large source of firebrands, in addition to the vegetation. Yet, due to a lack of quantitative information available on production of firebrands from structures, it cannot be determined if firebrand production from structures is a significant source of firebrands in WUI fires. In the case of urban fires in Japan, structures are responsible for firebrand production, but as in the case of WUI fires, little is known about firebrand size/mass distributions produced.

In this paper, efforts to investigate firebrand production from full-scale roofing assemblies is presented. These results are being incorporated into a comprehensive database the authors are developing related to firebrand production from structures.

2. Experimental Description

Experiments were performed in the Building Research Institute's (BRI) Fire Research Wind Tunnel Facility (FRWTF) by varying the wind speed. As part of our database development, wall assemblies and re-entrant corner assemblies were used in experiments; here roofing assemblies have been considered.

The roofing assemblies used in the experiments were 1.2 m by 1.2 m in dimension. Wood joists (2 x 6) were used as framing, and oriented strand board (OSB - 11 mm thickness) was used as the base sheathing material. As a first-step, specific roofing treatments, such as roof tiles, were not applied, and are the obvious next step. The assemblies were installed inside the test section of the FRWTF at BRI (see Fig. 1). A custom frame was used to support the roofing assemblies and roof angle was fixed at 25 degrees.

The FRWTF was equipped with a 4.0 m diameter fan to produce a wind field up to a 10 m/s (\pm 10 %). The wind velocity distribution was verified using a hot wire anemometer array. To track the evolution of the size and mass distribution of firebrands, a series of water pans was placed downstream of the assemblies.

The ignition method used in this study was developed based on lessons learned from igniting wall assemblies for firebrand collection [4]. This method provides repeatable conditions to collect firebrands during combustion of the assemblies. It is important to realize that it is very difficult to simulate the conditions of an actual WUI or urban fire in a controlled laboratory setting. The effect of wind speed on firebrand size and mass distribution generated from a given roofing assembly configuration is an important parameter to study. For this study, each assembly was ignited using a propane gas T-shaped burner with a heat release rate of 26 kW positioned adjacent to the assemblies for 10 min under conditions of no wind. The T-shaped burner was placed on the outside of the roofing assemblies, at the leading edge, since the purpose of this study was to simulate ignition from an outside fire. If the burner was applied under wind (e.g., 6 m/s), flaming combustion of the assembly was difficult to achieve due to convective heat loss from the applied wind flow. Another advantage of ignition under no wind, with the burner applied for the same duration for each assembly, was that it provided more repeatable initial conditions for the experiments. Specifically, the area exposed to direct flame contact was similar for a given assembly. If igniting under wind, in addition to large convective heat loss, the contact area of the flame onto the OSB surface of the assembly became unsteady.



Figure 1 Top view of BRI's FRWTF. The location of roofing assemblies and water pan array is shown.

With the application of the T-shaped burner to the assembly under no wind, flaming combustion was observed on the exterior of roofing assembly. Once the burner was switched off, the wind tunnel was switched on. Firebrands were collected until the assemblies were consumed to such a degree that they could no longer support themselves (loss of structural integrity). Fig. 2 (a) displays an image

Manzello, Samuel; Naruse, Tomohiro; Suzuki, Sayaka. "Experiments to Quantify Firebrand Production from Roofing Assembly Combustion." Paper presented at 55th Japanese Combustion Symposium, Toyama, Japan. November 13, 2017 - November 15, 2017.

after the ignition of roofing assembly, with a 6 m/s wind applied. Firebrands were collected using a series of water pans placed behind the roofing assembly. Water was necessary to quench the combustion of the generated firebrands. After deposition into the water pans, firebrands were filtered from the water using a series of fine mesh filters. Firebrands were dried in an oven at 104 °C for 24 h.



Figure 2 Roofing assembly combustion exposed to an applied wind speed of 6 m/s. The T-shaped burner is shown (gas supply terminated since 10 min ignition period has passed).

3. Results and Discussion

Experiments were conducted for wind speeds of 6 m/s and 8 Image analysis software was used to determine the projected area of a firebrand by converting the pixel area using an appropriate scale factor. It was assumed that deposited firebrands would rest flat on the ground and the projected areas with the maximum dimension and the second maximum dimension of three dimensions were measured. Images of well-defined shapes (e.g. circular objects) were used to determine the ability of the image analysis method to calculate the projected area [4]. Based on repeat measurements of different areas, the standard uncertainty in determining the projected area was \pm 10 %. The mass of each firebrand was measured by a precision balance with 0.001 g resolution. Repeat measurements of known calibration masses were measured by the balance which was used for the firebrand mass analysis. The standard uncertainty in the firebrand mass was approximately ± 1 %.



Figure 3 Size and mass distribution of collected firebrands from roofing assembly combustion. The roofing assembly was constructed from 2 by 6 joists lined with OSB as sheathing.

Fig. 3 displays a comparison of the collected firebrand size and mass distribution collected from these experiments. To more clearly show the results, the graph is scaled up to projected areas below 2.5 cm². Of the more than 250 firebrands collected, only two firebrands were larger than 2.5 g.

It is of interest to compare these roofing assembly data to other experiments, conducted previously that considered re-entrant corner wall assemblies as the firebrand ignition source [4]. In those experiments, re-entrant corners were constructed using wood studs (2 x 4), with OSB sheathing applied. Specifically, the wall dimensions were 1.2 m wide one each side, with a height of 2.44 m. The wood studs were spaced 40 cm on center [4] and the same wind speeds were applied (6 m/s and 8 m/s).



Figure 4 Size and mass distribution of collected firebrands from re-entrant corner combustion [4] directly compared to roofing assembly data in Fig 3. The wall dimensions were 1.2 m wide one each side, with a height of 2.44 m.

Similar to the roofing assembly results, Fig. 4 is scaled up to projected areas below 2.5 cm². A direct comparison of the different building components may be seen.

4. Summarv

An experimental protocol has been developed to ignite full-scale roofing assemblies and quantify the degree of firebrand production during the combustion process. As a first-step, roofing assemblies were constructed using only OSB as a sheathing material. While more work remains, when only OSB is applied as sheathing, a significant number firebrands collected from re-entrant corner assemblies and roofing assemblies were less than 1 g and 10 cm². Future work will of course consider various external sheathing types.

5. Acknowledgments

Mr. Marco G. Fernandez of NIST is appreciated for various assistance with this experimental campaign.

6. References

1. Vodvarka, F.J., Firebrand Field Studies-Final Report, IIT Research Institute, Chicago, IL, 1969.

2. Manzello, S.L., and Foote, E.I.D., Characterizing Firebrand Exposure During Wildland-Urban Interface Fires: Results of the 2007 Angora Fire, Fire Tech., 50 (2014) 105-124.

3. Suzuki., S., Fujisawa, S., and Manzello, S.L., Characteristics of Firebrands Collected from Urban Fire -Niigata Fire, December 22nd 2016, 12th IAFSS Poster Session, Lund, Sweden, 2017.

4. Suzuki, S., and Manzello, S.L., Firebrand Production from Building Components Fitted with Siding Treatments, Fire Saf. J., 80 (2016) 64-70.

Events for the Application of Measurement Science to Evaluate Ground, Aerial, and Aquatic Robots*

Adam Jacoff, Richard Candell, Anthony Downs, Hui-Min Huang, Kenneth Kimble, Kamel Saidi, Raymond Sheh and Ann Virts

Abstract—This paper reports on three measurement science field exercises for evaluating ground, aerial, and aquatic robots. These events, conducted from February to June 2017, were conducted in close co-ordination with the responder community, standards organizations, manufacturers, and academia. Test data from a wide variety of robot platforms were gathered in a wide variety of standard and prototypical test methods ranging from mobility and manipulation to sensors and endurance.

I. INTRODUCTION

Remotely operated robots enable emergency responders to perform extremely hazardous tasks from safer stand-off distances. Standard test methods help robot manufacturers and users objectively evaluate system capabilities to align with mission requirements. This improves the safety and effectiveness of emergency responders as they attempt to save lives and protect property in our communities. The ASTM International Standards Committee on Response Robots for Homeland Security Applications (E54.09) is developing the standards infrastructure necessary to inspire innovation, inform purchasing decisions, and focus training with measures of operator proficiency [1].

The underlying measurement science is being developed to quantitatively evaluate and compare robotic system capabilities. Along with the standards committee meeting, comprehensive test method validation exercises were organized with sponsorship from the Department of Homeland Security, Science & Technology Directorate, Capability Development Support Group, Office of Standards. These exercises provided an opportunity for robot manufacturers, emergency responders, and international collaborators to validate 50 test methods for ground, aerial, and aquatic systems. This paper reports on three of these exercises, conducted in the first half of 2017.

The first was a five day meeting and exercise, hosted at the Virginia Beach Fire Department Training Facility in February in conjunction with the January 2017 E54.09 meeting. This is home to the Federal Emergency Management Agency's (FEMA's) Virginia Task Force 2 Urban Search and Rescue (US&R) Team. Roughly 100 people participated each day from Canada, Germany, Israel, Japan, Korea, UAE, UK, and the USA. More than 30 robots provided by manufacturers, emergency responders, and military organizations evaluated their robotic capabilities and operator proficiencies within the test methods.

The second was a one day exercise hosted by the Virginia Department of Public Safety and took place in Charlottesville, VA in March as part of the National Unmanned Aerial Systems Summit. The purpose of the summit was to demonstrate the use of aerial systems in public safety applications and included speakers and panels on fire, search and rescue, and law enforcement. Roughly 20 manufacturers and 150 attendees participated and watched as the final day included aerial systems operating within various training scenarios.

The third event was a week-long inter-laboratory experiment attached to the June 2017 E54.09 meeting and hosted by the Canadian Explosive Technicians' Association at the Hamilton Port Authority near Toronto, Ontario, Canada. They were supported in conducting robot operator training, across ground, aerial and aquatic systems using the DHS-NIST-ASTM International standard test methods to measures and compare their proficiency across more than 20 ground, aerial, and aquatic systems.

The rest of this document reports on the specific outcomes of the three events before summarizing the overall impacts and the plans for the future.

II. VIRGINIA TASK FORCE 2

FEMA's new US&R Robotics Standards Committee met for the first time. They are tasked with selecting suites of standard test methods to evaluate ground, aerial, and aquatic systems to purchase for US&R Task Force equipment caches nationwide. The Virginia Beach Fire Department Training Facility has several US&R training scenarios that were embedded with standard test methods to enable quantitative evaluation of robots even within the uncontrolled and very difficult environmental variables.

Dozens of civilian and military bomb squad technicians in the region participated with their own ground robots to validate new approaches toward standardized robot operator training for countering improvised explosive devices (C-IED). This suite of 60 cm (24 in) and 120 cm (48 in) wide ground robot test methods have been used across the country at several such events and hosted in dozens of responder facilities. The test methods are shown in Figure 1. The smallest 30cm (12in) wide ground robot test methods were also deployed along with rapidly manufactured robot kits that were built on 3D printers and laser cutters to create truly disposable robots for the most hazardous tasks [2].

"Events for the Application of Measurement Science to Evaluate Ground, Arabia, and Aquatic Robots." Paper presented at The 15th IEEE International Symposium on Safety, Security, and Rescue Robotics 2017 (SSRR 2017), Shanghai, China.

October 11, 2017 - October 13, 2017.

^{*}This work was supported by the US Department of Homeland Security, Science & Technology Directorate, Capability Development Support Group, Office of Standards.

A. Jacoff, R. Candell, A. Downs, H.M. Huang, K. Kimble, K. Saidi, R. Sheh and A. Virts are with the National Institute of Standards and Technology, Gaithersburg, MD 20878 USA (e-mail: robottestmethods@nist.gov).

R. Sheh is also with the Intelligent Robots Group, Department of Computing, Curtin University, Kent St, Bentley WA 6102, Australia.

Candell, Richard; Downs, Anthony; Huang, Hui-Min; Jacoff, Adam; Kimble, Kenneth; Saidi, Kamel; Virts, Ann.

Underwater explosive ordinance disposal (EOD) teams from the local Navy base collaborated with robot manufacturers demonstrating the newest available remotely operated vehicles (ROVs) in an 18,000 gallon water tank with more than a dozen test methods plus apparatuses to add turbulence and various lighting conditions. The repeatable



Figure 1: Ground robot tests for Operator Proficiency – Basic Skills at 120 cm (front) and 60 cm (rear) scales.

tasks in various orientations, floating, and on the tank bottom, along with the clear water and videos of the tasks being performed provided key insights into the capabilities of the systems and helped measure important upgrades implemented by manufacturers since the previous exercise. Beyond the obvious training benefits, these test methods provide useful de-bugging tools and readiness assessments, especially after repairs.

Indoor and outdoor netted aviaries contained more than a dozen test methods to evaluate small unmanned aircraft systems (sUAS) in both still and turbulent air. These systems were provided and operated by manufacturers and responder organizations leading the effort to safely and effectively incorporate these emerging technologies. These test methods quantify essential safety features and airworthiness while providing useful practice tasks that measure proficiency for responders or hobbyists prior to flying in the national airspace.

Several embedded scenarios were deployed to capture quantitative scores within the uncontrolled environmental variables. These included a random maze inside the "smoke house" to test sensor degradation due to obscurants, a vehicle-borne improvised explosive device (truck bomb) and a rubble pile and surrounding wide area survey embedded with visual, thermal, and mapping targets.

III. NATIONAL UNMANNED AERIAL SYSTEMS SUMMIT

This outdoor event helped manufacturers and emergency responders validate the test methods within their own training scenarios conducted by fire departments, law enforcement, and others. In fact, most of the test method apparatuses were brought to the site by local responders, who used the aerial system test methods to train and evaluate firefighters in their departments. NIST set up a comprehensive evaluation station shown in Figure 2, with 10 test methods developed to measure maneuvering, endurance, sensing, and situational awareness. These test methods augmented the surrounding scenarios by providing objective, quantitative, repeatable, and reproducible measures enabling direct comparison of system capabilities and operator proficiency. The training scenarios for "Structure Fire" and "Hazardous Materials" included



Figure 2: Overview of the outdoor aerial tests. For scale, the center pillar is 2.4 m (8 ft) tall.

some of NIST's test apparatuses embedded to demonstrate how to capture quantitative scores even within uncontrolled environmental variables such as winds gusting over 20 knots. Other scenarios included "Disaster Mapping," "Search and Rescue," and "Law Enforcement."

IV. CANADIAN EXPLOSIVE TECHNICIANS' ASSOCIATION

Like the event held with the Virginia Task Force 2, this event engaged the responder and manufacturer communities with the standard test method process. The new FEMA US&R Robotics Standards Committee also used the event to select suites of standard test methods to evaluate ground, aerial, and aquatic systems to purchase for US&R Task Force equipment caches nationwide.

For the first time in Canada, ground robot exercises were conducted, focusing on evaluating bomb squad proficiency and using the 128 cm (48 in) suite of ground robot tests. Underwater robot tests were conducted both in a dedicated watertank as well as in Lake Ontario, with a focus on the evaluation of embedded test apparatuses for best-in-class sonar and vision. Aerial tests were deployed in both a netted aviary and an outdoor scenario. These included initial prototypes of a test for resilience of systems to water, as well as the latest iteration of a radio communications attenuation test for handheld operator interfaces.

V. CONCLUSION

These events, and those that preceded them, have resulted in responder organizations around the world adopting the test methods into their training and procurement procedures. These events also encourage the responders' involvement in the development of these standard test methods, as a crucial part of the standard test method process. These events continue to be refined and seeded with the latest prototyping test methods and initiatives.

REFERENCES

- [1] A. Jacoff, E. Messina, H. M. Huang et. al., "Guide for Evaluating, Purchasing, and Training with Response Robots using DHS-NIST-ASTM International Standard Test Methods", from www.nist.gov/el/isd/ks/upload/DHS_NIST_ASTM_Robot_Test_Meth ods-2.pdf, 2014.
- [2] R. Sheh, A. Eguchi, H. Komsuoglu, A. Jacoff, "The Open Academic Robot Kit", in *Robotics in STEM Education*, Springer, 2017, pp. 85-100.

Candell, Richard: Downs, Anthony: Huang, Hui-Min; Jacoff, Adam; Kimble, Kenneth; Saidi, Kamel; Virts, Ann,

"Events for the Application of Measurement Science to Evaluate Ground, Arabia, and Aquatic Robots." Paper presented at The 15th IEEE International Symposium on Safety, Security, and Rescue Robotics 2017 (SSRR 2017), Shanghai, China.

October 11, 2017 - October 13, 2017.

Investigation on Firebrand Generation from Reduced-scale Mock-ups

SUZUKI, Sayaka^{1*}, MANZELLO, Samuel, L.²

¹ National Research Institute of Fire and Disaster, 4-35-3 Jindaiji-higashimachi, Chofu, Tokyo, 182-8508, Japan
 ² National Institute of Standards and Technology, 100 Breau Dr. Gaithersburg, MD, 20899, USA

Abstract : Firebrands generated from structures are known to be a source of rapid flame spread within communities in large outdoor fires, such as wildland-urban (WUI) fires, and urban fires. It is important to better understand firebrand generation mechanism. To this end, a series of experiments were performed using mock-ups of full-scale walls exposed to wind. The objective of this study was to examine if experiments with mock-ups may provide insight into firebrand generation from structures. Specifically, generated firebrands were collected and compared with those collected from full-scale components. Results of these experiments will be presented and discussed.

Keywords : Firebrand, Generation, Large Outdoor Fires, WUI Fires, Urban Fires

1. Introduction

Large outdoor fires are becoming an important research area across the world. Japanese cities are densely populated and there exists the potential for large-scale urban fires [1]. These may or may not be produced after the occurrence of strong earthquakes. The Great Hanshin earthquake in Kobe, Japan in 1995 is one example. The recent 2016 Itoigawa City fire that occurred in Niigata, Japan, is an example where no earthquake was present, but a large-scale urban fire developed.

In many countries outside Japan, wildland fires that spread into communities, termed wildland-urban interface (WUI) fires, are frequently seen in media reports, and have resulted in loss of life, and property damage. While both urban fires and WUI fires are complex, it is possible to develop scientific-based mitigation strategies by attempting to understand how structures are ignited in these fires.

In both the USA and Japan, post-fire disaster investigations have pointed to the significance of firebrands, or embers, as a leading driver of *both* fire spread, and structure ignition [2-3]. A simple-minded approach is to attempt to develop test methods to screen building materials for firebrand production. In this manner, it may be possible to design structures that do not produce firebrands.

Conducting many full-scale experiments of structure firebrand production is costly, and not easy to undertake. As a result, the authors have embarked on an experimental program to determine if smaller-scale experiments may produce useful insight into the firebrand production process. The objective of this study was to investigate if experiments using mock-ups of full-scale wall assemblies may provide insight into firebrand generation from actual, full-scale wall assemblies. As a first-step, firebrands generated in the current paper were collected and compared with those collected from previous experiments that made use of full-scale walls constructed of common building materials [4-5].

2. Experimental Descriptions

A series of experiments with reduced-scale mock-ups of building materials were performed in a wind facility in National Research Institute of Fire and Disaster (NRIFD). NRIFD's wind facility has a 4 m diameter fan. The flow field was measured to be within \pm 10 % over a cross-section of 2.0 m by 2.0 m.

Mock-ups used in this study were wall corner assemblies, constructed with Oriented Strand Board (OSB) and 2×4 studs with the dimensions of 0.6 m (W) x 1.2 m (H) (Fig. 1). This is half height and width of full-scale corner walls in previous experiments [4].

Experiments were performed in the following manner. First, mock-up assemblies were ignited by a T-shaped burner for 5 minutes. After the burner was turned off, wind at desired wind speeds (6 or 8 m/s) was applied. Combustion was stopped when the assemblies were not able to support themselves anymore.

A series of water pans were used to collect firebrands

generated from the mock-ups. Water is needed to quench combustion otherwise only ash remains when experiments are finished. Water pans were placed downwind from reduced-scale mock-ups. After experiments, pans were collected, and firebrands were filtered and dried at 104 °C for 24 hours. Dried firebrands were measured with scale and pictures were taken for image analysis.



Figure 1 Schematic of mock-ups of corner assemblies.

3. Results and Discussions

Fig. 2 displays an image of an experiment with a mock-up corner assembly under an 8 m/s wind. After each experiment, firebrands were collected and analyzed.

The size distribution of firebrands generated from the mock-ups are analyzed and shown in **Fig. 3**. Image analysis methods, described elsewhere were used **[6]**. While some of firebrands were more than 1 g, a majority of firebrands collected in this study were relatively small with less than 1 g mass and 10 cm² projected area. This is the same as firebrands generated from full-scale wall assemblies. **Fig. 4** shows images of firebrands collected in this study under an 8 m/s wind.

Fig. 5 shows the comparison of the size and mass of firebrands generated both mock-up and full-scale corner walls under two different wind speeds (6 and 8 m/s). This shows that firebrands with similar size and mass classes may be observed.

4. Summary

A series of experiments were performed to investigate if experiments with much simpler mock-ups may provide the insight of firebrand generation from full-scale corner wall assemblies. The comparison of firebrands generated from mock-ups and full-scale corner walls showed firebrands with similar size and mass classes may be observed. While these results are encouraging, significantly more work is needed to determine if such simple experiments may capture more complex firebrand production effects, such as the addition of siding materials.



Figure 2 Image of an experiment under 8 m/s wind. Mock-ups used in this study were wall corner assemblies, constructed with Oriented Strand Board (OSB) and 2 x 4 studs with the dimensions of 0.6 m (W) x 1.2 m (H) (**Fig. 1**). This is half height and width of full-scale walls in previous experiments.



Figure 3 Comparison of size distribution collected in this study to those collected from full-scale corner assemblies. All comparisons are for wind speeds of 6 m/s.



Figure 4 Images of firebrands generated from reduced-scale mock-ups under an 8 m/s wind. For scale, the projected area of top left firebrand was 0.92 cm².



Figure 5 Comparison of size and mass of firebrands generated from mock-ups and from full-scale corner wall assemblies [4] under different wind speeds (6 and 8 m/s) This graph focuses up to 2.5 g mass and 20 cm² projected area for detailed comparison.

5. Reference

[1] Manzello, S.L., et al., Fire Saf J. 59 (2013) 122-131.

[2] Maranghides, A., et al., A Case Study of a Community Affected by the Waldo Fire – Event Timeline and Defensive Actions, NIST TN1910, 2015. http://dx.doi.org/10.6028/NIST.TN.1910

[3] National Research Institute of Fire and Disaster, Fire Spread Behaviors of The Sakata Fire, Technical Report of National Research Institute of Fire and Disaster, No.11, 1977 (in Japanese).
[4] Suzuki, S., et al., Proceedings of the Combustion Institute, Vol. 34, Issue 2, (2013) pp.2479-2485.

[5] Suzuki, S., et al., Fire Safety Journal, Vol. 63 (2014) pp. 43-51.
[6] Suzuki, S., et al., International Journal of Wildland Fire, Vol. 21, Issue 8, (2012) pp.961-968.

Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview

Jennifer Helgeson¹, Juan Fung¹, Chevnev O'Fallon¹, David Webb¹, Harvey Cutler²

¹ Applied Economics Office, Engineering Laboratory, National Institute of Standards and Technology

² Economics Department, Colorado State University

Abstract

This paper introduces the concept of accounting for the net co-benefits (the resilience dividend) associated with community-level resilience planning. Two solutions to the same resilience issue may often have different associated co-benefits that accrue on a day-to-day basis even if a disruptive event has not vet occurred. Thus, assessing potential community resilience projects requires taking (positive or negative) co-benefits (i.e., the resilience dividend) into account. Without including positive (negative) co-benefits, the total value of a resilience project may be underestimated (overestimated). But to date, quantification of the net co-benefits of resilience planning is not often addressed in the literature, as it is

adaptation, mitigation,

Keywords

not a straight-forward task. We overview a methodology developed using spatial computable general equilibrium (SCGE) models to guantify and assess the distributional effects of the resilience dividend arising from a proposed resilience plan. In turn, such assessments can be used in benefit-cost analyses (BCAs) and other economic project assessments when comparing among potential resilience projects. Economically, good decision-making requires prioritizing feasible projects with the greatest overall net-benefit to the community. We provide a way for co-benefits to be guantified and subsequently accounted for in formal assessment by communities choosing among resilience plans.

1. Introduction

The number of observed large-scale disruptive natural events is rising - by about five percent a year since 1960 (Schultz and Elliott 2013). Kunreuther and Michel-Kerjan (2009) note that costs of natural disaster-related losses jumped from \$93.3 billion in the 1960s to \$ 778.3 billion in the 1990s. Strömberg (2007) notes that population growth (meaning more people encounter disasters) explains only about half of this increase. After all, there has also been a marked reduction in lives lost due to natural disasters.40 An important

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

⁴⁰ From 1900 to 2003, 62 million deaths resulted from natural disasters throughout the world. But 85 percent of those deaths occurred between 1900 and 1950 (Bandyk 2010)



factor in the increased number of reported disasters is likely better reporting and more responsive aid organizations as well as changing climatic trends.

As weather-related covariate risks and the associated costs of losses increase in the future, households and businesses need resilient strategies and coping mechanisms that reduce the effects of such disasters, in terms of intensity and economic losses. Generally, as assets vulnerable to natural disasters increase in value, so do costs of protecting these assets and infrastructure through insurance and/or other means of planning. Thus, the concept of choosing resilience plans that encompass co-benefits to the community on a day-to-day basis in the absence of a disaster event has garnered increased interest recently (e.g., Rodin 2014). Accounting for the net co-benefits (i.e., the "resilience dividend") of a resilience project can often produce a convincing business case for undertaking the project. This is especially pertinent when the return on investment may be much lower if a disaster does not take place during the time frame of the analysis. Fung and Helgeson (2017) define the "resilience dividend as the net benefit (or cost) that accrues, from investments aimed at increasing resilience, in the absence of a disruptive incident over the planning horizon," and provide a comprehensive overview of the resilience dividend as a useful metric for community resilience planning and reviews measurement and assessment efforts.

This paper provides an overview of the importance of accounting for the net co-benefits of resilience planning and explores a novel approach to quantifying the resilience dividend. The methodological approach introduced uses a spatial computable general equilibrium (SCGE) model of the community being assessed for resilience planning to identify co-benefits (co-costs).

The remainder of this paper is organized as follows. Section 2 provides context by reviewing the literature and some approaches that strive to quantify the resilience dividend, which to date has been largely dealt with through qualitative case studies (Rodin 2014). The section highlights the importance of considering the economic flows from the resilience dividend in a dynamic, quantifiable manner. Section 3 provides an overview of CGE and SCGE modelling. Section 4 provides a detailed discussion of data required to use SCGE models with special focus on the characteristics of a CGE model designed to trace co-benefit-related flows and distributional effects. It discusses the complex nature of obtaining data for CGE models and the accompanying social accounting matrix (SAM). This section offers insight as to the ideal

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

192

data versus data that is sufficient in most cases. Section 5 takes the data discussion towards methodological implementation. Finally, Section 6 highlights next steps and future work to develop a full case study based on the SCGE net co-benefits methodology introduced in this paper.

2. Background and Motivation

2.1. The importance of considering resilience-related co-benefits

Economic valuation techniques, such as benefit-cost analyses (BCAs) for community resilience planning alternatives, are often not a straightforward process. In nearly all cases, measuring the economic impact associated with resilience planning requires a better understanding of the costs and indirect losses41 to maintain a full accounting of the major cost elements. On the loss side, an understanding of the cascading indirect losses is critical to true valuation of losses stemming from a natural disaster. Furthermore, guantifying and accounting for uncertainty in estimates related to these costs and losses is complicated due to the nature of disaster events and the uncertainty surrounding their occurrence and effects. Finally, measuring net co-benefits (i.e., the resilience dividend) is needed to articulate the business case for resilience planning. Often plans that could alleviate vulnerability to a large-scale disruptive event, but are not called into action due to the absence of the event (in a given time frame), are perceived as a poor investment. Consideration of co-benefits (co-costs) is generally good practice,42 as the impacts of these values can be pivotal in identification of the most effective and efficient resilience plan.

When quantification of co-benefits is possible, they should be folded into the net-present valuation (NPV) of resilience plans (see Gilbert et al. (2016) and Helgeson et al. (2017)). Yet, much like cascading indirect losses, there are likely cascading and wide-spread effects of identified co-benefits. Thus, the use of computable general equilibrium (CGE) models that employ actual economic data from a community to estimate how an economy might react to changes in policy, technology, or other resilience planning initiatives allow for a better understanding of distributive effects of net co-benefits. Specifically, spatial computable general equilibrium (SCGE) models can be employed to indicate the distinction of flows throughout different areas of a community, which may be more or less vulnerable to and affected by a disruptive event. CGE and SCGE models are overviewed in Section 3.

2.2. Defining and quantifying the resilience dividend - overview

To date, the literature is largely dominated by definitional discussions and gualitative assessments of co-benefits (co-costs) and the resilience dividend using case study examples (e.g., Rodin 2014). In a review of the literature, Fung and Helgeson (2017) found that co-benefits fall into three broad categories: 1. Objective-based, 2. Intent-based, and 3. Externality-based.43 The objective-based definition of co-benefits fits well into the methodology overviewed in this paper. Objective-based definitions regard co-benefits as benefits to secondary objectives of a policy (ibid.). For instance, changed zoning in a community may have a primary objective of shifting commerce away from the flood zone, while secondary objectives may include stimulating economic growth in an area of town that becomes favorable for re-locating businesses.

As noted in Fung and Helgeson (2017), research on the co-benefits of climate change mitigation and adaptation is substantial, while co-benefits in the context of resilience planning is still relatively nascent.

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

⁴¹ To date, direct losses tend to be better documented.

⁴² See Gilbert et al. (2016) and Helpeson et al. (2017)

⁴⁵ Externalities are defined by benefits (costs) that accrue to third parties. As such we treat them fundamentally different from values that are encompassed by the resilience dividend. For a discussion of externalities versus non-disaster related benefits (i.e., the resilience dividend), see Gilbert et al. (2016) and Helgeson et al. (2017).

Much of the literature on co-benefits of resilience planning is centered upon the developing country context. Furthermore, there appears to be relatively few scholarly works that deal with quantification, opposed to qualitative assessment, of co-benefits. This is understandable, as much of the work that explicitly encourages quantification of co-benefits when possible is based upon ex ante analysis, such as BCA, to determine effective investment decisions across a suite of options. In an ex ante BCA it is naturally complicated to capture full valuation of co-benefits, which often are apparent only after a plan is put in place. In other words, some co-benefits of significant value may not be readily obvious during the planning phase without a larger scale model that can incorporate spatial and/or distributive effects. But quantifying the full co-benefits ex post is not a simple task – modeling the economy is complex, it is likely very unclear how the co-benefits flow through the economy, and since the decision was already made, stakeholders may be less inclined to spend money and other resources on studying the issue.



A series of World Bank reports have presented the resilience dividend as arising from a "Triple Dividend of Resilience" as largely relevant to disaster risk management (DRM) (e.g., Tanner et al. 2015, Tanner et al. 2016, Mechler et al. 2016). This triple bottom line consists of: 1. avoided or reduced losses, in the event of a disruptive event occurring; 2. increased economic resilience from reduced disaster risk; and 3. co-benefits for development. Elements one and two make up the first dividend of resilience, while the third element makes up the second as shown in Figure 1. Though these three "dividend" sources do not map perfectly onto the developed country context, the prevailing message is that budgeting for contingent liabilities such as disaster risk, especially ex ante a disruptive event, is nearly impossible without accounting for the resilience dividend.

A recent RAND report (Bond et al. 2017) describes a Resilience Dividend Valuation Model (RDVM) and its application to six case studies in the developing country context. It should be kept in mind that Bond et al. (2017) define the resilience dividend as "the difference in net benefits from a project developed with a resilience lens versus one that is not." This definition is much broader than the definition we use (Fung and Helgeson 2017), which is concerned with net benefits above and beyond benefits expected to accrue directly to the goal of resilience to a disruptive event.

The RDVM largely looks at the resilience dividend as the positive net benefits generated between a resilience project and a business as usual (BAU) counterfactual. The elements of the RDVM are largely based on typical meso- and macro-economic elements within a production-oriented framework: 1. Capital stocks/assets, 2. Production functions and allocation mechanisms (i.e., institutions), 3. Social welfare function, 4. Shocks and stressors (both ex ante and ex post), and 5. Project interventions (based on resilience).

Of the six case studies considered in the RAND Report, three are ex ante and three are ex post assessments. Three of these six case studies resulted in no guantifiable resilience dividend assessment (two ex post and one ex ante) and three result in a partial quantitative assessment of the resilience dividend (two ex ante and one ex post). In many cases the lack of a full quantitative resilience dividend analysis is discussed in the context of too little data

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

194

being available through pre-existing documentation and data.44 Another challenge discussed is that only one state of the world is observed-the counterfactual is unobservable-it is then difficult to rely on observations made with or without a resilience intervention (i.e., plan) in place (ex ante or ex post).

The systems model approach we propose for assessing the resilience dividend is typically based at the meso-level of a community's economy and allows us to make assessments of the resilience dividend and the associated indirect flows throughout the economy. Many of the elements discussed in the RDVM in terms of a production-oriented framework are reflected in the SCGE model approach we describe in this paper. In our approach, we can theoretically obtain community-level data for any US-based community that may be engaged in resilience planning and assess ex ante potential resilience dividends as well as ex post performance. There are limitations inherent in this approach; this is especially the case for micro-level economic activity, at the household, opposed to community levels. This approach is a first step toward creating dynamic guantitative valuation of the resilience dividend and distributive effects.

3. Computable General Equilibrium Models -Introduction and Overview

3.1. CGE General Details and History

The characteristics of CGE models make them a reasonable choice for exploring the effect of large disruptive events on a community's economic activity as well as effects of resilience planning. This section provides a general outline of CGE and SCGE models. Specific use of an SCGE model and the relative data requirements is discussed in Section 4 of this paper.

There are two major reasons for exploring the use of CGE models to quantify the resilience dividend. The first being that while gualitative results are useful, understanding the relative effects in magnitude of a shock and the associated resilience plan as well as the resilience dividend is important. The second being that solid micro-foundations enhance our understanding of resilience planning and how a resilience dividend affects consumers, producers, and government in an economy. Overall, the aim of the CGE model approach is to convert the abstract representation of the community's economy into a realistic, solvable approximation to assess direct and indirect benefits of resilience planning. In turn, these assessments can help inform values used for co-benefits (co-costs) of resilience planning in (ex ante) BCA during planning phases.

Input-output (I-O) analysis (Leontief 1941) has been used for assessing the impact of a change in the demand conditions for a given sector of the economy.⁴⁵ I-O models/ coefficients assume constant returns-to-scale for associated production functions and prices are also assumed to remain constant. Extension of the I-O model to a social accounting matrix (SAM) framework is performed by partitioning the accounts into endogenous and exogenous accounts. Sadoulet and de Janvry (1995) note that endogenous accounts are those for which changes in the level of expenditure directly follow any change in income, while exogenous accounts are those for which we assume that the expenditures are set independently of income.

The CGE model encompasses both the I-O and SAM frameworks; this occurs because demand and supply of commodities and factors are assumed to be dependent on prices. Figure 2 provides a schematic overview of the typical elements of a CGE model. A CGE model simulates the working of a market economy in which prices and quantities adjust to clear all markets. For example, households maximize their welfare, the government is assumed to have a balanced budget, and resources are limited and costly. Effectively, a CGE model

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

⁴⁴ This data was not necessarily collected initially for use with the RDVM in most cases (Bond et al. 2017).

⁴⁵ See Section 4.7 for further discussion.

specifies expected behavior of optimizing consumers and producers, as well as including community government (e.g., taxes) as an agent to capture transactions in circular flow of income (Robinson et al. 1990).46

Figure 2. Schematic of main components in a CGE model. Note that ROW refers to the "Rest of the World", that is the aggregation of all economic transactions between the selected region under consideration and those not within the selected region.



3.2. Comparative-Static or Dynamic CGE Models?

Many CGE models are comparative-static; they are used to model the reactions of the economy at only one point in time. In such cases, the model is interpreted as demonstrating the reaction of the economy in a future period to one (or more) external shocks, policy changes, and/or resilience planning efforts-in our application, the resilience dividend. That is, the results show the difference (usually reported in percent change form) between two alternative future states (with and without the resilience plan in place). The process of adjustment to the new equilibrium is not explicitly represented in such a model, as the temporal element of a CGE model is not well defined. But it is possible to distinguish between short-run and long-run equilibria (e.g., looking at whether capital stocks are allowed to adjust in a given run of the model).

By contrast, dynamic CGE models (e.g., Pereira and Shove 1988) explicitly trace each variable through time—often at annual intervals. These models are more (temporally) realistic than comparative-static models; however, the data requirements are greater and they are generally more challenging to construct and solve. Furthermore, in the case of resilience planning which already encompasses a great deal of uncertainty, they require that future estimations are made for all exogenous variables—not just those affected by the shock,

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

196

⁴⁶ The CGE model takes a Walrasian neoclassical general equilibrium approach—the main equations that need to maintain equilibrium are derived from constrained optimization of the neoclassical production and consumption functions. Producers operate at a level as to maximize profits (minimize costs). Production factors – labor, capital, and land – are paid in accordance with their respective marginal productivities. Consumers are assumed to be subject to budget constraints, but otherwise maximize their utility. At equilibrium, the model solution at equilibrium provides a set of prices to clear commodity and factor markets within the modelled community's economy (see Bandara 1991).

policy change, and/or resilience plan. Furthermore, consistency problems may arise because variables that change from one equilibrium period to the next may not be consistent with each other in the fixed period of change.

Thus, we propose using a comparative-static model set-up. In some cases, the data required for the CGE assessment of the resilience dividend (see Section 4) will be available in different years. Thus, creating a CGE model for a period before the resilience plan integration and another CGE model following the integration may be a realistic way to provide a dynamic understanding of the resilience dividend.47

3.3. Spatial CGE Models

SCGE models deal with distributive effects in a manner that makes a great deal of sense when dealing with resilience planning against large-scale shocks (e.g., natural disasters). To date, SCGE models have been used to assess economic impacts of infrastructure investments and policies, especially in the area of transportation (e.g., Ivanova et al. 2007 and Miyagi et al. 2006). Multi-regional input-output models are the closest relatives to SCGE models, but they are not able to fully capture price and quantity effects as they do not allow for substitution effects.

Thus, SCGE models are a natural fit for exploring the resilience dividend and the geographic distribution of the relative effects. In our discussion of data requirements and setting up the resilience dividend assessment we assume use of SCGE modelling.⁴⁸ Section 4 describes the specific data required to create a SCGE model to quantify the resilience dividend and determine distributive effects.

4. Data Required

4.1. Social Accounting Matrix (SAM)

The primary goal of data collection is to develop the social accounting matrix (SAM). A SAM can generally be described as "an organized matrix representation of all transactions and transfers between different production activities, factors of production, and institutions ... within the economy and with respect to the rest of the world" (Hirway et al. 2008). In short, it quantifies all cash flows between pertinent actors within an economy. The SAM serves as the core of the CGE analysis, as it defines the base relationships between sectors, households, labor markets, and other key actors in the economy that the CGE model uses to determine the impacts of policies and shocks. The World Bank (Round 2003) notes that there are three key features to a SAM: 1) they are square matrices, 2) they are comprehensive, including all economic activities of the system, and 3) they are flexible in how they may be disaggregated and what parts of the economic system are emphasized. The following sections provide available sources for acquiring the data needed to build a SAM based on the method for constructing a spatial SAM and CGE model developed in Cutler et al. (2017). The subsequent data sources are not the only ones available, but are more commonly used than others or are most capable of filling data needs. There is a comprehensive discussion of methods for SAMs and CGE models in Cutler et al. (2017). Example applications and case studies using CGE modeling can be found in Cutler et al. (2017) and Schwarm and Cutler (2003).

 $[\]frac{47}{10}$ This solution implies a retroactive study opposed to a perspective study in which the outcomes of the economy after the resilience

plan is enacted is completely unknown

⁴⁸ For more on SCGE, see Bröcker and Korzhenevych (2011).

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

4.2. Quarterly Census of Employment and Wages (QCEW) data

4.2.1. Summary

The quarterly census of employment and wages (OCEW) is a Bureau of Labor and Statistics (BLS) program that reports the guarterly count of employment and wages for employers, broken down at the industry (defined by the North American Industry and Classification System (NAICS) code) level and geographically at the county, Metropolitan Statistical Area (MSA), state, and national levels. The QCEW covers roughly 95 % of all U.S. jobs (BLS Website). This data is an excellent source to determine wage payments, employment, and number of firms by industry.

4.2.2. Challenges

Because the data contains commercially identifiable information (CII) and, potentially, personally identifiable information (PII), firm-level QCEW data is not publicly available and must be requested through an appropriate state or federal government agency. This process can be time consuming and may require payment to cover the cost of labor. There are also restrictions on how it can be used and reported, namely steps must be taken to mask any CII or PII. This is typically done through ensuring a minimum number of firms in each industry and making sure that no single industry has a large percentage of its data coming from one firm, regardless of how many firms are in the industry.

While one of the best sources of data for building a SAM, there are other ways to obtain the same information, though the data will typically be pre-aggregated to address CII and PII concerns and thus, less refined. The advantage of obtaining firm-level data is that a researcher can customize how the data is aggregated. In particular, the data can be aggregated with respect to sectors defined by the researcher, potentially breaking these out spatially using the establishment address. See Section 5.1 for more details.

4.3. LEHD Origin-Destination Employment Statistics (LODES) data

The Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES) data collected by the Center for Economic Studies at the US Census Bureau details employers, their employees, and the flow of jobs over time and space. This data allows for the mapping of labor flows between regions within and beyond the scope of a given CGE model. This data can be especially useful in modeling the commuting patterns of employees in and out of town as well as movement between a city's districts. The value of this data is in its ability to specify the transportation needs of the community under analysis and evaluate how that community would be impacted by various disaster scenarios or other shocks. For instance, the severity (measured in economic damages) of a hazard event that results in a bridge closure is likely to be informed by the extent to which the local community relies on that piece of infrastructure to commute to work or flee the ill-effects of the hazard in guestion.

4.4. Public Use Microdata Sample (PUMS) data

4.4.1. Summary

Public use microdata sample (PUMS) data is collected by the U.S. Census Bureau and reported at various intervals. The dataset relies on the use of American Community Survey (ACS) data. Unlike the decennial census, ACS surveys are yearly and not nationwide. Roughly one in thirtyeight households are invited to take the survey every year (U.S. Census Bureau 2017). The data collected in the ACS is very similar to the data collected during the decennial census. The household income distribution can be obtained from this dataset at varving geographic levels ranging from as large as the United States, down to ZIP code tabulated areas. The primary data set of interest from the PUMS data is the employment by sector and the aggregated wage payments by sector. These allow the SAM to differentiate between different labor groups.

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.
4.4.2. Challenges

Access to individual level data is not available without the permission of the U.S. Census Bureau due to the large amount of PII and its access and use is subject to severe restrictions. At present, getting access to Census data requires showing that the use of the data would benefit the U.S. Census Bureau in some way. This is not necessarily easy to prove and, even if access is given, may take a significant amount of time to obtain. If obtained, restrictions on use, where and how the data can be accessed, and how data can be reported add further barriers to use. The publicly available data through the ACS website comes with no restrictions, but is pre-aggregated in a way that may not match one-to-one with the way industries are defined in other datasets. This issue may or may not be important depending on how industries are aggregated in the SAM, but nevertheless it is the most readily available data. The use of microsample data means the given value is extrapolated from the subset of U.S. homes that took part in the survey. If the desired year happens to coincide with a decennial census, then the use of decennial census data is possible.

4.5. County and City Assessor data

4.5.1. Summary

The development of a CGE model involves the construction of an accurate snapshot of a specific economy at a given point in time so that the resulting model may be calibrated to represent the community under investigation. One key component of the CGE model is the accurate representation of the value of land and capital within the regional economy. Estimates of household expenditures on various classes of housing services for disaggregated groups of households is also a key attribute of the regional economy that must be modeled with the greatest level of fidelity possible. City and County Assessors offices collect, maintain, and make available to the public this information on the building stock within their respective political boundaries.

4.5.2. Challenges

The challenges inherent in working with public data are generally present when working with property tax assessment data. While very accessible, property tax assessment data is freely available for many communities, usually through the county assessor's office, it can and often does entail typographical errors that complicate the matching of the built environment to the businesses and residences therein. Missing data can be a problem for some variables reported in the property tax assessment data. There is considerable variance in the degree of detail and historical support of reported data across communities. Data may be reported in a manner that is not consistent across all years of interest to a given project. The classification and categorization of the built environment may change over time as data systems are improved and expanded.

4.6. City Budgets and the Comprehensive Annual Financial Report (CAFR)

4.6.1. Summary

Comprehensive Annual Financial Reports (CAFR) are documents containing details of the financial state of a given governmental entity such as a state or municipality. These documents are useful resources for the determination of local government tax revenue, expenditures, and employment. The CAFR provides the information necessary to decompose employment and expenditures into constituent government "industries"; education, public health, public safety, park and recreation, and others. This information is critical to efforts to properly size and disaggregate the government sector within the CGE model. CAFRs tend to be different across communities; one constant tends to be that the CAFR provides an excellent source of tax revenue and expenditures. The CAFR can also be a reliable source of data on the expenditure of federal funds in the local economy. Within some CAFRs, there is a Schedule of Expenditure of Federal Awards which lists information for each federal grant

awarded to the city (or other government entity) organized by the granting agency and program title.⁴⁹ This is a useful source of information for corroborating the timing of federal assistance programs that target disaster response among other pressing community needs. While the data in the CAFR on federal assistance may not be sufficiently disaggregated to model at the establishment or industry level, it is useful for ensuring the magnitude of relevant programs.

4.6.2. Challenges

The CAFRs are data-rich documents, but they generally contain information that must be reformatted or reorganized if it is to be of further use to the CGE modeler. While there are standards of presentation and content associated with the CAFRs, the exact format of the reports can differ over time, complicating long-term trend analysis.⁵⁰ It is possible that the CAFR for any single year may include federal grants that are only present in that year. Care should be taken to avoid treating grant awards as recurring components of local government finance within the CGE model

4.7. Bureau of Economic Analysis (BEA) data

4.7.1. Summary

Bureau of Economic Analysis (BEA) data is vital in building the SAM. The BEA data set provides the necessary tables to determine I-O coefficients and the values required to develop the relationship between investment and the stock of capital. The I-O data is generally taken at the national level and, in its raw form, gives the raw dollar amounts of input from each industry and the total output from each industry. These values can be used to determine I-O coefficients, which represent how much input each industry requires from every other industry in order to produce a dollar's worth of output. I-O coefficients define the flow of money between industries, and thus the linkages between industries necessary for the CGE model to determine how impacts on one industry flow to another.

The data for the investment capital linkage (CAPCOM) matrix comes from the BEA "Capital Flow" data. This data tracks the investment in new structures, equipment, and software by using industries. In essence, it measures how many commodities a specific industry purchases for investment from another industry. Like the I-O data, the CAPCOM tracks the interdependencies between industries; however, it focuses on new investments instead of required input. The raw data is taken from the I-O commodity categories (as opposed to the National Income and Product Account categories), which are in terms of producers' prices.

Other useful data from the BEA includes the BEA employment estimates and the BEA income estimate, which are available at varying geographical levels. While other datasets offer data on these values that are better suited for use in the SAM, the BEA estimates provide a useful check for their totals.

4.7.2. Challenges

As the BEA data is derived from multiple sources using CII, including the U.S. Census Bureau, all publicly available data is pre-aggregated, meaning industry classifications may not match one-to-one between other data sets. The more detailed underlying data is subject to the similar requirements for access, and restrictions on use, as mentioned for the ACS data.

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

200

⁴⁹ A CAFR may not include a Schedule of Expenditure of Federal Awards if the total amount of federal awards expenditures by a non-federal entity is less than \$750 000 dollars in the reporting year. Title 2 U. S. Code of Federal Regulations Part 200, Uniform administrative requirements, cost principles, and audit requirements for Federal awards provides the guidance on whether or not a CAFR must contain an audit of the expenditure of federal grant awards.

Accounting and financial reporting standards for state and local governments are codified by the Government Accounting Standards Board (GASB)

4.8. Informal data from community leadership and agencies

4.8.1. Summary

Depending on the exact research question and scope of the resilience dividend, the community being studied itself may prove to be an invaluable source of information. In the context of resilience, local officials can offer a unique and comprehensive perspective on the impact of a natural disaster on the community. Conversations with the City Manager's Office, Emergency Management, and (public or private) Economic Development teams can reveal priorities with respect to both the immediate response to a disruptive event type faced by the community, as well as short- and long-term recovery efforts and community goals. For instance, while a researcher may be aware that a community is investing in flood resilience, it is not obvious to an outsider where and how a community is investing its resources. Moreover, community officials can help a researcher compile a more complete picture of funding sources, both private and public.

Conversations with community officials can also provide perspective on local economic trends and goals, both irrespective of the potential disaster and specific to the disaster occurrence. While official data may suggest that manufacturing is an important sector to a community, the community itself may emphasize information technology as a growing sector being targeted with economic incentives such as tax breaks. Moreover, the community can provide insight into regional trends. For instance, business improvement districts may be integral to long-term community resilience. Certain neighborhoods may be of particular interest to a community (e.g., revitalization of downtown commerce). Such trends may inform the modeling step, in terms of how a researcher defines the productive sectors—especially spatially—and consequently the aggregation of official data for constructing the SAM.

4.8.2. Challenges

While the information gathered from conversations with community officials comes from authoritative sources, the "data" collected is informal. Incorporating the array of information into constructing a CGE model is less about collecting input data and more about guiding research direction. The biggest challenge arises from knowing what to ask. As an outsider, a researcher may have preconceived notions of what issues matters most, and community officials may be more than happy to answer questions about such issues. It is important to remember that what matters most to a community may differ from what a researcher thinks matters most. Gaining an understanding for a community's priorities can provide the proper context for analyzing a community with a CGE model. Moreover, it is important to keep in mind that not all communities may be organized enough to provide the necessary data, and some may be reluctant to the idea of providing the information. Even when community officials are willing to share information, they may be constrained by regulations, budget, or time.

4.9. Third Party data

4.9.1. Summary

If other data sources are not viable for use in the SAM, third party data may also be used. Third party datasets typically will provide the requested data aggregated as requested for a fee. Impact Analysis for Planning (IMPLAN) data (IMPLAN Group LLC) is a commonly used third party dataset derived for economic analysis. Their data includes premade SAMs at the national, state, and county level that can be augmented by the user with different data or relationships (RESI 2006). Other datasets are available, for example Thomson Reuters (Thomson Reuters 2015) and FactSet (FactSet 2017).

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

4.9.2. Challenges

Due to the proprietary nature of third party datasets, it is impossible to know all of the details of how the data were developed. While companies do describe processes and underlying sources, they invariably do not include everything in order to preserve any business advantages they might possess. The data also must be purchased and the fees may be prohibitive depending on the nature of the analysis and the party or parties required to purchase it. Care also must be taken to ensure that the data available from the third party is the actual data required.

4.10. Geographic Data

4.10.1. Summary

Economies have long been modeled as systems disembodied from their physical components. The increased adoption of geographic information systems (GIS) by firms and government entities allows for the spatial disaggregation of economic data with location records. Geographic data enables the introduction of explicit spatial considerations into the CGE model, which brings it towards an SCGE model. It is reasonable to assume that similar shocks may propagate through an economy in patterns that are informed by the topology of the built environment and regional geography. In many cases, data used for the CGE model includes spatial identifiers such as street address. GIS tools such as geocoders that produce longitude and latitude coordinates when fed address information, allow for the geolocation of individual business establishments and residences. In addition to matching firms and parcels, geocoding is instrumental to the process of defining the districts into which the local economy is divided. Once the geographic coordinates of each parcel are obtained, the parcels can be plotted and sorted into their districts using ArcGIS software. The importance of spatial linkages to overall impacts from a hazard may differ with the economy and hazard in question. There is a fundamental tradeoff between increased spatial disaggregation using GIS data and reduced complexity within the SAM. Establishing a distinct district for each establishment or residence would intractably complicate the SAM. Neglecting to incorporate any spatial information into the SAM may aggregate contravening trends, delivering results that mask important underlying trends in economic growth and hazard recovery.

4.10.2. Challenges

The fundamental challenge of working with GIS data is rooted in its variable quality and availability. GIS data may be missing for some public records and can be difficult to extract from data with messy variable coding. Improperly assigning establishments to the wrong district, as a result of bad address data, could impact the validity of a spatial CGE model. Different geocoding tools can produce geographic coordinates for the same record that disagree by small or large distances. The judgement calls that must be made to render this GIS data usable may ultimately be unjustifiable. Furthermore, GIS data can be inherently identifying when merged with other sources of data. Care must be taken when working with GIS data to avoid the unintended disclosure of CII and PII.

5. Methodology

5.1. Combining the Data

Combining the data from Section 4 into the SAM offers several benefits. First, in many cases it is often necessary. No single data set from Section 4 contains all of the required data for the SAM, with the possible exception of that provided by a third party vendor. Second, all of the data can be verified by the model builder. Moreover, each dataset can be verified independently by the model builder and better tailored to particular assumptions. Using third party data limits how much verification and customization is available for the analysis. Most

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

202

third-party datasets are heavily vetted; however, it can be beneficial to be able to check the underlying data. Third, working with the data directly can allow further insights outside the original scope of the model. Trends may appear in one dataset that wouldn't be visible in working with only the final SAM.

While there are benefits to using multiple sources of data, the use of the varied sources in Section 4 can create challenges when folding them into the final SAM. One example of this complication is attempting to derive the I-O and CAPCOM data at the PUMS sector level using the PUMS defined industry codes. The BEA and PUMS data sets are both based on NAICS codes; however, they aggregate those NAICS codes into larger industry categories that do not match one-to-one with each other. If the industries are broadly defined then this is not necessarily an issue. For instance, if manufacturing industry data is provided without disaggregation, then the industry codes from the PUMS data and BEA data, while different, still fall entirely within the larger aggregated manufacturing sector. If manufacturing industry data is disaggregated, then there is no guarantee that the each PUMS industry code will have a corresponding BEA code, or codes, that match in terms of NAICS codes covered. In such cases a fuzzy match is required which will possibly lead to a NAICS code from a sector not in a specific PUMS industry code being in the IO table for that PUMS industry code due to the inconsistency. The alternative version where a PUMS industry code loses a corresponding NAICS code is also possible.

Another challenge comes from datasets not necessarily covering the same geographical area. For instance, the smallest division of the county assessor may be at the city level, while the smallest division in the PUMS data may be at the MSA level. In such cases, it may be necessary to scale numbers up or down based on some distribution of relevant data, to get geographic areas to match up. An example of this would be scaling down MSA level data for industry specific employment by labor group down to the city level based on the known distribution of industry employment in the city.

Using different data sets means the totals for some values obtained for the same geographic area, such as total employment, should be the same (if all data were perfect), but end up being different between data sets. Such differences are to be expected between data sets, as differences in what is or is not included and methods may end up resulting in different estimates. Still, the CGE requires consistency between key values in order to balance the SAM and run analysis. Similar to the situation of differing geographic areas, scaling numbers up or down to match may be required. However, these differences between data sources under such circumstances should be relatively close. Otherwise, there may be an unaddressed issue with the data.

Spatializing the SAM adds further complications. One issue that arises is the need to match industry level data to the spatialized components. This process is meant to allot the capital land value from the county property tax assessment data and the QCEW employment and wages to the appropriate industry sector. Matching on addresses is known to be a nontrivial task, as abbreviations, misspellings, date entry errors, and other consistency problems make getting the desired match difficult. Address standardization and fuzzy matching can alleviate this, but typically does not fully address the issue. The other complication with using the OCEW data in this context is that there are requirements on making sure all data is aggregated to the point that CII becomes masked. This is typically achieved by ensuring every industry has a minimum number of firms included to make it impossible to trace back the information to a specific firm. This means that industry sectors may need to be aggregated into larger sectors if they contain too few firms.

Spatialization also complicates the entry of sector related data into the SAM. Ordinarily, industries are assumed to be in the area of study and that is all. Spatialization divides industries into sub-regions of the study area. This means labor, households, capital value, and the I-O and CAPCOM tables need to reflect this division. For the I-O table, it can be done fairly simply if one assumes that the firms in any sub-region are essentially the same as the firms in the larger area. Under this assumption, all I-O coefficients are identical to the non-spatialized I-O table for every industry. Otherwise, effort must be put into understanding how firms differ in terms of inputs and outputs in each sub-region. The spatialized CAPCOM

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

can be obtained by determining a distribution of investments based on available data, for instance, the distribution of workers, firms, or wages for all sub-regions, and distributing them accordingly.

5.2. CGE Coverage of the Resilience Dividend

The ultimate goal of the proposed SCGE modeling method is to guantify the resilience dividend. Therefore, it is important to understand what the SCGE model can and cannot quantify. A CGE model provides distributional impacts of shocks, policy changes, and the current status of the region. Distributional impacts allow the analyst to understand not only the overarching net impacts, but to whom and where those impacts fall and are distributed. Large economic effects will be easily discerned and the impacts can be selected to see how different scenarios may have played out in the region. Any effects of resilience actions that have co-benefits can be modeled to identify how those co-benefits manifest themselves throughout the economy and where they go. Thus, the resilience dividend can be quantified as a grand total, as well as determining who gets these benefits and where they go spatially. SCGE models may not capture the entirety of the resilience dividend in many cases. Nonmarket benefits that never actually materialize as real cash flows are not necessarily captured. Minor impacts may also be lost as the overall economic conditions may overwhelm them

5.3. Additional Considerations

There are additional considerations that are important when using a SCGE model to quantify the resilience dividend. There are limitations to the CGE approach and full assessment of the resilience dividend may be best achieved using CGE methods in tandem with other economic methods.

5.3.1. The use of two CGE models

One critique of CGE models is that they are unable to fully capture the dynamics of an economy's response to a shock. Whether a community responds in acute fashion or slowly over a longer time period can have a considerable influence on the impacts of a given shock. The speed and persistence of a shock may be more informative than its magnitude. An advantage of having cross-sectional and panel data for a community is that response trajectories generated without full time specification can be calibrated using trends observed in the temporal data. The speed of recovery is of interest to communities considering their various options. The use of multiple CGE models may facilitate the corroboration of findings across approaches. Furthermore, if static CGE models are built using different baseline years that coincide with periods before and after a hazard event of interest, it is possible to see how the economy's response to an unrelated shock has changed over time. Of course, care must be taken to avoid the Post hoc ergo propter hoc fallacy if one is to employ two CGE models timed to before and after a hazard event. It is quite possible that other important structural changes are occurring simultaneously with the hazard.

5.3.2. Net Present Value, The EDGeS Tool, and CGE

As the CGE methodology uses I-O data there is some debate as to how time plays in CGE models. I-O tables generally represent a snapshot in time. However, CGE models use them to obtain the equilibrium following shocks to a system. How long it takes to reach that equilibrium after a shock is not a simple question to answer. In that regard, the time varying nature of the transition period from base state to post-shock state is currently difficult to model.

On the other hand, if one views the post-resilience action equilibrium as the base state for one case and the pre-resilience equilibrium as the base state for another case, it is possible to use the Economic Decision Guide (EDG) (Gilbert et al. 2016) methodology to examine these options based on Net Present Value (NPV). If a shock representing a disaster of an

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

assumed magnitude is applied to both cases, the on-event indirect losses required using the EDG methodology can be obtained. Direct losses, such as structural losses, and response and recovery losses, such as temporary shelters, would need to be added in separately. The non-event related benefits can be estimated by examining the two cases' base states, with non-market benefits and externalities added separately from the CGE analysis, assuming these are not impacted by disaster related shock. The costs for each case should be known, thus all inputs required for the Economic Decision Guide Software (EDGeS) Tool (Helgeson et al. 2017) should be available.

6. Next Steps

The next step in this process is to complete an SCGE model based upon a community that has made changes based on resilience planning against a natural hazard event. A flooding event was chosen for the initial case study. Flood situations cannot be entirely prevented, but steps can generally be taken to prevent and minimize loss of property, interruption of business, and loss of life. Furthermore, floods are a leading cause of death from natural disasters in the United States. Flood-related fatalities are reported around 200 per year with about half caused directly by individuals attempting to drive through flood waters (Ashley and Ashley 2008).

Given the uncertain nature of most hazard events, in terms of timing, magnitude and path, we find flooding to be one disturbance event that may be more predictable than are other events, at least in terms of areas potentially affected (i.e., within a flood plain). Flood situations are variable and are often a by-product of other natural hazards, such as hurricanes. But there are instances when floods are standalone disturbance events (e.g., snowmelt, severe thunderstorms, prolonged rains) versus a bi-product or co-consequence of other disturbance events. In such cases of flood as a singular event, a geographic area in a given community may be affected more than other areas given soil, height above sea level, and flood protections. This is the case in general for Cedar Rapids, Iowa and the community's flooding events (p.c. S. Fowler, 13 March 2017).

We are in the process of finalizing the construction of the SAM for Cedar Rapids, Iowa with consideration for Linn County, Iowa. To date, we have collected detailed data for each category noted in Section 4 of this paper. This community made a number of deliberate choices in terms of zoning, retrofit construction, and new construction in the period since the major flood event of September 2008. Tate et al. (2016) assess the government buy-out process undertaken in Cedar Rapids. There are a number of additional projects that have now had time to mature since 2008, such as the revitalization of the downtown district and development of the McGrath Amphitheater (p.c. H. Stiffler 27 June 2017) that can be assessed using SCGE modeling to understand the full resilience dividend and distributional effects throughout the Cedar Rapids economy.

In turn, these findings may be compared to estimates of ROI and the NPV metrics calculated for the projects at the initial time of development and when the choice was made for which projects to take-on and further develop.

We are aware that the SCGE process is data-driven and unique to each community and its associated economy, the disturbances faced and the resilience options available (e.g., subject to budget constraints, social factors, etc.). It is clear that the SCGE resilience dividend quantification methods discussed in this paper may be better suited for the developed country context because of the extensive data requirements. Yet, once the methodology is demonstrated in a case study, it may be possible to assess the level of specificity required in the data to obtain meaningful estimates of the resilience dividend.

Citations

Ashley, S. and W.S. Ashley. 2008. Flood Fatalities in the United States, Journal of Applied Meteorology and Climatology, 47, 805-818.

Bandara, SJ. (1991) An investigation of "Dutch Disease" economics with a miniature CGE model'. Journal of Policy Modeling 13(1), 67-92.

Bandyk, M. Why Natural Disasters Are More Expensive-But Less Deadly Haiti and Chile guakes show importance of economic growth. U.S. News. March 24, 2010, Available at: https://money.usnews.com/money/ business-economy/articles/2010/03/24/why-naturaldisasters-are-more-expensivebut-less-deadly

BLS Website. Quarterly Census of Employment and Wages. Retrieved from <https://www.bls.gov/cew/> on September 7, 2017

Bond, Craig A., Aaron Strong, Nicholas Burger and Sarah Weilant. Guide to the Resilience Dividend Valuation Model. Santa Monica, CA: RAND Corporation, 2017. https://www.rand.org/pubs/research_reports/RR2130.html.

Bröcker, Johannes and Korzhenevych, A., 2011. "Forward looking dynamics in spatial CGE modelling," Kiel Working Papers 1731, Kiel Institute for the World Economy (IfW).

Cutler, H., Shields, M., Tavani, D., & Zahran, S. (2016). Integrating engineering outputs from natural disaster models into a dynamic spatial computable general equilibrium model of Centerville. Sustainable and Resilient Infrastructure, 1(3-4), 169-187.

Cutler, H., J. Davis, Y. Hu, K. Kakpo, S. Mckee, M. Shields and S. Zahran (2017). "Developing a Methodology to Build Spatial SAMs and CGE Models. White Paper, Colorado State University.

EM-DAT (25th January 2016) : The OFDA/CRED -International Disaster Database www.emdat.be Université catholique de Louvain Brussels – Belgium

FactSet (2017). Third-Party Economic Databases. Retrieved from < https://www.factset.com/data/market_ data/economic > on September 7, 2017

Fung, J.F. and J.F. Helgeson, 2017. Defining the Resilience Dividend: Accounting for Co-benefits of Resilience Planning. NIST Technical Note, 1959. Available at: http://nvlpubs.nist.gov/nistpubs/TechnicalNotes/NIST. TN.1959.pdf.

Gilbert, S.W., D. Butry, J. Helgeson, and R. Chapman, 2016, Community Resilience Economic Decision Guide for Buildings and Infrastructure Systems. NIST Special Publication, 1197. Available at: http://dx.doi.org/10.6028/ NIST.SP.1197.

Helgeson, J.F., S. A. Grubb, D.H. Webb, 2017. The Economic Decision Guide Software (EDGeS) Tool: User Guidance. NIST Technical Note, forthcoming.

Hirway, I., M.R. Saluja, B. Yadav (2008). The impact of public employment strategies on gender equality and pro-poor economic development. The Levy Economics Institute of Bard College, New York

IMPLAN Group LLC, IMPLAN System, 16905 Northcross Dr., Suite 120, Huntersville, NC 28078 www.IMPLAN.com

Ivanova, O., Heyndrickx, C., Spitaels, K., Tavasszy, L., Manshanden, W., Snelder, M., Koops, O.: RAEM Version 3.0. Transport Mobility Leuven, Leuven (2007).

Kunreuther, H. and E. Michel-Kerjan (2009). At War with the Weather. Cambridge, MA: MIT Press.

Leontief, W. (1936). Quantitative Input-Output Relations in the Economic System of the United States.

Review of Economics and Statistics, 18(3), 105-125.

Leontief, W. (1941). The Structure of American Economy, 1919-1929. New York: Oxford University press.

Reprinted, White Plains: International Arts and Sciences Press 1976

Mechler, R., J. Mochizuki, and S. Hochrainer-Stigler, 2016. Disaster Risk Management and Fiscal Policy:

Narratives, Tools, and Evidence Associated with Assessing Fiscal Risk and Building Resilience., p.39.

Meng S. and M. Siriwardana (2017) An Introduction to CGE Modelling. In: Assessing the Economic Impact of Tourism. Palgrave Macmillan, Cham.

T. Miyagi, M. Nishimura and S. Mitui. 2006. Measuring indirect economic impacts arising from transportation investment by a SCGE model. WIT Transactions on The Built Environment, Vol. 89, WIT Press.

Pereira, A.M. and J.B. Shove, (1988). Survey of dynamic computable general equilibrium models for tax policy evaluation. Journal of Policy Modeling, 10, 401-436.

RESI (2006). What is IMPLAN? Towson University, White paper

Robinson, S., M. Kilkenny, and K. Hanson, The USDAIERS Computable General Equilibrium Model of the United States, Washington, DC: U.S. Department of Agriculture, Economic Research Service, staff report no. AGES 9049, June 1990.

Rodin, J., 2014. The Resilience Dividend: Being Strong in a World Where Things Go Wrong, PublicAffairs.

Round, J. (2003). Social accounting matrices and SAMbased multiplier analysis. The impact of economic policies on poverty and income distribution: Evaluation techniques and tools 261-276

Sadoulet E. and A.de Janvry (1995) Quantitative Development Policy Analysis. Baltimore: Johns Hopkins University Press.

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

Schwarm, W. and H. Cutler. (2003) Building small city and town SAMs and CGE Models, Review of Urban & Regional Development Studies. 15(2), 132-46.

Schultz, J. & Elliott, J.R. "Natural disasters and local demographic change in the United States." Popul Environ (2013) 34: 293.

Strömberg, D. "Natural Disasters, Economic Development, and Humanitarian Aid." Journal of Economic Perspectives—Volume 21, Number 3—Summer 2007— Pages 199 –222

Tanner, T. et al., 2015. The Triple Dividend of Resilience: Realising development goals through the multiple benefits of disaster risk management, Overseas Development Institute, The World Bank.

Tanner, T., S. Surminski, E. Wilkinson, R. Reid, J. Rentschler, S. Rajput. The triple dividend of resilience, December 2016

Tate, E. A. Strong, T. Kraus, H. Xiong, 2016. Flood recovery and property acquisition in Cedar Rapids, Iowa. Nat Hazards (2016) 80:2055-2079. DOI 10.1007/s11069-015-2060-8

Thompson Reuters (2015). Thompson Reuters Economic Data. Retrieved from < https://financial.thomsonreuters. com/content/dam/openweb/documents/pdf/financial/ economics-data.pdf> on September 7, 2017

U.S. Census Bureau (2017). American Community Survey: how it works for your community. Retrieved from < https://www.census.gov/content/dam/Census/ programs-surveys/acs/about/2017%20How%20 the%20ACS%20Works%20for%20Your%20 Community_508.pdf> on September 7, 2017.

Uniform administrative requirements, cost principles, and audit requirements for Federal awards, 2 C.F.R. § 200 (2014).

Wu, H., R. F. Adler, Y. Tian, G. J. Huffman, H. Li, and J. Wang (2014), Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model, Water Resour. Res., 50, 2693.2717, doi:10.1002/2013WR014710.

Wu H., R. F. Adler, Y. Hong, Y. Tian, and F. Policelli (2012), Evaluation of Global Flood Detection Using Satellite-Based Rainfall and a Hydrologic Model. J. Hydrometeor, 13 1268 1284

Wu H., J. S. Kimball, H. Li, M. Huang, L. R. Leung, R. F. Adler (2012), A new global river network database for macroscale hydrologic modeling, Water Resour. Res., 48, W09701, doi:10.1029/2012WR012313.

Wu, H., J. S. Kimball, N. Mantua, and J. Stanford (2011), Automated upscaling of river networks for macroscale hydrological modeling, Water Resour. Res., 47, W03517, doi:10.1029/2009WR008871.

Cutler, Harvey; Fung, Juan; Helgeson, Jennifer; O'Fallon, Cheyney; Webb, David. "Identifying and Quantifying the Resilience Dividend using Computable General Equilibrium Models: A Methodological Overview." Paper presented at 2nd International workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, Ispra, Italy. December 13, 2017 - December 15, 2017.

Automatic Localization of Casting Defects with Convolutional Neural Networks

Max Ferguson Engineering Informatics Group Civil and Environmental Engineering Stanford University Stanford, United States maxferg@stanford.edu

Ronay Ak Systems Integration Division National Institute of Standards and Technology (NIST) Gaithersburg, United States ronay.ak@nist.gov

Yung-Tsun Tina Lee Systems Integration Division National Institute of Standards and Technology (NIST) Gaithersburg, United States yung-tsun.lee@nist.gov

Kincho H. Law Engineering Informatics Group Civil and Environmental Engineering Stanford University Stanford, United States law@stanford.edu

Abstract—Automatic localization of defects in metal castings is a challenging task, owing to the rare occurrence and variation in appearance of defects. Convolutional neural networks (CNN) have recently shown outstanding performance in both image classification and localization tasks. We examine how several different CNN architectures can be used to localize casting defects in X-ray images. We take advantage of transfer learning to allow state-of-the-art CNN localization models to be trained on a relatively small dataset. In an alternative approach, we train a defect classification model on a series of defect images and then use a sliding classifier method to develop a simple localization model. We compare the localization accuracy and computational performance of each technique. We show promising results for defect localization on the GRIMA database of X-ray images (GDXray) dataset and establish a benchmark for future studies on this dataset.

Keywords — Casting Defect Detection, Defect Localization, Convolutional Neural Network, Computer Vision

I. INTRODUCTION

Quality control is an important aspect of modern manufacturing processes [1]. With an increased level of competition in the manufacturing market, manufacturers must increase their production rate while maintaining stringent quality control limits. In order to meet the growing demand for high-quality products, the use of intelligent visual inspection systems is becoming essential in production lines.

The casting process can often introduce defects in the product which are detrimental to the final product quality [2]. The early detection of these defects can allow faulty products to be identified early in the manufacturing process, leading to time and cost savings [3]. It is beneficial to automate quality control where possible to ensure consistent and cost-effective inspection, especially if defect detection is to be carried out at many points along the production line. The primary drivers for automated inspection systems include faster inspection rates, higher quality demands, and the need for more quantitative product evaluation that is not hampered by the effects of human fatigue.

There are a number of nondestructive examination (NDE) techniques available for producing two-dimensional and threedimensional images of an object. Real-time X-ray imaging technology is widely used in defect detection systems in industry, such as on-line weld defect inspection [3]. Ultrasonic inspection and magnetic particle inspection can also be used to measure the size and position of casting defects in cast components [4, 5]. An alternative method is three-dimensional X-ray computed tomography, that can be used to visualize the internal structure of materials. Recent developments in high resolution X-ray computed tomography have made it possible to gain a three-dimensional characterization of porosity [6, 7].

The defect detection process can be framed as either an image segmentation problem or an object localization problem. In the image segmentation approach, the problem is essentially one of pixel classification, where the goal is to classify each image pixel as a defect or not. In the object localization approach, the goal is to place a tight-fitting bounding box around each defect in the image. As a related task, defect classification can then be performed to classify different types of defects. Common casting defects include air holes, foreign-particle inclusions, shrinkage cavities, cracks, wrinkles and casting fins [8]. Some defect detection systems attempt to locate and classify defects, but, in this study, we choose to focus purely on defect localization.

In this work, we develop a fast and accurate casting defect localization system, utilizing recent advances in computer vision. More specifically, we develop a model that can localize casting defects in two-dimensional X-ray images. We train and evaluate the proposed casting defect localization models using the GRIMA database of X-ray images (GDXray) dataset, published by Grupo de Inteligencia de Máguina (GRIMA) [9]. The GDXray Castings subset contains 2727 X-ray images of metal parts, many of which contain casting defects.

Object localization is a widely studied but difficult problem in computer vision [10]. In many tasks, including defect detection, the aspect ratio, size, location and the number of objects in each image are typically unknown variables. There is

Ak, Ronay; Ferguson, Max; Law, Kincho; Lee, Yung-Tsun. "Automatic Localization of Casting Defects with Convolutional Neural Networks." Paper presented at 2017 IEEE International Conference on Big Data (BigData 2017), Boston, MA, United States. December 11, 2017 -

not a large amount of literature describing the application of modern computer vision techniques for casting defect localization [11, 12]. It has been suggested that many advances in defect detection remain proprietary [3].

Most object detectors contain two important components: a feature extractor and an object classifier. In traditional methods, the feature extractor is usually a hand-engineered module, such as histogram of oriented gradients (HOG) or local binary patterns (LBP). The classifier is commonly a support vector machine (SVM) or a non-linear boosted classifier [3]. In more modern object detectors, the feature extractor and object classifier are often replaced by neural networks (NNs).

Recently, deep neural networks (DNNs) have attained impressive performance in many fields such as image classification, object detection, and semantic segmentation [13-16]. The performance of these deep networks has begun to exceed human performance in many tasks. For example, the human top-5 classification error rate on the large scale ImageNet dataset has been reported to be 5.1 %, whereas a state-of-the-art NN achieves a top-5 error rate of 3.57 % [13, 17]. The top-5 error rate indicates the fraction of the test images for which the correct label is not among the five labels considered most probable by the classification model. Large performance improvements have also been realized by training deep convolutional neural networks (CNN) for object detection [14, 15, 18]. While many NN architectures have been proposed for the object localization task, there is still no consensus on the best method. Therefore, we choose to train and compare a number of different architectures on the defect detection task. To summarize, our main contributions are as follows:

- We provide a concise summary of modern casting defect detection algorithms.
- We adapt several modern object detection networks to the casting defect task. We use these networks to conduct several experiments that explore the speed/accuracy tradeoff.
- We demonstrate that transfer learning can be used to improve defect localization on the relatively small GDXray dataset.

To the best of our knowledge, this study presents the first benchmark results for defect localization using the publicly available GDXray dataset.

RELATED WORK П

The detection and localization of casting defects with traditional computer vision techniques is a well-studied task. One popular method is background subtraction, where an estimated background image (which does not contain the defects) is subtracted from the preprocessed image to leave a residual image containing the defects and random noise [19-21]. However, background subtraction tends to be very sensitive to the positioning of the image, as well as random image noise. A range of matched filter techniques have also been proposed, with modified median (MODAN) filtering being a popular choice [22]. The MODAN-Filter is a median filter with adapted filter masks, that is designed to differentiate structural contours of the casting piece from casting defects

[23]. A number of other authors have proposed wavelet-based techniques with varying levels of success [8, 24]. Interestingly, the convolutional feature extraction technique proposed in this work can be seen as a generalization of many traditional filterbased techniques, where instead of specifying the filter type a priori, we attempt to learn the optimum filters [25].

Feature-based detection is another useful technique, where each image pixel is classified as a defect or not depending on features that are computed from a local neighborhood around the pixel. Common features include statistical descriptors (mean, standard deviation, skewness, kurtosis) and localized wavelet decomposition [8]. Selecting a threshold value to determine if a pixel is part of a defect region remains a large challenge of this method, although adaptive threshold methods have been proposed [26]. Several fuzzy logic approaches have also been proposed, but these techniques have been largely superseded by modern NN-based computer vision techniques [27].

Many state-of-the-art object detection systems developed from the region-based convolutional neural network (R-CNN) architecture [28]. R-CNN creates bounding boxes, or region proposals, using a process called selective search. At a high level, selective search looks at the image through windows of different sizes, and for each size tries to group together adjacent pixels by texture, color, or intensity to identify objects. Once the proposals are created, R-CNN warps the region to a standard square size and passes it through a feature extractor. A SVM classifier is then used to predict what object is present in the image, if any. In more recent object detection architectures, such as region-based fully convolutional networks (R-FCN) each component of the object detection network is replaced by a DNN [18].

III. OBJECT DETECTION ARCHITECTURES

Our main approach draws on recent successes of CNN for image feature extraction and object localization [13-15, 17]. We focus primarily on applying three recent object detection architectures to the casting defect detection task, namely: Faster R-CNN [14], R-FCN (Region-based Fully Convolutional Networks [18]) and SSD (Single Shot Multibox Detector [15]). While these architectures were originally presented with a particular feature extractor, we review each method using both the Visual Geometry Group (VGG) and Residual Network (ResNet) type feature extractors. This allows us to make a fair comparison between the performance of object detection networks on the casting defect localization task.

At a high level, all three architectures consist of a single convolutional network, trained with a mixed regression and classification objective. Each architecture defines a preset list of anchor boxes at different spatial locations in the image. These anchors often vary in aspect-ratio and scale, such as to contain any potential object in the image. The model is trained to make two predictions for each anchor: the offset between the anchor and the ground truth bounding box, and the class of the object inside the anchor, if any.

Ak, Ronay; Ferguson, Max; Law, Kincho; Lee, Yung-Tsun. "Automatic Localization of Casting Defects with Convolutional Neural Networks." Paper presented at 2017 IEEE International Conference on Big Data (BigData 2017), Boston, MA, United States. December 11, 2017 -

Each of the object detectors described in this work minimize a combined classification and regression loss, that we now describe. For each anchor, a, the best matching defect bounding box *b* is selected. The precise method for selecting the bounding box depends on the object detection architecture [14, 15, 18]. If such a match is found, then we assume a contains a defect and we assign it a ground-truth class label $y_a = 1$. In this case, we also assign a vector encoding of box b with respect to anchor a, denoted $\phi(b_a; a)$. If no match is found, we assume that a does not contain a defect and we set the class label $y_a = 0$. The location-based loss for a is expressed as a function of the predicted box encoding $f_{loc}(I; a, \theta)$ and $\phi(b_a; a)$, where I is the image and $\boldsymbol{\theta}$ is the model parameters. The classification loss is expressed as a function of the predicted class $f_{class}(I; a, \theta)$ and y_a . The total loss for a is expressed as the weighted sum of the location-based loss and the classification loss [29]:

$$L(a, \mathbf{I}; \boldsymbol{\theta}) = \alpha \cdot y_a \cdot l_{loc} \big(\phi(b_a; a) - f_{loc}(\mathbf{I}; a, \boldsymbol{\theta}) \big) + \beta \cdot l_{class} \big(y_a, f_{class}(\mathbf{I}; a, \boldsymbol{\theta}) \big),$$
(1)

where α , β are weights chosen to balance localization and classification losses [30]. The location loss function l_{loc} captures the distance between the true location and the estimated one. We use the Smooth L₁ loss function as l_{loc} with all architectures, even though the L2 loss function was originally proposed in [15, 30]. The classification loss l_{class} measures the difference between the predicted class and the true class, for each anchor. We use the cross-entropy loss function for l_{class} [17]. Table I summarizes the box encoding and loss functions used with each architecture. To train the object detection model, (1) is averaged over the set of anchors and minimized with respect to parameters $\boldsymbol{\theta}$.

A. Faster R-CNN

Faster R-CNN is a state-of-art object detection system composed of two modules. The first module is a deep fullyconvolutional neural network that proposes regions, and the second module is a feed-forward neural network that attempts to classify the objects in each region [14]. We will refer to the first module as the region proposal network (RPN) and the second as the region-based detector (RBD). The RBD makes two predictions for each region: a discrete class prediction for each region, and a continuous prediction of an offset by which the region needs to be shifted to best fit the ground truth bounding box.

Unlike [14], we choose to decouple the feature extractor from the RPN. This allows us to review the performance of R-CNN with two different feature extractors. This also makes it much easier to initialize the feature extractor with pretrained weights, especially when using standardized feature extractors such as VGG-16 or ResNet-101 [17, 31].

We now describe the configuration of the Faster R-CNN object detector used in this work. We configure the RPN to generate 300 region proposals for each image, as in [14].

TABLE I.	LOSS AND BOX ENCODING FUNCTIONS FOR EACH
	ARCHITECTURE, AS USED IN THIS WORK.

Architecture	$\begin{array}{c} \textit{Box Encoding}^{a} \\ \boldsymbol{\phi}(\boldsymbol{b}_{a}; \boldsymbol{a}) \end{array}$	Class Loss l _{class}	Loc. Loss l _{loc}
Faster R-CNN	$\left[\frac{x_c}{w_a}, \frac{y_c}{h_a}, \log w, \log h\right]$	Cross Entropy	Smooth L_1
R-FCN	$\left[\frac{x_c}{w_a}, \frac{y_c}{h_a}, \log w, \log h\right]$	Cross Entropy	Smooth L_1
SSD	$[x_0, y_0, x_1, y_1]$	Cross Entropy	Smooth L_1

Boxes are encoded with respect to a matching anchor a via a function ϕ , where $[x_0, y_0, x_1, y_1]$ are the min/max coordinates of the box, x_c , y_c are its center coordinates, and w, h are its width and height. The terms w_a and h_a denote the width and height of the matching anchor, and log is used to denote the natural logarithm

Anchor boxes are generated using 1:1, 1:2, and 2:1 aspect ratios at three different scales. As with many object detection networks, we find that faster R-CNN produces duplicate object proposals. To reduce redundancy, we apply the non-maximum suppression (NMS) algorithm on the proposed outputs [33].

We train and test the Faster R-CNN model using the images from the GDXray Castings dataset. Images are scaled to 600 pixels on the shorter edge. Additionally, we randomly flip the images horizontally at training time. This data augmentation technique is applied to artificially increase the size of the training dataset. We do not apply any other form of preprocessing to the images at training or testing time.

B. R-FCN

In contrast to the Faster R-CNN architecture, the R-FCN region-based detector is fully convolutional with almost all computation shared on the entire image. In Faster R-CNN, region-specific RBD must be applied several hundred times per image, greatly reducing performance [18]. In R-FCN, the feature map is cropped at the last layer of features prior to prediction, instead of cropping it at the layer where region proposals are generated [18]. This approach of pushing cropping to the last layer minimizes the amount of per-region computation that must be done.

We now describe the configuration of the R-FCN object detector used in this work. Similar to the Faster R-CNN network, we train and test R-FCN with the original unprocessed images from the GDXray dataset. We use the same anchor scales and aspect ratios as for Faster R-CNN. Again, we generate 300 region proposals for each image. Images are scaled to 600 pixels on the shorter side. Random horizontally flipping of the images is applied as a data augmentation technique. Again, we use NMS to reduce redundancy in the predicted output.

C. Single Shot MultiBox Detector (SSD)

Recently, a new family of object detection networks were proposed, in which the region proposal step is completely eliminated [15, 33]. One of the popular architectures, SSD, discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales [15]. At prediction time, the network generates scores for the presence

Ak, Ronay; Ferguson, Max; Law, Kincho; Lee, Yung-Tsun. "Automatic Localization of Casting Defects with Convolutional Neural Networks." Paper presented at 2017 IEEE International Conference on Big Data (BigData 2017), Boston, MA, United States. December 11, 2017 -

of each object category in each default box and produces adjustments to the box to better match the object shape. The SSD architecture uses feature maps from multiple different layers of the feature extractor to handle objects of various sizes. SSD is conceptually simpler than other methods as it eliminates the proposal generation, and subsequent feature resampling stage. The simplified architecture of the SSD network makes it faster than Faster R-CNN and R-FCN at test time [15].

We now describe the configuration of the SSD object detector used in this work. Unlike the Faster R-CNN and R-FCN architectures, the input size of the SSD architectures is fixed. We choose to use an input size of 300×300 pixels, as is common in many SSD implementations [15, 29]. During the training process, we randomly crop the input images to $300 \times$ 300 pixels. If the cropped image does not contain any groundtruth-labelled defects, it is discarded. Cropped images are randomly flipped, both horizontally and vertically at training time. The amount of benefit obtained from horizontal and vertical flipping the training images often depends on the type of images in the dataset. For example, horizontally flipping an image of a boat will give an equally probable image, while vertically flipping the same image will yield an improbable image. Flipping an image from the GDXray Castings dataset either horizontally or vertically will vield an equally probable image. We chose to vertically flip input images when training the SSD model, as the SSD architecture is believed to be less robust to image transformation than methods such as Faster R-CNN [15]. However, future work could explore the benefits of each data augmentation technique in more detail.

When evaluating a test image with an SSD model it is common to scale the image to the model input size (in this case 300×300 pixels). The aspect ratio is often preserved by padding the image with black pixels. However, we found that down-scaling the GDXray images made the defects very difficult to detect. Therefore, at test time, we slide the SSD classifier over the original image with a stride of 150 pixels. We post-process the resulting predictions using non-maximum suppression (NMS) to remove duplicate boundary boxes.

IV. SLIDING CLASSIFIER

A simple approach to the object localization task is to slide a classifier over the image, generating a heat-map of potential object locations. The heat-map can then be post-processed, to generate object location predictions. Although this method has been largely superseded by modern computer vision techniques, it can provide useful insight into the object localization task.

We develop a convolutional defect classifier based on the Xnet architecture proposed for a similar defect classification task [35]. The Xnet classifier is modified to reduce overfitting, leading to the architecture shown in Table II. We refer to this modified architecture as XnetV2. We add L_2 regularization to the last four convolutional layers of the neural network. Conceptually, the lower convolutional layers use the output from the upper layers to classify each image; applying regularization to the lower layers can prevent the classifier overfitting. This is done by adding a complexity penalty, sum of

TABLE II. PROPOSED XNETV2 ARCHITECTUR	RE
---------------------------------------	----

ı	Layer type	Filter Size (w×h, n)	Stride	Output Size w × h × d
0	Input	-	-	$32 \times 32 \times 1$
1	Convolution	7 × 7, 32	1	32 × 32 × 32
2	Convolution	3 × 3, 32	2	$16 \times 16 \times 32$
3	ReLU	-	-	$16 \times 16 \times 32$
4	Convolution	3 × 3, 64	1	$16 \times 16 \times 64$
5	Convolution	3 × 3, 64	2	$8 \times 8 \times 64$
6	ReLU	-	-	$8 \times 8 \times 64$
7	Dropout	-	-	$8 \times 8 \times 64$
8	Convolution+ L_2	3 × 3, 128	1	$6 \times 6 \times 128$
9	Convolution+L ₂	3 × 3, 128	2	$2 \times 2 \times 128$
10	ReLU	-	-	$2 \times 2 \times 128$
11	Convolution+L ₂	2 × 2,64	1	$1 \times 1 \times 64$
12	Convolution+L ₂	1 × 1,2	1	$1 \times 1 \times 2$
13	Softmax	-	-	$1 \times 1 \times 2$

squares of weights, to the loss function. For detailed information regarding regularization and overfitting, we refer the reader to [36]. The max-pooling layers in Xnet are replaced with convolutional layers with stride 2. A dropout probability of 0.25 is used during training. The proposed network architecture is "fully convolutional" in that it only contains convolutional and ReLU layers [37]. This allows the network to take an input of arbitrary size and produce correspondingly-sized output with efficient inference and learning. However, we deliberately design the network so that a 32×32 pixel image is mapped to a $1 \times 1 \times 2$ sized tensor.

Our key insight is to train the classification model on a large dataset of defect patches. We obtain a public dataset containing 47.520 cropped X-ray images of size 32×32 pixels along with the corresponding labels [35]. The dataset contains 23,760 patches with label defect and 23,760 patches with label nodefect. Example patches from this dataset are shown in Fig. 1. The dataset was originally synthesized by cropping defect images from the GDXray Castings series. When generating the dataset, the original cropped images were rotated at 6 different angles $(0^\circ, 60^\circ, 120^\circ \dots 300^\circ)$ and flipped both horizontally and vertically. This data augmentation technique was applied to artificially increase the size of the dataset, and encourage rotation invariance in classifiers that are subsequently trained on the dataset.

Ak, Ronay; Ferguson, Max; Law, Kincho; Lee, Yung-Tsun. "Automatic Localization of Casting Defects with Convolutional Neural Networks." Paper presented at 2017 IEEE International Conference on Big Data (BigData 2017), Boston, MA, United States. December 11, 2017 -



Fig 1. Examples of patches containing defects (left) and no-defects (right).



Fig 2. Original X-ray image containing a single defect (left), and the corresponding defect heat-map produced by the sliding classifier (right).

The XnetV2 architecture returns a vector of scores $\hat{y} \in \mathbb{R}^2$. with each component corresponding to the likelihood of a particular class. For the casting defect detection task, we only consider two classes: defect and no-defect.

We train the XnetV2 architecture on the patch dataset using the binary cross-entropy loss function [29]:

$$L(w) = -\frac{1}{N} \sum_{n=1}^{N} y_n \log \hat{y}_n + (1 - y_n) \log(1 - \hat{y}_n) , \quad (2)$$

where N is the number of training examples and $y_n \in \mathbb{R}^2$ is a vector containing the true class labels.

Applying the trained classifier to the images in the GDXray dataset produces a defect heat-map, as shown in Fig. 2. The heat-map can be seen as an estimate of the probability that a defect exists at the corresponding spatial location in the original image. The selective search algorithm is used to generate a set of bounding box proposals from the heat-map [38]. Bounding boxes with an average heat-map value greater than 0.5 are labelled as defects.

V. FEATURE EXTRACTORS

It has been shown that the choice of feature extractor can strongly influence the output of the object detection network. For casting defect detection, the choice of feature extractor is likely very important:

- Casting defects such as air bubbles tend to be very small, and can easily be mistaken for random image noise.
- Casting can generate large oddly shaped voids as shown in Fig. 3. The feature extractor must operate over a large enough space to capture these formations.

Most state-of-the-art feature extractors are designed and trained to detect well-defined objects such as and cars. We suspect that the features used to detect these objects may differ from the features used to detect casting defects.

We choose to base our feature extractors on the VGG and ResNet architectures [21, 26]. We initialize the feature extractors with pretrained weights created using the ImageNet dataset. This initialization strategy represents a form of transfer learning, where knowledge from the ImageNet dataset is transferred to the casting defect detection task.

A. VGG-16

The VGG architecture has been widely used in computer vision over the last few years. It consists of stacked convolutional and max pooling layers. We choose to use the smaller, and hence faster, 16-layer architecture known as VGG-16 [32]. The VGG-16 architecture is detailed in Fig. A1 in the Appendix. The outputs of the upper convolutional layers are used as feature maps in the object detection networks:

- Faster R-CNN: We extract features from the "conv5" layer whose stride size is 16 pixels.
- SSD: Following [15], we extract feature maps from the "conv4 3", and "fc7" layers. The "fc6" and "fc7" layers from the VGG-16 architecture are converted to convolutional layers as described in [15]. Five additional convolutional layers with decaying spatial resolution are appended to the VGG network, as described in [30].

Presently, we have not obtained the results from the R-FCN network with VGG-16 feature extractor, so we do not discuss this model in further detail here.

B. ResNet-101

The ResNet architecture was designed to avoid many of the issues that plagued very deep neural networks. Most predominately, the use of residual connections helps to overcome the vanishing gradient problem [17]. We choose to use the relatively large ResNet-101 variant which has 101 trainable layers [17]. The ResNet-101 architecture is detailed in Table A1 in the Appendix. We freeze the batch normalization parameters to be those estimated during ImageNet pretraining. Again, the outputs of the upper convolutional layers are used as feature maps in the object detection networks:

- Faster R-CNN: We extract features from the last layer of the "conv4" block with stride size 16 pixels. Feature maps are cropped and resized to 14×14 then max-pooled to 7×7 .
- **R-FCN:** We extract features from "block3" layer whose ٠ stride size is 16 pixels.
- SSD: We use the feature map from the last layer of the "conv4" block, which has a stride size of 16 pixels. Five additional convolutional layers with decaying spatial resolution are appended, with depths 512, 512, 256, 256, 128, as described in [30].

Ak, Ronay; Ferguson, Max; Law, Kincho; Lee, Yung-Tsun. "Automatic Localization of Casting Defects with Convolutional Neural Networks." Paper presented at 2017 IEEE International Conference on Big Data (BigData 2017), Boston, MA, United States. December 11, 2017 -



Fig. 3. Examples of X-ray images in the GDXray Castings dataset. The red boxes show the ground-truth labels for casting defects

VI. IMPLEMENTATION DETAILS

Transfer learning is used to reduce the total training time and improve the accuracy of the trained models. In our early experiments, we found it was difficult to train the object localization networks solely on the GDXray dataset. The main issue was that the localization models tended to overfit the training dataset, without learning any meaningful features. As a remedy, we initialize the localization model feature extractors using weights from feature extractors that were trained on the ImageNet dataset. In this way, the top layers of the localization network are already initialized to extract image features before we start training on the GDXray dataset. We then pretrain each localization model on the Microsoft Common Objects in Context (COCO) dataset [39]. When pretraining the model, we adjust the learning rates according to the schedule outlined in [30]. Training on the relatively large COCO dataset ensures that each model is initialized to localize common objects before it is trained to localize defects. Finally, we fine-tune the localization models on the GDXray Castings dataset.

The GDXray Casting dataset contains 2727 X-ray images mainly from automotive parts, including aluminum wheels and knuckles. The casting defects in each image are labelled with tight fitting bounding-boxes. The size of the images in the dataset ranges from 256×256 pixels to 768×572 pixels. Fig. 3 shows a random collection of images from the GDXray Casting dataset. The images in the dataset were randomly split into a training and testing set using an 80/20 split. We publicly release the list of images in the training and test set to facilitate fair comparison in future work [40].

When training the models, we choose to exclude images without defects, as we found this makes the training process much more stable. With non-defective samples excluded, the training set contains a total of 2308 images with 2334 annotated defects.

Training is completed on the Google Cloud Machine Learning Engine using a cluster with 5 worker nodes and 3 parameter servers [41]. Asynchronous stochastic gradient descent with momentum is used to train each model on the distributed hardware. Each worker node performs graph calculations on a NVIDIA Tesla Kepler K80 GPU.

Training the sliding classification architecture was completed in less than an hour. Fine-tuning the weights for the larger object localization architectures took up to 6 hours per model. The training loss for the fine-tuning process is shown in Fig. 4.



Fig 4. Training loss (smoothed) for the object localization networks, during the fine-tuning process. Note that the loss function for the SSD network varies from that of the R-FCN and Faster R-CNN networks, so the relative magnitudes of loss is not relevant.



Fig. 5. Mean average precision (mAP) of each object detection network on the test set, given different sized training sets.

Ak, Ronay; Ferguson, Max; Law, Kincho; Lee, Yung-Tsun. "Automatic Localization of Casting Defects with Convolutional Neural Networks." Paper presented at 2017 IEEE International Conference on Big Data (BigData 2017), Boston, MA, United States. December 11, 2017 -December 14, 2017.

VII. EXPERIMENTS

We conduct a number of experiments with the object localization models. The accuracy of each model on the test set is compared using the mean of average precision (mAP) [42]. We use the intersection over union metric (IoU) to determine whether a bounding box prediction is to be considered correct. To be considered a correct detection, the area of overlap a_0 between the predicted bounding box B_p and ground truth bounding box B_{at} must exceed 0.5 according to the formula:

$$a_o = \frac{\operatorname{area}(B_p \cap B_{gt})}{\operatorname{area}(B_p \cup B_{gt})}, \qquad (3)$$

where $B_p \cap B_{gt}$ denotes the intersection of the predicted and ground truth bounding boxes and $B_p \cup B_{gt}$ denotes their union.

The mAP for each architecture is shown in Table III. The Faster R-CNN architecture achieves a very high mAP of 0.921 with the ResNet feature extractor. The mAP of the SSD architecture is considerably lower than the Faster R-CNN and R-FCN architectures, regardless of the feature extractor. This is unsurprising, as the SSD architecture was designed to prioritize evaluation speed over classification accuracy [15].

In general, the defect localization models using the ResNet feature extractor performed better than those using the VGG feature extractor. This is not surprising, as the ResNet architecture has been shown to outperform the VGG architecture on a number of different tasks [17, 29]. However, most of these tasks involve identifying common objects such as cars, planes and bicycles. Additionally, the ResNet-101 feature extractor has a much larger number of layers than the VGG-16 feature extractor, which might not be beneficial in the defect localization task. For these reasons, we were unsure whether the defect localization models with the ResNet feature extraction architecture would yield better results than the models with the VGG feature extraction architecture.

The evaluation speed of each architecture is also measured. The models are evaluated on a 2.2 GHz Intel Xeon E5 virtual machine with 8 virtual CPU cores, 32 GB RAM, and a single NVIDIA Tesla Kepler K80 GPU. The models are evaluated

with the GPU being enabled and disabled. Every image in the testing data set is processed individually (no batching). The average evaluation time per image in the testing set is reported. Both the image preprocessing and post-processing are included in the reported evaluation time.

Table III shows that the evaluation time varies significantly between the networks. The SSD VGG-16 network is the fastest with an evaluation time of 0.088 seconds/image using the CPU and 0.025 seconds/image when using the GPU. The Faster R-CNN ResNet-101 architecture is the slowest, requiring 0.512 seconds/image when using GPU. Interestingly, the evaluation time seems to be inversely correlated with the mAP, showing the speed/accuracy tradeoff.

The localization accuracy of the sliding window method is significantly lower than that of the other object detection architectures. At a high level, this demonstrates the significant amount of progress that has been made in object localization over the last few years. However, there are a number of practical reasons why the performance of the sliding window method is so low. Firstly, the receptive field of the sliding window method is 32×32 pixels; that is, the network must generate a class score based on a 32×32 pixel image tile. Therefore, it is unlikely that the sliding window method can classify defects that are larger than this receptive field. Secondly, we use a naïve thresholding method to convert the defect heat-map into bounding box predictions. We hypothesize that this method is suboptimal, especially when compared to the way that the state-of-the-art object detectors are trained.

As with many deep learning tasks, it takes a large amount of labelled data to train an accurate classifier. We train the best performing classifiers with different amounts of training data, and observe the performance of each classifier. Fig. 5 shows how the amount of training data affects the accuracy of the classifier. We notice that the accuracy improves significantly when the size of the training dataset is increased from ~1100 to 2308 images. Extrapolating from Fig. 5 suggests that a higher mAP could be achieved with a larger training dataset.

Method	Evaluation time / image using CPU [s]	Evaluation time / image using GPU [s]	mAP
Sliding window method	2.231	0.231	0.461
Faster R-CNN VGG-16	7.291	0.438	0.865
Faster R-CNN ResNet-101	9.319	0.512	0.921
R-FCN ResNet-101	3.721	0.375	0.875
SSD VGG-16	0.088	0.025	0.697
SSD ResNet-101	0.141	0.051	0.762

TABLE III. COMPARISON OF THE ACCURACY AND PERFORMANCE OF EACH MODEL ON THE CASTING DEFECT LOCALIZATION TASK

Ak, Ronay; Ferguson, Max; Law, Kincho; Lee, Yung-Tsun. "Automatic Localization of Casting Defects with Convolutional Neural Networks." Paper presented at 2017 IEEE International Conference on Big Data (BigData 2017), Boston, MA, United States. December 11, 2017 -

VIII. FUTURE WORK

The object detection models described in this work are accurate enough and can be evaluated fast enough to be useful in a real manufacturing setting. However, the training process for these models is complex and computationally expensive. Future work could focus on developing a standardized method of representing these models, making it easier to distribute the trained models. It would also be interesting to apply the object detection models from this work to other defect detection tasks, such as welding defect detection. Finally, the models developed in this work could be adapted to detect casting defects in threedimensional X-ray computed tomography images.

IX. CONCLUSION

In this work, we have studied how several state-of-the-art object detectors can be used to localize casting defects in the GDXrav casting dataset. By decoupling the feature extraction laver from the object detection architecture, we could evaluate each object detection architecture with different feature extractors. Using an adapted version of the Faster R-CNN architecture, we were able to achieve a mAP of 0.921 on the testing dataset. This represents an extremely strong result, especially given the inherit difficultly of defect localization. The results show that there is a tradeoff between localization accuracy and inference time, with the Faster R-CNN models taking much longer to generate predictions than the other models.

To the best of our knowledge, this work represents the first attempt at casting defect localization with the GDXray dataset. We hope that the results presented here will serve as a benchmark for future work in casting defect localization.

ACKNOWLEDGMENT

The authors acknowledge the support by the Smart Manufacturing Systems Design and Analysis Program at the National Institute of Standards and Technology (NIST), US Department of Commerce. This work was performed under the financial assistance award (NIST Cooperative Agreement 70NANB17H031) to Stanford University.

DISCLAIMER

Certain commercial systems are identified in this paper. Such identification does not imply recommendation or endorsement by NIST; nor does it imply that the products identified are necessarily the best available for the purpose. Further, any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NIST or any other supporting U.S. government or corporate organizations.

REFERENCES

- T. R. Rao, Metal casting: Principles and practice. New [1] Age International, 2007.
- R. Raikolhe and J. Khan. "Defects, causes and their [2] remedies in casting process: A review," International Journal of Research in Advent Technology, vol. 2, no. 3. pp. 375-383, 2014.
- [3] S. Ghorai, A. Mukherjee, M. Gangadaran, and P. K. Dutta, "Automatic defect detection on hot-rolled flat steel products," IEEE Transactions on Instrumentation and Measurement, vol. 62, no. 3, pp. 612-621, 2013.
- [4] I. Baillie, P. Griffith, X. Jian, and S. Dixon, "Implementing an ultrasonic inspection system to find surface and internal defects in hot, moving steel using EMATs," Insight-Non-Destructive Testing and Condition Monitoring, vol. 49, no. 2, pp. 87-92, 2007.
- M. Lovejoy, Magnetic particle inspection: a practical [5] guide. Springer Science & Business Media, 2012.
- J.-Y. Buffiere, S. Savelli, P.-H. Jouneau, E. Maire, and [6] R. Fougeres, "Experimental study of porosity and its relation to fatigue mechanisms of model Al-Si7-Mg0. 3 cast Al alloys," Materials Science and Engineering: A, vol. 316, no. 1, pp. 115-126, 2001.
- E. Masad, V. Jandhyala, N. Dasgupta, N. Somadevan, [7] and N. Shashidhar, "Characterization of air void distribution in asphalt mixes using X-ray computed tomography," Journal of materials in civil engineering, vol. 14, no. 2, pp. 122-129, 2002.
- [8] X. Li, S. K. Tso, X.-P. Guan, and O. Huang, "Improving automatic detection of defects in castings by applying wavelet technique." IEEE Transactions on Industrial Electronics, vol. 53, no. 6, pp. 1927-1934, 2006.
- [9] D. Mery, V. Riffo, U. Zscherpel, G. Mondragón, I. Lillo, I. Zuccar, H. Lobel, and M. Carrasco, "GDXray: The database of X-ray images for nondestructive testing," Journal of Nondestructive Evaluation, vol. 34, no. 4, p. 42, Nov. 2015.
- [10] C. H. Lampert, M. B. Blaschko, and T. Hofmann, "Beyond sliding windows: Object localization by efficient subwindow search," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8.
- [11] H. Rathod and P. Maniar, "Prediction and investigation of shrinkage porosity defect in sand casting process - A review," 2016.
- [12] A. Sata and B. Ravi, "Bayesian inference-based investment-casting defect analysis system for industrial application," The International Journal of Advanced Manufacturing Technology, pp. 1–15, 2017.
- [13] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in Advances in neural information processing systems, 2015, pp. 91-99.

Ak, Ronay; Ferguson, Max; Law, Kincho; Lee, Yung-Tsun. "Automatic Localization of Casting Defects with Convolutional Neural Networks." Paper presented at 2017 IEEE International Conference on Big Data (BigData 2017), Boston, MA, United States. December 11, 2017 -

- [15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," European conference on computer vision, pp. 21--37, 2016.
- [16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431-3440.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.
- [18] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in Advances in neural information processing systems, 2016, pp. 379-387.
- [19] A. Gayer, A. Saya, and A. Shiloh, "Automatic recognition of welding defects in real-time radiography," NDT international, vol. 23, no. 3, pp. 131-136, 1990.
- [20] W. Daum, P. Rose, and H. Heidt, "Automatic recognition of weld defects in X-ray-inspection,' Materialpruefung, vol. 28, no. 6, pp. 177-180, 1986.
- [21] J. Munro, R. McNulty, W. Nuding, H. Busse, H. Wiacker, R. Link, K. Sauerwein, and R. Grimm, "Weld inspection by real-time radioscopy," NDT and E International, vol. 3, no. 29, p. 191, 1996.
- [22] D. Mery, T. Jaeger, and D. Filbert, "A review of methods for automated recognition of casting defects," INSIGHT-WIGSTON THEN NORTHAMPTON-, vol. 44, no. 7, pp. 428-436, 2002.
- [23] D. S. MacKenzie and G. E. Totten, Analytical characterization of aluminum, steel, and superalloys. CRC press, 2005.
- [24] Y. Tang, X. Zhang, X. Li, and X. Guan, "Application of a new image segmentation method to detection of defects in castings," The International Journal of Advanced Manufacturing Technology, vol. 43, no. 5-6, pp. 431-439, 2009.
- [25] T. Wiatowski and H. Bölcskei, "A mathematical theory of deep convolutional neural networks for feature extraction," arXiv preprint arXiv:1512.06293, 2015.
- [26] A. Kehoe and G. Parker, "An intelligent knowledge based approach for the automated radiographic inspection of castings," NDT & E International, vol. 25, no. 1, pp. 23-36, 1992.
- [27] V. Lashkia, "Defect detection in X-ray images using fuzzy reasoning," Image and vision computing, vol. 19, no. 5, pp. 261-269, 2001.
- [28] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE

conference on computer vision and pattern recognition, 2014, pp. 580-587.

- [29] A. Buja, W. Stuetzle, and Y. Shen, "Loss functions for binary class probability estimation and classification: Structure and applications," Working draft, November, 2005.
- [30] J. Huang et al., "Speed/accuracy trade-offs for modern convolutional object detectors," arXiv preprint arXiv:1611.10012, 2016.
- [31] J. E. Gentle, Numerical linear algebra for applications in statistics. Springer Science & Business Media, 2012.
- [32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
- [33] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," in Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, 2006, vol. 3, pp. 850-855
- [34] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779-788.
- [35] D. Mery and C. Arteta, "Automatic defect recognition in X-ray testing using computer vision," in Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on, 2017, pp. 1026-1035.
- [36] N. Buduma and N. Locascio, Fundamentals of Deep Learning: Designing Next-generation Machine Intelligence Algorithms, O'Reilly Media, Inc., 2017.
- [37] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, 2011, pp. 315–323.
- [38] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," International journal of computer vision, vol. 104, no. 2, pp. 154-171, 2013.
- [39] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in European conference on computer vision, 2014, pp. 740-755.
- [40] M. Ferguson, "Casting defect localization source code," 12-Sep-2017. [Online]. Available: https://github.com/maxkferg/casting-defect-detection. [Accessed: 12-Sep-2017].
- Google Inc., "Google Cloud Machine Learning [41] Engine," 29-Sep-2016. [Online]. Available: https://cloud.google.com/ml-engine/. [Accessed: 19-Oct-2017].
- G. Salton and M. J. McGill, "Introduction to modern [42] information retrieval," 1986.

Ak, Ronay; Ferguson, Max; Law, Kincho; Lee, Yung-Tsun. "Automatic Localization of Casting Defects with Convolutional Neural Networks." Paper presented at 2017 IEEE International Conference on Big Data (BigData 2017), Boston, MA, United States. December 11, 2017 -

APPENDIX: FEATURE EXTRACTION NETWORK ARCHITECTURES



Fig. A1. The standard VGG-16 network architecture as proposed in [32]. Note that only layers "conv1" to "fc7" are used in the feature extractor.

Layer name	Filter Size (width × height, number filters)	Output Size (width × height × depth)
conv1	7 × 7, 64, stride 2	112 × 112 × 64
	3×3 , max pool, stride 2	
conv2_x	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	56 × 56 × 256
conv3_x	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	28 × 28 × 512
conv4_x	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	14 × 14 × 1024
conv5_x	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	7 × 7 × 2048
output	average pool, fully connected, softmax	$1 \times 1 \times 1000$

ResNet-101 Architecture as proposed in [17]. Only layers "conv1" to "conv4_X" are used in the feature extractor TABLE A1.

Ak, Ronay; Ferguson, Max; Law, Kincho; Lee, Yung-Tsun. "Automatic Localization of Casting Defects with Convolutional Neural Networks." Paper presented at 2017 IEEE International Conference on Big Data (BigData 2017), Boston, MA, United States. December 11, 2017 -December 14, 2017.



Eleventh U.S. National Conference on Earthquake Engineering Integrating Science, Engineering & Policy June 25-29, 2018 Los Angeles, California

INFLUENCE OF FIRE ON THE SEISMIC SHEAR CAPACITY OF COLD-FORMED STEEL SHEAR WALLS

B. Andres¹ and M. S. Hoehler²

ABSTRACT

Lightweight construction using cold-formed steel (CFS) study represents roughly 20 % of the multi-story nonresidential building market in the United States. During an earthquake, these buildings rely on the performance of their lateral force-resisting system (LFRS). While there is extensive information about the structural performance and fire resistance of cold-formed steel construction, there is scarce knowledge about the performance of CFS-LFRS under combined hazards, such as mechanical and fire loading. Initial tests on earthquake-damaged steel-sheathed cold-formed steel shear walls under fire load (Phase 1) showed a change in failure mode from local to global buckling and highlighted the significance of the response of the gypsum on the overall fire and load-bearing behavior. A second phase of the project extends the study to multiple levels of fire severity and two additional CFS-LFRSs: (1) Oriented Strand Board shear panels, and (2) steel-strap braced cold-formed steel walls. Walls are tested under sequential mechanical and thermal loading. The mechanical loading setup is similar to ASTM E2126, and the fire load is provided by a natural gas burner located in a compartment placed next to the wall. Tests are performed at ambient conditions and under three fire loads representing (1) the ASTM E119 standard fire, (2) a severe fire, and (3) a mild fire. Results from this study are intended to provide data for a range of system performance under realistic conditions and will inform fire compartmentation design when significant lateral deformation of the building is anticipated, post-fire assessment to repair or replace a structure, as well as first responder decisions to enter a building when earthquake aftershocks are likely.

Andres Valiente, Blanca; Hoehler, Matthew. "Influence of Fire on the Seismic Shear Capacity of Cold-Formed Steel Shear Walls." Paper presented at Eleventh U.S. National Conference on Earthquake Engineering, Los Angeles, CA, United States. June 25, 2018 - June 29,

¹Research Consultant, Danish Institute of Fire and Security Technology, Jernholmen 10, 2650 Hvidovre, Denmark (email: bav@dbi-net.dk)

²Research Structural Engineer, National Fire Research Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899

Andres B, Hoehler M, Influence of fire on the seismic shear capacity of cold-formed steel shear walls. Proceedings of the 11th National Conference in Earthquake Engineering, Earthquake Engineering Research Institute, Los Angeles, CA. 2018.

Influence of Fire on the Seismic Shear Capacity of Cold-Formed Steel Shear Walls

B. Andres¹ and M. S. Hoehler²

ABSTRACT

Lightweight construction using cold-formed steel (CFS) studs represents roughly 20 % of the multi-story nonresidential building market in the United States. During an earthquake, these buildings rely on the performance of their lateral force-resisting system (LFRS). While there is extensive information about the structural performance and fire resistance of cold-formed steel construction, there is scarce knowledge about the performance of CFS-LFRS under combined hazards, such as mechanical and fire loading. Initial tests on earthquake-damaged steel-sheathed cold-formed steel shear walls under fire load (Phase 1) showed a change in failure mode from local to global buckling and highlighted the significance of the response of the gypsum on the overall fire and load-bearing behavior. A second phase of the project extends the study to multiple levels of fire severity and two additional CFS-LFRSs: (1) Oriented Strand Board shear panels, and (2) steel-strap braced cold-formed steel walls. Walls are tested under sequential mechanical and thermal loading. The mechanical loading setup is similar to ASTM E2126, and the fire load is provided by a natural gas burner located in a compartment placed next to the wall. Tests are performed at ambient conditions and under three fire loads representing (1) the ASTM E119 standard fire, (2) a severe fire, and (3) a mild fire. Results from this study are intended to provide data for a range of system performance under realistic conditions and will inform fire compartmentation design when significant lateral deformation of the building is anticipated, postfire assessment to repair or replace a structure, as well as first responder decisions to enter a building when earthquake aftershocks are likely.

Introduction

In recent decades, the building industry has reduced construction costs by using building typologies that use less material and require less time to construct compared to traditional methods. This has led to the increased use of pre-fabricated and lightweight framing systems. Today, lightweight frame structures represent the vast majority of single and multi-family residential housing in the United States [1], as well as a large proportion of the non-residential building market. Lightweight framing systems are commonly used as the primary structural system in single-story buildings. However, it is increasingly common to find lightweight framing systems used as structural elements in multi-story buildings [2]. In multi-story non-residential structures, lightweight frame structures represent 40 % of the U.S. construction market with approximately 20 % cold-formed steel (CFS) and 20 % timber; the remainder of the market being hot-rolled steel and concrete [3]. While only about 5 % of the cold-formed steel material is

Andres Valiente, Blanca; Hoehler, Matthew. "Influence of Fire on the Seismic Shear Capacity of Cold-Formed Steel Shear Walls." Paper presented at Eleventh U.S. National Conference on Earthquake Engineering, Los Angeles, CA, United States. June 25, 2018 - June 29,

2018

¹Research Consultant, Danish Institute of Fire and Security Technology, Jernholmen 10, 2650 Hvidovre, Denmark (email: bav@dbi-net.dk)

²Research Structural Engineer, National Fire Research Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899

Andres B, Hoehler M, Influence of fire on the seismic shear capacity of cold-formed steel shear walls. Proceedings of the 11th National Conference in Earthquake Engineering, Earthquake Engineering Research Institute, Los Angeles, CA. 2018.

used to construct the building's lateral force-resisting system (LFRS), the building relies on these systems to resist wind or earthquake loads. Over the past 30 years, extensive research has examined the structural performance and fire resistance of cold-formed steel construction [4–6]. There is, however, little information about the performance of CFS-LFRS [7] for cold-formed steel framed structures under combined mechanical and fire loading in spite of their increasing prevalence in the building market.

In 2016, a limited number of tests were conducted at the National Institute of Standards and Technology (NIST) National Fire Research Laboratory to characterize the response of earthquake-damaged CFS sheet-steel shear walls to fire [8]. The Phase 1 work was performed in conjunction with the project *Earthquake and Post-Earthquake Fire Performance of Mid-Rise Light-Gauge Cold-Formed Steel Framed Buildings* conducted at the University of California San Diego [9]. The results showed a change in failure modes in the walls from local to global buckling of the sheet steel, and highlighted the importance of the response of the gypsum on both the fire and structural behavior. Phase 2 extends the results to two additional types of CFS shear walls and two additional levels of fire exposure. The aim is to understand the thermal response of these systems to fires and characterize changes in the load-deformation behavior after various degrees of fire exposure and mechanical loading.

Experimental Program

The test program is designed to provide information about:

- 1. Post-fire lateral capacity strength to withstand horizontal loads (wind, earthquake) after a fire;
- 2. *Post-earthquake fire behavior* ability of the building to limit the spread of fire in the case of fire following earthquake; and
- 3. *Post-earthquake, post-fire lateral capacity* residual lateral strength after a fire following earthquake event; e.g. an aftershock or another earthquake before repairs are made.

The study investigates the following CFS walls [2] designed according to AISI S400-15 [10]:

- Type I: Steel-sheathed shear panels (similar to Phase I)
- Type II: Oriented Strand Board (OSB) shear panels
- Type III: Steel strap braced walls

The fire exposures were selected with a focus on risk associated with the severity of the fires:

- 1. Ambient: Baseline load-deformation behavior and lateral capacity.
- 2. *Standard ASTM E119* [11]: Characterized by a monotonic increase of temperature with time (Figure 1). Tests performed under standard ASTM E119 will target a 1-hour fire resistance rating.
- 3. *Severe Fire Exposure:* Designed to represent a post-flashover compartment fire which will compromise the structural capacity. Test results will provide information about failure mechanisms in fire damaged structures.
- 4. *Mild Fire Exposure:* This exposure also represents a post-flashover compartment fire, but with a lower maximum temperature and a short duration of the fire. Test results will help assess the impact of less severe fire on the structural capacity.



The thermal load is achieved placing a compartment with a natural gas diffusion burner capable of generating fires up to 4 MW heat release rate next to the wall (Figure 2). The thermal exposure to the wall is monitored using: (1) two vertical thermocouple arrays placed in the compartment with nine bare-bead, K-type (Chromel-Alumel) thermocouples spaced 28 cm apart, (2) six plate thermocouples and, (3) two heat flux gauges. The cyclic mechanical loading follows ASTM E2126-11 Method C (CUREE Basic Loading Protocol). Horizontal and vertical displacements are measured on the corners of the shear walls, and the out-of-plane displacement is measured at the mid-upper point of the wall. Temperatures on the unexposed side and through the cross-section are recorded by K-type thermocouples. The test program is shown in Table 1. In ambient tests, walls are cycled mechanically until failure. In standard ASTM E119 and the severe and mild fires, the walls are mechanically cycles after the fire to assess for their residual strength. In the severe fires the walls are also mechanically loaded prior to fire exposure to predamage the walls and to assess the residual capacity in case of fire following an earthquake.

> m

Wall	Specimen	Loading			
Туре	Name	Cycling (before fire)	Fire	Cycling (after fire)	
	SB01	Cyclic until failure	-	-	
Steel sheathed	SB02	-	ASTM E119 (1-hour)	Cyclic until failure	
	SB03	-	Severe	Cyclic until failure	
	SB04	-	Mild	Cyclic until failure	
	OSB01	Cyclic until failure	-	-	
Oriented Strand Board (OSB)	OSB02	-	ASTM E119 (1-hour)	Cyclic until failure	
	OSB03/OSB04	-		Cyclic until failure	
	OSB05/OSB6/OSB7	3 Drift levels (0.5%, 1%, 1.5%)	Severe	Cyclic until failure	
	OSB08	-	Mild	Cyclic until failure	
	S01	Cyclic until failure	-	-	
Strap braced	S02	-	ASTM E119 (1-hour)	Cyclic until failure	
	S03/S04	-		Cyclic until failure	
	S05/S06/S07	3 Drift levels (0.5%, 1%, 1.5%)	Severe	Cyclic until failure	
	S 08	-	Mild	Cyclic until failure	

Andres Valiente, Blanca; Hoehler, Matthew. "Influence of Fire on the Seismic Shear Capacity of Cold-Formed Steel Shear Walls." Paper presented at Eleventh U.S. National Conference on Earthquake Engineering, Los Angeles, CA, United States. June 25, 2018 - June 29,



Figure 2. Mechanical and Fire Loading Test Setup [8].

Conclusions

The experimental program identifies specific potential failure modes and begins to quantify the loss of strength of cold-formed steel shear walls under a combination of earthquake load and fire exposure. Initial tests on earthquake-damaged steel-sheathed shear walls under fire load showed a shift in failure mode from local to global buckling of the sheet steel and highlighted the significance of the response of the gypsum on the overall fire and load-bearing behavior. The levels of drift prior to fire loading did not significantly reduce their post-fire mechanical resistance. The Phase 2 testing is ongoing. The data from this study can provide insight into seismic design and for fire compartmentation design when significant lateral deformation of the building is anticipated.

References

- 1. U.S. Department of Commerce. Characteristics of New Housing. 2015;
- 2. Madsen RL, Castle TA, Schafer BW. NEHRP Seismic Design Technical Brief No . 12 Seismic Design of Cold-Formed Steel Lateral Load-Resisting Systems A Guide for Practicing Engineers. (12). doi:10.6028/NIST.GCR.16-917-38.
- Steel Framing Industry Association (SFIA). Market Share and Comparative Report, 2016; 3.
- 4. Sultan MA. A Model for Predicting Heat Transfer through Noninsulated Unloaded Steel-Stud Gypsum Board Wall Assemblies Exposed to Fire. Fire Technology [Internet]. 1996;32(1):239 – 257. Available from: http://link.springer.com/article/10.1007/BF01040217
- 5. Keerthan P, Mahendran M. Numerical modelling of non-load-bearing light gauge cold-formed steel frame walls under fire conditions. Journal of Fire Sciences. 2012;30(5):375–403. doi:10.1177/0734904112440688.
- 6. Takeda H. A model to predict fire resistance of non-load bearing wood-stud walls. Fire and Materials. 2003;27(1):19-39. doi:10.1002/fam.816.
- 7. Madsen RL, Castle TA, Schafer BW. Seismic Design of Cold-Formed Steel Lateral Load-Resisting Systems: A Guide for Practicing Engineers. Gaithersburg, MD: 2016.
- 8. Hoehler MS, Smith CM. Influence of Fire on the Lateral Load Capacity of Steel-sheathed Cold-Formed Steel Shear Walls - Report of Test. 2016.
- 9. Wang X, Hutchinson TC, Hegemier G, Gunisetty S, Kamath P, Meacham B. Earthquake and fire performance of a mid-rise cold-formed steel framed building - test program and test results: rapid release (preliminary) report (SSRP-2016/07). 2016;(December).
- 10. AISI S400-15, North American Standard for Seismic Design of Cold-Formed Steel Structural Systems. 2015;
- 11. ASTM International. ASTM E119-16a Standard Test Methods for Fire Tests of Building Construction and Materials. 2016; doi:http:// dx.doi.org/10.1520/E0119.



Eleventh U.S. National Conference on Earthquake Engineering Integrating Science, Engineering & Policy June 25-29, 2018 Los Angeles, California

COST ESTIMATES FOR THE SEISMIC **RETROFIT OF FEDERALLY OWNED AND** LEASED BUILDINGS

J.F. Fung¹ and D.T. Butry² and S. Sattar³ and S.L. McCabe³

ABSTRACT

Presidential Executive Order 13717, Establishing a Federal Earthquake Risk Management Standard, encourages federal agencies to "enhance resilience...[to] future earthquakes" by evaluating and retrofitting existing federal buildings based on current existing building codes. We develop a methodology for predicting seismic retrofit costs based on observable building characteristics and the desired performance standard. Our approach is to train a series of regression models on a database collected for FEMA 156, Typical Costs for the Seismic Rehabilitation of Existing Buildings. The models vary in the level of data required; e.g., a decision-maker may not have information on building construction type. We use prediction error to quantify the effect of data availability on estimates.

Butry, David; Fung, Juan; McCabe, Steven; Sattar, Siamak. "Cost Estimates for the Seismic Retrofit of Federally Owned and Leased Buildings." Paper presented at Eleventh US National Conference on Earthquake Engineering, Los Angeles, CA, United States. June 25, 2018 - June 29,

¹Applied Economics Office, Engineering Laboratory, National Institute of Standards and Technology (email: juan.fung@nist.gov)

²Applied Economics Office, Engineering Laboratory, National Institute of Standards and Technology ³Earthquake Engineering Group, Engineering Laboratory, National Institute of Standards and Technology

Fung JF, Butry DT, Sattar S, McCabe SL. Cost Estimates for the Seismic Retrofit of Federally Owned and Leased Buildings. Proceedings of the 11th National Conference in Earthquake Engineering, Earthquake Engineering Research Institute, Los Angeles, CA. 2018.



Eleventh U.S. National Conference on Earthquake Engineering Integrating Science, Engineering & Policy June 25-29, 2018 Los Angeles, California

Cost Estimates for the Seismic Retrofit of Federally Owned and Leased Buildings

J.F. Fung¹ and D.T. Butry² and S. Sattar³ and S.L. McCabe³

ABSTRACT

Presidential Executive Order 13717, Establishing a Federal Earthquake Risk Management Standard, encourages federal agencies to "enhance resilience...[to] future earthquakes" by evaluating and retrofitting existing federal buildings based on current existing building codes. We develop a methodology for predicting seismic retrofit costs based on observable building characteristics and the desired performance standard. Our approach is to train a series of regression models on a database collected for FEMA 156, Typical Costs for the Seismic Rehabilitation of Existing Buildings. The models vary in the level of data required; e.g., a decision-maker may not have information on building construction type. We use prediction error to quantify the effect of data availability on estimates.

Introduction

Presidential Executive Order 13717 (EO 13717), Establishing a Federal Earthquake Risk Management Standard, encourages agencies to "enhance resilience...[to] future earthquakes by evaluating and retrofitting existing federal buildings based on current existing building codes. To ensure resilience, EO 13717 suggests that agencies go beyond the minimum life safety performance standard. However, while guidance on evaluation and retrofit practices is readily available (e.g., ASCE/SEI 41-13 [1] and FEMA 547 [2]), a standard approach to estimating seismic retrofit costs does not exist. Moreover, the absence of easily obtainable estimates can make seismic retrofits, especially beyond the minimum of life safety, a prohibitive option for decision makers.

We present a methodology for estimating seismic retrofit costs based on observable

¹Applied Economics Office, Engineering Laboratory, National Institute of Standards and Technology (email: juan.fung@nist.gov)

²Applied Economics Office, Engineering Laboratory, National Institute of Standards and Technology

³Earthquake Engineering Group, Engineering Laboratory, National Institute of Standards and Technology

Fung JF, Butry DT, Sattar S, McCabe SL. Cost Estimates for the Seismic Retrofit of Federally Owned and Leased Buildings. Proceedings of the 11th National Conference in Earthquake Engineering, Earthquake Engineering Research Institute, Los Angeles, CA. 2018.

building characteristics (such as building construction type, age, and area) and the desired performance standard. Our approach relies on using a database collected for FEMA 156 [3] and FEMA 157 [4], hereafter FEMA 156/157, to train linear regression models to predict retrofit costs. The data is freely available, making our approach to estimating retrofit costs a viable and easy to implement option for decision makers.

Moreover, we quantify the cost of data limitations in terms of prediction error. Decision makers may not have access to, or the ability to collect, the data required for prediction; e.g., determining building type for a large building inventory may be infeasible. Our approach provides decision makers with guidance in prioritizing data collection for making retrofit cost predictions, with a direct measure in dollars of the trade-off from using less data.

The paper is structured as follows. We first provide some background on FEMA 156/157 and briefly describe the methodology. We then describe the data, presenting select summary statistics. This is followed by the main results, including estimates of prediction error and an application that compares retrofit costs for different performance standards. Finally, we conclude with a discussion of limitations and future directions.

Methodology

Seismic retrofit costs vary with a building's unique characteristics, such as building construction type, age, and area, as well as the desired performance standard. The most obvious, and perhaps most reliable, approach to estimating retrofit costs is to hire a consulting engineering or construction firm with the appropriate expertise. However, this is both time consuming and expensive, especially for large building inventories.

An alternative to collecting such primary data is to estimate retrofit costs based on secondary data on retrofits for other buildings. Naturally, an estimate based on retrofits for other buildings will be less accurate than an estimate made specifically for the building of interest. Nevertheless, using secondary data is relatively inexpensive and thus offers both a feasible approach and a reasonable frame of reference for cost estimates.

In NIST TN 1973, A Methodology for Estimating Retrofit Costs [5], we develop a methodology based on secondary data. The data was originally collected for FEMA 156/157, Typical Costs of Seismic Rehabilitation of Existing Buildings [3, 4]. The FEMA 156/157 reports present both a unique data collection effort and a set of methods for estimating retrofit costs using this data.

According to FEMA 156/157, the data set represents actual or estimated ("by an experienced design professional") retrofit project costs and contains extensive information on building characteristics. Importantly, the data is "validated" through follow-ups with survey respondents, and assigned a quality control rating (or "quality factor") to ensure that each response is "objective and reliable." Cost estimates with low quality factors are removed from the data set, with the intent to improve the overall quality of the data. The final database contains 2088 cost estimates.

Moreover, FEMA 156/157 present three options for estimating retrofit costs, differing in their data requirements. All three options estimate the mean structural retrofit cost, conditional on some observable characteristics. Option 1 estimates the mean cost conditional on building type, while Option 2 estimates the mean cost conditional on building type, building seismicity, and desired performance standard. Option 3 estimates the mean cost conditional on additional building characteristics and is thus presented as a linear regression model. The coefficients from a regression fit to the database can be used to predict costs.⁴

Our approach differs in two critical dimensions. First, we do not use the entire database to fit the regression. We split the data into a training set of size *n* to estimate the coefficients, and a test set of size m to estimate the prediction error. This approach provides a plausible estimate of prediction error and allows decision makers to compare different models by their accuracy.

Second, our model specification differs from Option 3 in FEMA 156/157. We use building type fixed effects rather than fitting a regression separately for each building type. In NIST TN 1973 [5], we show that this approach can reduce prediction error (defined as Mean Square Error).⁵ We also include information on whether the building is deemed historic since retrofits are likely to be very different for historic and non-historic buildings. Most importantly, we model retrofit costs using a Generalized Linear Model (GLM).⁶ Let $Y \equiv \log(E[C])$, where E[C] is the expected cost per square foot.⁷ Our **main model** is

$$Y = \beta_0 + \beta_1 Area + \beta_2 Age + \beta_3 Stories + \beta_4 Occup + \beta_5 Hist + \gamma + \eta + \delta + \zeta$$
(1)

where Area is building area in square feet; Age is building age in years; Stories is the number of above and below ground stories; Occup is building occupancy during the retrofit; Hist is a categorical variable denoting the building's historic status; γ is the building type fixed effect; η is the seismicity fixed effect; δ is the performance standard fixed effect; and $\zeta \equiv \eta \times \delta$ is the interaction of seismicity and performance.

While the implications of Area, Age, Stories, and Hist are straightforward, the rest of the predictors merit further discussion. Occup denotes whether the building occupants remain in place during construction (IP), are temporarily relocated to another part of the building during construction (TR), or whether the building is vacant during construction (V). In terms of construction costs, IP is the most expensive option, because construction takes place around the occupants, while V is the least expensive. Building type denotes the building model type (e.g., Unreinforced Masonry, or URM). Seismicity, which refers to the seismic hazard at the building site, is discussed in the next section, Data.

⁴More precisely, Option 3 estimates a linear regression, separately for each building type, on a *synthetic* version of the original database. More details and a comparison with our approach, including the danger of overfitting on the synthetic data, are given in NIST TN 1973 [5].

⁵Technically, for consistency with FEMA 156/157 [3, 4], we group similar building model types together rather than using all building types directly. See NIST TN 1973 [5] for more details.

⁶In contrast, the approach illustrated in NIST TN 1973 [5] models the log of retrofit costs, log(C), a standard linear model as in FEMA 156/157 [3, 4].

⁷We use log to refer to the natural logarithm.

Butry, David; Fung, Juan; McCabe, Steven; Sattar, Siamak. "Cost Estimates for the Seismic Retrofit of Federally Owned and Leased Buildings." Paper presented at Eleventh US National Conference on Earthquake Engineering, Los Angeles, CA, United States. June 25, 2018 - June 29,

The performance standards, chosen by the decision maker, are defined as follows in FEMA 156 [3]:

- Life Safety (LS): Allows for unrepairable damage as long as life is not jeopardized and ٠ egress routes are not blocked.
- Damage Control (DC): Protects some feature or function of the building beyond lifesafety, such as protecting building contents or preventing the release of toxic material.
- Immediate Occupancy (IO): Allows only minimal post-earthquake damage and disruption, with some nonstructural repairs and cleanup done while the building remains occupied and safe.

These definitions suggest that Damage Control is equivalent to Life Safety plus nonstructural improvements and is more likely to correspond with current Life Safety standards (e.g., as defined in RP-8 [6]).

Note that Eq. 1 only specifies the first element of a GLM, the model for expected cost, E[C]. The model links the mean, E[C], to the predictors through the log function; that is, the mean is modeled as $E[C] = exp(X\beta)$, where X is the matrix of predictors. The second element of a GLM is the distribution of C. Since C should be nonnegative, and the distribution of C is unlikely to be symmetric, we assume that C follows a Gamma distribution. One advantage of this specification is that it does not assume constant variance.⁸

We use Root Mean Square Error (RMSE) as our measure of out-of-sample prediction error. Let $\hat{\beta}$ denote the coefficients obtained in the training step. The RMSE is estimated from the test set as follows:

$$RMSE \equiv \left(\frac{1}{m}\sum_{i=1}^{m} \left(exp(\hat{\beta}^{T}x_{i}) - C_{i}\right)^{2}\right)^{\frac{1}{2}}$$

$$\tag{2}$$

Note that we estimate RMSE on the original scale of the response variable, C. Thus, RMSE as defined in Eq. 2 is directly interpretable as prediction error in dollars per square foot.

Data

A version of the data collected for FEMA 156/157 is freely available online through FEMA's Seismic Rehabilitation Cost Estimator (SRCE), a web-based app for estimating retrofit costs that is no longer maintained.⁹ Similarities and differences between the data presented in FEMA 156/157 and the data available through SRCE are discussed in NIST TN 1973 [5].

The data set, hereafter the SRCE data, contains 1978 cost estimates for buildings in the United States and Canada. Although the data contains extensive information on building

⁸The variance of C_i for observation *i* is modeled as proportional to V(E[C_i]).

⁹See https://www.fema.gov/media-library/assets/documents/30220.

Butry, David; Fung, Juan; McCabe, Steven; Sattar, Siamak. "Cost Estimates for the Seismic Retrofit of Federally Owned and Leased Buildings." Paper presented at Eleventh US National Conference on Earthquake Engineering, Los Angeles, CA, United States. June 25, 2018 - June 29,

characteristics, it represents a snapshot into retrofit costs in the early 1990s and is thus outdated. Nevertheless, the absence of reliable and comprehensive data on seismic retrofits makes this database an attractive option.

Costs are normalized, to account for variation across time and location, to 1993 US dollars (USD) for construction in the state of California. The adjustment factors, derived from the Engineering News Record's Building Cost Index (BCI), are included in the SRCE data. In order to obtain approximate *current* costs, we obtained historical BCI values to construct an adjustment factor that normalizes costs to 2016 USD for national average construction costs.

The SRCE data also contains information on the seismic hazard the building faces: its seismicity. The measure used, loosely based on peak ground acceleration, is derived from an outdated hazard map. To better approximate seismic hazard for current building inventories, we assign an updated measure of seismicity, derived from the 2014 US Geological Survey (USGS) National Seismic Hazard Maps, to the buildings in the SRCE data.¹⁰

Table 1 below presents select summary statistics for the SRCE data used in this paper, including cost (in 2016 USD), area, age, and number of stories. It is worth noting that the summary statistics suggest the distribution of cost is not symmetric. Moreover, the buildings are relatively short, in terms of number of stories, and therefore may not be representative of buildings in the United States. Further summary statistics are presented in NIST TN 1973 [5].

	Cost per sq ft (sq m)	Area 1000 sq ft (sq m)	Age years	Stories
Minimum	0.50 (3.1)	0.2 (0.0)	0	1
Mean	49.58 (320.1)	64.8 (6.0)	44	3
Median	26.71 (172.4)	25 (2.3)	40	2
Maximum	1688.54 (18175.94)	1430.3 (132.9)	153	38
Standard deviation	78.97 (849.78)	109.1 (10.1)	22	3

Table 1. Summary statistics for SRCE data: cost per square foot and select building characteristics (N = 1716).

Note that our sample size is 1716 and not 1978, the size of the SRCE data. This is because we drop data for Canadian buildings, as well as other minor restrictions.¹¹ We use a

¹⁰Since the SRCE data only includes state and county location information, we use a population-weighted county average of peak ground acceleration with a 10 % probability of exceedance in 50 years. See NIST TN 1973 [5] for more details.

¹¹Given that our objective is to estimate retrofit costs for federal buildings, we restrict the SRCE data to include only those buildings located in the United States. See NIST TN 1973 [5] for a discussion of the other restrictions on the SRCE data, as well as an application that uses the Canadian building data as an additional test set.

Butry, David; Fung, Juan; McCabe, Steven; Sattar, Siamak. "Cost Estimates for the Seismic Retrofit of Federally Owned and Leased Buildings." Paper presented at Eleventh US National Conference on Earthquake Engineering, Los Angeles, CA, United States. June 25, 2018 - June 29,

75/25 % split on our sample of the SRCE data to obtain our training and test sets, resulting in a training set of size n = 1287 and a test set of size m = 429.

Results

We use the training set described in the last section to fit our main model, Eq. 1, as well as versions of Eq. 1 that do not use all the predictors. Table 2 presents actual cost and predicted cost for the test data, as well as prediction error, in 2016 USD. In addition to the main model, Eq. 1, we train versions of the main model that do not include information on: building age ("No age"); building type ("No bldg type"); number of stories ("No stories"); and combinations of these cases. For ease of comparison, the last column presents the percentage change in RMSE relative to the main model for the cases subject to data limitations.

Model	Predicted cost per sq ft (sq m)	RMSE	RMSE relative to main model
Actual cost	44.60 (480.07)	-	-
Main model	57.67 (620.81)	54.25 (583.91)	0 %
No age	59.40 (639.37)	56.55 (608.76)	4.26 %
No bldg type	58.15 (625.98)	55.24 (594.66)	1.84 %
No stories	58.05 (624.89)	54.12 (582.59)	-0.23 %
No age, stories	59.66 (642.16)	56.23 (605.25)	3.66 %
No age, stories, bldg type	60.01 (646.01)	55.99 (602.65)	3.21 %

Table 2. Predicted cost and RMSE in 2016 USD per sq ft (sq m) for Main model, Eq. 1, and the model subject to data limitations.

The RMSE may be interpreted as the variation of the true values around our predictions. Thus, for instance, under the main model, true cost per square foot is within one standard deviation of our predicted mean, 57.67 ± 54.25 , since 44.60 lies in the interval (3.42, 111.92). Note that this deviation is expressed in the same units, dollars per square foot. While the true cost is within one standard deviation of the prediction, the magnitude of the RMSE represents a large degree of uncertainty around the prediction. Thus, in dollar terms, our prediction interval suggests true retrofit costs could be 6 times smaller or twice as large as predicted cost.

Lower values of RMSE imply better prediction. Omitting data on number of stories does not appear to impact prediction error. In fact, RMSE *decreases*, though the change is negligible. In all other cases, prediction error increases. Surprisingly, building age appears to be more important as a predictor of cost than building type.¹² These results suggest that building age and stories are sufficiently correlated with building type to serve as proxies in cases where building

¹²In contrast, our results in NIST TN 1973 [5] suggest building type is the most important predictor.

Butry, David; Fung, Juan; McCabe, Steven; Sattar, Siamak. "Cost Estimates for the Seismic Retrofit of Federally Owned and Leased Buildings." Paper presented at Eleventh US National Conference on Earthquake Engineering, Los Angeles, CA, United States. June 25, 2018 - June 29,

type is not available.

The cost predictions in Table 2 are averaged across all building types and performance standards in the test set. To better illustrate our approach, we use the trained main model, Eq. 1, to make predictions for a particularly vulnerable building type, Unreinforced Masonry (URM), broken down by performance standard. This provides a simple way for a decision maker to compare how retrofit costs increase with stricter performance standards.

In Table 3, we present cost predictions for URM buildings by performance standard. Our predictions assume that the building is not historic and that occupancy during construction is TR. We use the test set average values for Area, Age, Stories, and seismicity.

Performance	Mean cost per sq ft (sq m)	Predicted cost per sq ft (sq m)	RMSE
LS	19.07 (205.25)	29.27 (315.09)	10.20 (109.83)
DC	25.55 (274.93)	37.68 (405.62)	12.14 (130.69)
IO	25.27 (2271.88)	46.95 (505.43)	21.70 (233.54)

Table 3. Predicted cost and RMSE in 2016 USD per sq ft (sq m) by performance standard for URM buildings; based on Main model, Eq. 1.

Predicted costs are larger than actual costs. Moreover, the difference between DC and IO (and between LS and IO) is predicted to be larger than in the test data. However, one interpretation is that predicted costs are *conservative*, erring on the side of caution.

Consistent with the results in Table 2, actual costs for each performance standard are within one standard deviation of our predictions. In the case of LS and DC, actual cost is exactly one standard deviation below our predicted mean. One critical difference between these predictions and the aggregate predictions presented in Table 2 is that prediction error is much smaller relative to the predicted value; in the case of LS, it is almost three times smaller. In contrast, prediction error in Table 2 is roughly the same magnitude as the predicted value.

While the predicted mean values are not directly in line with actual values, RMSE provides an adequate measure of predictive uncertainty. Moreover, this approach provides a sensible starting point for estimating retrofit costs, especially for large building inventories. This example also illustrates the importance of breaking predictions down by building type and performance standard.

Finally, it is worth noting that the larger prediction error for DC and IO is likely due to the relative amount of observations with performance standard LS in the SRCE data. In our training set, for instance, 610 observations are for performance standard LS, compared to 334 for DC and 201 for IO.

Conclusion

In a forthcoming report, we apply our methodology to obtain retrofit cost estimates for federal buildings. While the methodology is developed to obtain the federal estimates, it is applicable to any building inventory. The major caveat is that the data is rather outdated. Our cost adjustments provide a suitable approximation to current retrofit costs, but predictive accuracy could be improved with more recent data.

We focus on predicting construction costs. The FEMA data includes non-construction costs (e.g., permits, fees, and relocation costs), which can easily be incorporated in a model to predict total cost. More importantly, our focus on construction costs ignores indirect costs, such as loss of productivity during a retrofit. Construction projects may also impose externalities on neighbors, from noise and disruption to traffic. Nevertheless, obtaining reasonable construction cost estimates is an important first step. In future work, we attempt to quantify both direct and indirect costs of seismic retrofits.

References

- 1. American Society of Civil Engineers (ASCE). Seismic evaluation and retrofit of existing buildings. 2014; ASCE/SEI 41-13.
- 2. Federal Emergency Management Agency (FEMA). Techniques for the seismic evaluation of existing buildings. 2006: FEMA 547.
- Federal Emergency Management Agency (FEMA). Typical costs for seismic rehabilitation of existing 3. buildings, Volume 1: Summary. 1994; FEMA 156.
- Federal Emergency Management Agency (FEMA). Typical costs for seismic rehabilitation of existing 4. buildings, Volume 2: Supporting documentation. 1995; FEMA 157.
- 5. Fung JF, Butry DT, Sattar S, McCabe SL. A methodology for estimating seismic retrofit costs. NIST Technical Note 1973. 2017; NIST TN-1973. DOI: https://doi.org/10.6028/NIST.TN.1973
- Holmes W, Bonowitz D, Cagley J, Poland C. Standards of seismic safety for existing federally owned and 6. leased buildings – ICSSC Recommended Practice 8. 2011; NISTGCR-11-917-12 (RP-8).

Butry, David; Fung, Juan; McCabe, Steven; Sattar, Siamak. "Cost Estimates for the Seismic Retrofit of Federally Owned and Leased Buildings." Paper presented at Eleventh US National Conference on Earthquake Engineering, Los Angeles, CA, United States. June 25, 2018 - June 29,

Estimation of Online Tool Wear in Turning Processes Using Recurrence Quantification Analysis (RQA)

Srinivasan Radhakrishnan¹, Yung-Tsun Tina Lee² and Sagar Kamarthi^{1*}

Abstract-In this work we exploit the underlying dynamics of a turning process captured in force measurements for online flank wear estimation. We transform the sensor signals into feature vectors using recurrence quantification analysis and then estimate flank wear using a gradient boosted regression model. The data is collected by conducting two sets of turning experiments. The first set of data, which has 168 records, is used for training the machine learning model. The second set of data, which has 95 records, is used for testing the performance of the flank wear estimation method. The results indicate that the proposed method gives accurate flank wear estimates. The root mean square of the flank wear estimation for the test data is 9.87×10^{-5} mm.

Keywords—Tool wear, flank wear, turning process, recurrence quantification analysis, gradient boosting

I. INTRODUCTION

Flank wear is a friction related degradation that occurs on the surface of a cutting tool due to the relative motion between the tool and the workpiece. Flank wear increases with time and when the wear progresses beyond a threshold, the tool is either replaced or reworked to ensure optimum performance. Flank wear effects the finishing quality of the workpiece. In the current practice the machine operator replaces the cutting tool before it reaches the end of life or manually examines the tool condition offline by stoppage of the cutting process. In contrast, the sensor-assisted tool monitoring methods examine the tool quality without stopping the turning process. Flank wear estimation involves estimating the width of flank wear by indirect measurements from sensors to assess the state of the system. Online flank wear estimation is one of the main required tasks in automating turning processes. Adequate research efforts have been devoted to develop effective online tool monitoring systems. Most approaches can be classified under physics-based models (analytical models) or machine learning models. Analytical models focus on the underlying laws of physics such as diffusion, adhesion, and abrasion that govern the tool wear in a cutting process. Early, important advancement in analytical models are documented in publications by Koren and Lenz [1], Koren [2], and Usui et al.[3]. On the other hand machine learning methods use direct and indirect measurements of the process state to extract features and build models and then predict a response variable using the model. Historically, multilayer neural networks, adaptive resonance theory networks, and Kohonen's feature maps are most widely used machine learning models to address tool wear monitoring problems. The application of multilayer neural networks for tool wear monitoring was investigated by Rangwala and Dornfeld [4], Emel [5], Elanayar et al. [6], and Chryssolouris et al. [7]. Burke [8] studied the performance of adaptive resonance networks for identification of flank wear states (fresh or worn). Kamarthi et al. [9] investigated the application of Kohonen's feature maps for classification of flank wear levels in turning process. In recent years proliferation of data centric approaches, and advent of Internet of Things (IOT) and superior data processing capabilities have enabled wide scale implementation of machine learning models for tool wear monitoring. Appendix A describes the abbreviations used in this paper.

II. METHOD

An overview of the tool wear estimation method is shown in Figure 1. The force measurements are collected along three directions. Phase synchronization is used to ascertain if the force measurements have captured the dynamics of the process. Once confirmed, one of the force measurements is converted into a recurrence plot (RP). From the RP, features are extracted using recurrence quantification analysis (RQA). The features are used as an input to a gradient boosted regression (GBR) model to estimate tool wear. In the following sections we will discuss in detail the tool wear estimation methodology.

A. Sensor Data Selection

Most of the methods for tool wear estimation described in literature use cutting force, temperature, vibration, or acoustic emission (AE) measurements, or some combination of these measurements. Force and vibration signals can be measured in three machine-tool coordinate directions: cutting, feed, and radial. Since machining forces in all three coordinate directions are likely to be sensitive to tool wear, though to a varying degree, forces in all three directions are measured in this work. Figure 2 shows the three forces that are generated during a turning operation. Vibration signals in feed direction are more sensitive to flank wear than vibration signals in radial and cutting directions. Studies[10], [11] indicate that vibration in radial and cutting directions exhibit similar characteristics with progressing flank wear. In this work we use force measurements as the input data for flank wear estimation.

In order to determine if the force measurements along all the three machine-tool directions represent the state of the turning process, we use the phase synchronization method.

¹Mechanical and Industrial Engineering, Northeastern University, Boston, USA sagar@coe.neu.edu

²Systems Integration Division, National Institute of Standards and Technology, Gaithersburg, MD, USA



Fig. 1. Overview of the flank wear estimation method



Fig. 2. Force measurements for turning operation

B. Phase-Synchronization of Sensor Signals and Recurrence plot

A dynamical system like a turning process can be represented in the phase space. The points in the phase space represent the states of the system using Takens algorithm [12], the state of a system at time instance i can be specified by m embedded dimensions and τ time delays. The dynamics of a time series can be reconstructed in the phase space $\overrightarrow{x}(i)$:

$$\vec{x}(i) = (x_i, x_{i+2\tau}, \dots x_{i+(m-1)\tau})$$
 (1)

Studies indicate that often m = 1 and $\tau = 1$ are sufficient to represent the underlying dynamics of the system.

Once a pair of signals (time series) are converted into phase vectors, we use a phase synchronization method to study the underlying dynamics of sensor signals. A set of signals are considered phase synchronized when the difference among their respective phases is bounded [13], the probability P(s) that the system returns to the ϵ -neighborhood of a former point $\overrightarrow{x_i}$ on the trajectory after s time interval is computed as [13]:

$$P(s) = \frac{\sum_{i=1}^{N-s} \Theta(\epsilon - \|\overrightarrow{x}_i - \overrightarrow{x}_{i+s}\|)}{N-s}, \qquad (2)$$

where $s = 1, 2, ... K \leq N - 1$, N is is the number of measured points \vec{x}_i ; ϵ is a closeness threshold distance; Θ is the Heaviside function; ||.|| is a suitable norm in the phase space under consideration. In this work K = 100 since the time delay does not exceed 100 time steps.

The correlation between $P_1(s)$ and $P_2(s)$ for s = $1, 2, ..., K \leq N - 1$ reflects the phase synchronization between the two signals originating from either the same or different systems. It is possible to detect phase synchronization by using the coincidence of the positions of the maxima for the two signals. In this work, the correlation of probability of recurrence (CPR), which is a correlation coefficient, is used as a criterion to quantify the phase synchronization. The CPR is defined as follows [13]:

$$CPR = \frac{\sum_{s=1}^{K} (P_1(s) - \bar{P}_1)(P_2(s) - \bar{P}_2)}{\sqrt{\sum_{t=1}^{K} (P_1(t) - \bar{P}_1)^2 \sum_{t=1}^{K} (P_2(t) - \bar{P}_2)^2}} \quad (3)$$

where \bar{P}_1 and \bar{P}_2 are the mean of $P_1(s)$ and $P_2(s)$ respectively $(s = 1, 2, \dots, K)$. If the two signals are synchronized, the CPR is close to positive or negative one; and if they are not synchronized, the CPR is close to zero.

We compute CPR values between each pair of the three force signals. If the three force signals represent the same dynamical process, then the CPR values should be close to 1. In that case we can proceed to use any one of the force measurements for flank wear estimation. We use the time series of one of the force measurements and convert it into a RP to visualize recurrence in phase space. A recurrence plot represents all recurrences in the form of a binary matrix R, where $R_{i,j} = 1$ if the state x_j is in the neighborhood of x_i in phase space, otherwise $R_{i,j} = 0$. $R_{i,j}$ is computed as follows:

$$R_{i,j}(\epsilon) = \Theta(\epsilon - (\|\overrightarrow{x}_i - \overrightarrow{x}_j\|)), \quad i, j = 1, ..., n$$
(4)

where $\overrightarrow{x}_i, \ \overrightarrow{x}_j \in \Re^m$ are points in the phase space. Once the recurrence plot is obtained, relevant features are extracted using Recurrence Quantification Analysis (RQA).

C. Feature Extraction using RQA

Zbilut and Webber [14] introduced RQA for measuring quantitative information hidden in a RP. ROA performs nonlinear data analysis to quantify the number of occurrences and the duration of recurrence of a dynamical system represented by its state space vector. It is actually a measure of information complexity. The main advantage of ROA is that it can provide useful information even for nonlinear and multivariate data. Marwan et al. [15] introduced new measures of complexity to quantify RPs by looking at the small-scale structures such as dots and lines. RPs mostly contains single dots and lines which are vertical/horizontal

or parallel to the mean diagonal referred to as line of identity (LOI). Since a RP matrix is symmetric, horizontal, and vertical lines correspond to each other (the upper right triangle of RP is equal to the lower left triangle of RP); only vertical lines are considered for RQA. The lines capture a typical behavior of the phase space trajectory. While the diagonal lines represent some segments of the phase space trajectory which run parallel for a period of time, the vertical lines represent some segments which remain in the same phase space region for some time duration. The RQA measures are computed based on the recurrence point density, and diagonal and vertical line structures in a RP. The RQA measures were defined by Zbilut and Webber [14] and Marwan et al. [15] and are summarized here for reference:

1) Recurrence rate (RR): RR is the density of recurrence points in a RP. This coincides with the correlation sum. It is the estimator of the correlation integral, which is the mean probability that the states at two different times are close. It is computed as:

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}$$
(5)

where $R_{i,j}$ is an element of the RP and N is the number of points on the phase space trajectory.

2) Determinism (DET): DET is the fraction of recurrence points forming diagonal lines. Diagonal lines represent epochs of similar time evolution of states of the process. DET is computed as:

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{i,j}^{N} R_{i,j}}$$
(6)

where P(l) is the histogram of the length l of the diagonal lines, l_{min} is the minimum acceptable diagonal line length, and N is the number of points on the phase space trajectory.

3) Laminarity (LAM): LAM is the percentage of recurrence points forming vertical lines. Vertical lines indicate intermittency, which is the alternation of phases of apparently periodic and chaotic dynamics. Laminarity is computed as:

$$LAM = \frac{\sum_{v=v_{min}}^{N} vp(v)}{\sum_{v=1}^{N} vp(v)}$$
(7)

where p(v) is the histogram of the lengths v of the vertical lines, v_{min} is the minimum acceptable vertical line length, and N is the number of points on the phase space trajectory.

4) Mean diagonal line length (L): L is the mean prediction time or the inverse of the divergence of the system (K2-entropy). Mean of the diagonal line lengths is computed as follows:

$$L = \frac{\sum_{l=l_{min}}^{N} lp(l)}{\sum_{l=l_{min}}^{N} p(l)}$$
(8)

where p(l) is the histogram of the length l of the diagonal lines, l_{min} is the minimum acceptable diagonal line length, and N is the number of points on the phase space trajectory.

5) Trapping Time (TT): TT measures the mean time that the system is trapped in a particular state with very slow change. Trapping time is computed as follows:

$$TT = \frac{\sum_{v=v_{min}}^{N} vp(v)}{\sum_{v=v_{min}}^{N} p(v)}$$
(9)

6) Longest diagonal line (L_{max}) : L_{max} is the length of the longest diagonal line:

$$L_{max} = \max\{l_i | i = 1, 2, \dots, N_l\}$$
(10)

where N_l is number of diagonal lines in the recurrence plot.

7) Longest vertical line (V_{max}) : V_{max} is the length of the longest vertical line:

$$V_{max} = \max\{v_i | i = 1, 2, \dots, N_v\}$$
(11)

where v_l is number of vertical lines in the RP.

8) Entropy (ENTR): ENTR is the Shannon entropy of the probability distribution of the diagonal lengths. The entropy of the line distribution measures the complexity of the recurrence structure:

$$ENTR = -\sum_{l=l_{min}}^{N} p(l) \ln(p(l))$$
(12)

where p(l) is the histogram of the length l of the diagonal lines, l_{min} is the minimum acceptable diagonal line length, and N is the number of points on the phase space trajectory.

The aforementioned RQA measures are used as input features to build the machine learning model.

III. IMPLEMENTATION OF PROPOSED METHOD

The method shown in Figure 1 is applied to flank wear estimation in a turning process. An extensive set of turning process experiments were conducted to collect data. The details of the workpiece, cutting inserts, sensors, and instrumentation are discussed in the following subsections.

1) Workpiece and Cutting Tools: AISI 6150, a chromiumvanadium steel alloy, is chosen as the workpiece material for the experiments. The cutting stock is in the form of cylinders of 914 mm long and 177 mm diameter. The hardness of the workpiece at a radius of 82.55 mm is 425 Bhn and at the center of the cylinder is 360 Bhn. The workpiece is cut with uncoated carbide grade K68 (C2) inserts. K68 is a tough Tungsten-Cobalt (WC-Co) unalloyed grade carbide. The geometric specifications of the insert and the tool holder are SPG-422 and KSBR-164C respectively.

2) Sensors and Instrumentation: Experiments are conducted on a 20 HP LeBlond 1610 heavy duty lathe. A three-axis Kistler Z3392/b piezo-electric force dynamometer is used to measure machining forces in cutting, feed, and radial directions. This dynamometer is located underneath the tool post. Force signals are passed through an appropriate set of amplifiers and data acquisition systems to collect the digitized data.

Kamarthi, Sagar; Lee, Yung-Tsun; Radhakrishnan, Srinivasan. "Estimation of Online Tool Wear in Turning Processes Using Recurrence Quantification Analysis (RQA)." Paper presented at 2017 IEEE Big Data, Boston, MA, United States. December 11, 2017 - December 14, 2017.
3) Data Collection: Cutting is started with a fresh insert edge. Every 60 seconds force measurements are sampled online; then flank wear is measured off-line using a toolmaker's microscope. The data collection at 60 second intervals is carried out until the cutting edge develops an average flank width of at least 0.45 mm, at which point the cutting edge is considered worn out. This limit is chosen according to the criteria recommended by the ISO to define the effective tool life for carbide tools (ISO 3685:1993) [16]. The experiments have shown that an insert edge usually runs for about 10 to 14 minutes before it is considered worn out. The sampling rates for digitization of sensor signals are determined by observing the power spectra of those signals from the preliminary experiments. Force signals are digitized with a sampling frequency of 3 kHz. All the signals are sampled for a length of 4096 points, but only the first 1024 points are considered for computation and analysis purposes.

4) Experimental Design: In this work, two sets of experiments, referred to as Set 1 and Set 2 are conducted. In all the experiments the depth of cut is kept constant at 1.27 mm; only the cutting speed and feed are varied. Several preliminary experiments were conducted to decide on the range of operating conditions. It is observed that when the cutting speeds are between 22.86 and 53.34 smpm, there is no formation of build-up edge, no chatter, and no formation of crater wear. Beyond 53.34 smpm, cutting tools start to develop crater wear. When cutting speeds exceeded 83.82 smpm, the tool tip catastrophically begins to deform. Based on these observations, the cutting speeds between 30.48 smpm and 67.06 smpm are used for experimentation.

The Set 1 experiments are conducted according to a fullfactorial design with three levels of speed (30.48, 39.62, 48.768 smpm) and five levels of feed (0.16256, 0.22352, 0.28448, 0.34544, 0.39116 mmpr). Accordingly, each one of the 15 (= 3×5) tools is used at exactly one combination of speed and feed. The Set 2 experiments are conducted according to a full-factorial design with three levels of speed (30.48, 39.62, 48.768 smpm) and three levels of feed (0.2235, 0.2845, 0.3454 mmpr). So, each one of the 9 (= 3×3) tools is used at one combination of speed and feed. It should be noted that the operating conditions used in Set 2 experiments are a subset of the operating conditions used in Set 1 experiments. The operating conditions in both Set 1 and Set 2 experiments cause mostly flank wear on the cutting tools. Set 1 has 168 instances collected over 15 cutting tools and Set 2 has 95 instances collected over 9 cutting tools.

5) Machine Learning Model: We investigated Multilayer Neural Network and Gradient Boosted Regression (GBR) as a learning model for estimating flank wear. We found that GBR outperformed multilayer neural networks in key parameters concerned with computation time and estimation error. Hence we used GBR as the machine learning model. Set 1 (training set) is used for training a GBR model and Set 2 is used for validation (test set). The results of the implementation are discussed in the next section.

IV. RESULTS

Each of the 168 instances from Set 1 generated three force measurements. The force signals were transformed from time domain into phase domain using Equation (1) with m = 1and $\tau = 1$. Phase synchronization defined in Equation (2) was applied to the transformed phase space vectors and the CPR values were computed for pairs of force signal components. We found that all the values of CPR were close to 1 indicating that any of the three force signals can effectively represent the system state. We use force in main cutting direction in the present work. The cutting force signals in phase space are converted to recurrence plots using Equation (4). A set of features are computed from the RPs using Equations (5-12). We end up with 168 feature vectors that represent the input vectors for the machine learning model. The aforementioned procedure was also followed for the instances from Set 2 which yielded 95 feature vectors. The root mean squared error (RMSE) of Set 2 (test data) estimation is 0.0497 mm. Figure 3 shows the actual (target) and predicted values of flank wear for all nine experiments. Figure 4a shows the progression of actual and predicted flank wear values for the first test tool and Figure 4b shows the box plot of RMSE values for all nine test tools in Set 2.



Fig. 3. Actual vs. predicted values for all nine experiments in Set 2 (test data)



Fig. 4. (a) Progression of actual and predicted flank wear values for the first test tool in Set 2, (b) box plot of root mean squared error for all the nine test tools in Set 2

V. CONCLUSIONS

The cutting force measurements in a turning process contain information about gradually progressing flank wear. This information can be successfully extracted and used for online flank wear estimation. The proposed flank wear estimation method provides accurate flank wear estimates within the range of operating conditions that were used during the training data generation. The average RMSE of 9 test tools is 0.0497 mm, which is about 11% deviation from the average flank wear level (0.45 mm) for replacing a cutting tool. This low estimation error makes the proposed estimation method very attractive for real-world applications. In future work we consider the option of using vibration signals and combination of force and vibration signals to further improve the estimation accuracy. In addition we will investigate other machine learning algorithms and feature representation of the measured signals in order to compare the performance of RQA based features with other methods.

APPENDIX

A. AbbreviAbbreviations

- Bhn = Brinell hardness
- CPR = Correlation of probability of recurrence
- GBR = Gradient boosted regression
- mm = Millimeter
- mmpr = Millimeter per revolution
- RMSE = Root mean squared error
- RP = Recurrence plot
- RQA = Recurrence quantification analysis
- smpm = Surface meters per minute

ACKNOWLEDGMENT

Funding for this research was provided by the National Institute of Standards and Technology under sponsor award number 70NANB15H028.

DISCLAIMER

Certain commercial software or equipment is identified in this article in order to aid understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

REFERENCES

- Y. Koren and E. Lenz, "Mathematical model for the flank wear while turning steel with carbide tools," *CIRP Proceedings on Manufacturing Systems*, vol. 1, no. 2, pp. 127–139, 1972.
- [2] Y. Koren, "Flank wear model of cutting tools using control theory," *Journal of Engineering for Industry*, vol. 100, no. 1, pp. 103–109, 1978.
- [3] E. Usui, T. Shirakashi, and T. Kitagawa, "Analytical prediction of cutting tool wear," Wear, vol. 100, no. 1-3, pp. 129–151, 1984.
- [4] S. Rangwala and D. Dornfeld, "Sensor integration using neural networks for intelligent tool condition monitoring," *Journal of Engineering for Industry*, vol. 112, no. 3, pp. 219–228, 1990.
 [5] E. Emel, "Tool wear detection by neural network based acoustic
- [5] E. Emel, "Tool wear detection by neural network based acoustic emission sensing," ASME, Dynamic Systems and Control Division Publication, vol. 28, pp. 79–85, 1991.
- [6] S. Elanayar, Y. C. Shin, and S. Kumara, "Machining condition monitoring for automation using neural networks," in ASME Winter Annual Meeting, Dallas, TX, USA, 1990, pp. 85–100.
- [7] G. Chryssolouris, M. Domroese, and P. Beaulieu, "Sensor synthesis for control of manufacturing processes," *Journal of Engineering for Industry(Transactions of the ASME)(USA)*, vol. 114, no. 2, pp. 158– 174, 1992.

- [8] L. I. Burke, "An unsupervised neural network approach to tool wear identification," *IIE transactions*, vol. 25, no. 1, pp. 16–25, 1993.
- [9] S. Kamarthi, G. Sankar, P. Cohen, and S. Kumara, "On-line tool wear monitoring using a kohonens feature map," in *Proceeding of the First Artificial Neural Networks in Engineering Conference, St. Louis*, 1991, pp. 639–644.
- [10] S. Rao, "Tool wear monitoring through the dynamics of stable turning," J. Eng. Ind. (Trans. ASME), vol. 108, no. 3, pp. 183–190, 1986.
- [11] A. Sokolowski, J. Liu, and J. Kormol, "On the correlation between the vibration measurement and tool wear in turning. japan," vol. 2, pp. 1075–1081, 1992.
- [12] F. Takens, "Detecting strange attractors in turbulence. in dynamical systems and turbulence, vol 898 of springer lecture notes in mathematics, 366-381," 1981.
- [13] M. C. Romano, M. Thiel, J. Kurths, I. Z. Kiss, and J. Hudson, "Detection of synchronization for non-phase-coherent and non-stationary data," *EPL (Europhysics Letters)*, vol. 71, no. 3, p. 466, 2005.
- [14] J. P. Zbilut and C. L. Webber, "Embeddings and delays as derived from quantification of recurrence plots," *Physics letters A*, vol. 171, no. 3-4, pp. 199–203, 1992.
- [15] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics reports*, vol. 438, no. 5, pp. 237–329, 2007.
- [16] ISO3685:1993, "Tool-life testing with single-point turning tools."

A multiscale framework for the prediction of concrete self-desiccation

M. Pathirage

Department of Civil and Environmental Engineering, McCormick School of Engineering and Applied Science Northwestern University, Evanston, IL, USA.

D.P. Bentz

Engineering Laboratory National Institute of Standards and Technology, Gaithersburg, MD, USA.

G. Di Luzio Department of Civil and Environmental Engineering Politecnico di Milano, Milano, Italy.

E. Masoero

School of Civil Engineering and Geosciences Newcastle University, Newcastle upon Tyne, United Kingdom.

G. Cusatis

Department of Civil and Environmental Engineering, McCormick School of Engineering and Applied Science Northwestern University, Evanston, IL, USA.

November 14, 2017

ABSTRACT: Cement hydration in concrete and mortar has been studied thoroughly over the past 50 years. To fully understand hydration in concrete and predict the evolution of the hygral, thermal, and mechanical properties at the structural level, one needs to study not only the reaction kinetics but also the development of the microstructure. Many models have been developed for this purpose, some of them looking only at the microscale or at the macro-scale and others tackling the fundamental nature of the issue, which can be qualified as a multiscale problem. This paper proposes a novel approach that consists of combining a cement hydration model at the microstructural level, the CEMHYD3D model, with a macroscopic hygro-thermo-chemical model, the HTC model. The coupling is performed by post-processing the output of the CEMHYD3D model, in particular with reference to cement hydration degree, silica fume reaction degree, and amounts of evaporable water and chemically bound water in order to identify through a curve fitting routine the parameters of the HTC formulation. This approach allows the possibility of predicting concrete behavior at multiple scales based on the actual chemical and microstructural evolution, thus enhancing the capabilities of the so-called HTC-CEMHYD3D model. This paper focuses on 1) introducing the concepts behind the formulation of self-desiccation and 2) demonstrating the predictive capabilities of the coupled model using some available experimental data.

INTRODUCTION 1

The behavior of concrete in terms of its mechanical, hygral, and thermal properties has always been of a great interest to researchers and engineers, and many computational models have been proposed.

Models at the micro-scale level are numerous. The first model that described the hydration kinetics and development is the Jennings and Johnson microstruc-Bentz, Dale; Cusatis, Gianluca; Di Luzio, Giovanni; Masoero, Enrico; Pathirage, Madura. "A multiscale framework for the prediction of concrete self-desiccation." Paper presented at EURO-C 2018 Conference - Computational Modelling of Concrete and Concrete Structures, Bad Hofgastein, Austria. February 26, 2018 - March 1, 2018.

ture simulation model (Jennings & Johnson 1986). Cement particles are represented as spheres, and hydration is simulated through a concentric growth of Calcium-Silicate-Hydrate (CSH) layers surrounding unhydrated particles. The model is able to capture key hydration features, including particle size distribution and evolution of volume due the formation of reaction products and the consumption of reactants. However, several model limitations must be cited. For example,

the model is only capable of simulating hydration of tricalcium silicate (C3S), and does not take into account transport phenomena.

The HymoStruc model (Van Breugel 1995, Koenders & Van Breugel 1997, Ye, Van Breugel, & Fraaij 2003) appeared as a continuation of this effort. It uses the same approach as in Jennings and Johnson's model, and statistical analysis is introduced. However, the previous limitations on transport phenomena, as well as the actual chemistry in the pore solution are still constraints to an extensive use of this model.

The CEMHYD3D model (Bentz 1997, Bentz & Garboczi 1990, Garboczi & Bentz 1992, Bentz 2005) proposed a new approach that is based on cellular automata algorithm. CEMHYD3D stands for Cement Hydration in three dimensions and was developed at the National Institute of Standards and Technology (NIST). This open-source code is a digital imagebased model and uses fundamental knowledge about hydration reactions and products. The model starts with an initial 3D microstructure, reconstructed from information on the actual cement powder via a scanning electron microscope (SEM) image for the phase distribution and the measured particle size distribution. Hydrating cement paste is represented by a cube of voxels. Each voxel is associated with a phase and can move and transform into another phase based on an iterative process through cycles. A database of cement types is available such that one needs only to calibrate a single parameter controlling the relationship between cycle lengths and real time. This model was successfully used in the literature and captures all aspects of the microstructure development, from chemical reactions to transport phenomena. It was legitimately pointed out (Thomas et al. 2011) that the parameter mapping the cycle lengths to the real time does not have a physical meaning and that the particle size limit of 1 voxel (1 μ m³) in the CEMHYD3D model brings a limitation. The HydratiCA model (Bullard 2007a, Bullard 2007b) and μ ic model (Navi & Pignat 1996, Bishnoi & Scrivener 2009) were developed in response to the aforementioned constraints. However, it is discussed later in this paper that such limitations have little or no effect on the proposed multiscale approach developed here. At the macro-scale, several models can be found in the literature that simulate not only cement hydration but also hygro-thermo-mechanical properties of concrete. Among others, the models of Ulm & Coussy (1995) and De Schutter & Taerwe (1996) are able to describe strength increase due to cement hydration. The temperature effect on hydration is taken into account in the work of Cervera, Oliver, & Prato (1999). More complex models provide better predictive capabilities in terms of hydration process, flow phenomena, chemical shrinkage and self-desiccation (Gawin, Pesavento, & Schrefler 2006a, Gawin, Pesavento, & Schrefler 2006b, Pan et al. 2017, Rahimi-Aghdam, Bažant, & Qomi 2017). Di Luzio & Cusatis (2009)

proposed a hygro-thermo-chemical model, called the HTC model, that takes into account moisture transport governing the pore relative humidity and heat transfer. It also accounts for the heat released by hydration, silica fume reaction and cement hydration degrees, adsorption/desorption isotherms (by estimating the chemically bound water and the evaporable water), and concrete permeability. This model is considered in this paper.

It must be noted that in most of these models, the use of empirical or semi-empirical formulations at the macro-scale do not allow a fundamental description of physics and chemistry at a lower scale. Moreover, for almost all the models, lengthy and sometimes complex calibration processes are needed, which limit the use of such models to cases in which available and up-to-date experimental data exist.

The present study introduces a multiscale approach, coupling the CEMHYD3D model to the HTC model, that overcomes the aforementioned limitations.

THE HTC-CEMHYD3D MODEL 2

2.1 Overview of the CEMHYD3D model

The model considers the cement particles as digitized spherical elements. These spheres are created to replicate the measured particle size distribution and phase distribution obtained from SEM analysis on a generic volume of one million cubic micrometers. Each of the spheres are made of adjacent one micrometer cube voxels that represent either cement clinker phases, calcium sulfates, or supplementary cementitious materials. All the particles lie in water, also represented with voxels; the microstructure is then computed to form a representative volume element. The hydration process is computed through a cellular automata algorithm, where cement phase voxels dissolve, then diffuse where capillary water exists, and collide with other reactants to produce solid hydration products. This process occurs in a series of cycles that cumulatively represent the termination time of the simulation. The volume stoichiometry is conserved for the hydration reactions in the unit volume, and periodic boundaries are used to ensure an overall homogeneous distribution of reactants and hydration products. To control the kinetics of the chemical reactions, a real time is mapped to the number of hydration cycles through a parameter that usually needs to be calibrated from experimental data on hydration degree or non-evaporable water content over time. In order to make the model parameter-free, a commonly used conversion factor of 0.00035 hours/cycle² (Bentz 2005) is taken and fixed for all the simulations shown later in the study.

2.2 Self-desiccation formulation in the HTC model

The HTC model is a macro-scale model. It considers the moisture transport and heat transfer through coupled diffusion equations (Di Luzio & Cusatis 2009). The present work focuses only on self-desiccation, which involves the adsorption/desorption isotherms, the cement hydration degree α_c , evaporable water w_e , and non-evaporable water w_n . Silica fume reaction degree α_s is also taken into account. Since the gradient of temperature and humidity in space are minimal or zero regarding the size of the samples used to study self-desiccation, one can formulate the following moisture mass balance equation at any time t as

$$\frac{\partial w_e}{\partial h}\frac{\partial h}{\partial t} + \frac{\partial w_e}{\partial \alpha_c}\dot{\alpha}_c + \frac{\partial w_e}{\partial \alpha_s}\dot{\alpha}_s + \dot{w}_n = 0 \tag{1}$$

where h is the internal moisture or relative humidity. The cement hydration degree is formulated as

$$\dot{\alpha}_c = A_{c1} (A_{c2} + \alpha_c) \langle \alpha_c^{\infty} - \alpha_c \rangle e^{-\frac{\eta_c \alpha_c}{\alpha_c^{\infty}}} e^{-\frac{E_{ac}}{RT}}$$
(2)

$$\alpha_c^{\infty}(h) = \alpha_c^{\infty,sat} \gamma(h) \tag{3}$$

$$\gamma(h) = e^{h_c^0 \left(1 - \frac{1}{h}\right)} \tag{4}$$

where A_{c1} , $A_{c2} = 0.005$, η_c and h_c^0 are material parameters, α_c^{∞} is the asymptotic hydration degree, $\alpha_s^{\infty,sat}$ is the asymptotic hydration degree at complete saturation, E_{ac} is the hydration activation energy, R is the universal gas constant. One can notice that in equation 2, the term $\alpha_c^{\infty} - \alpha_c$ can be negative when h is small enough, which would lead to an unphysical reduction in hydration degree over time. To ensure a positive rate of cement hydration degree, the Macaulay brackets $\langle . \rangle = \max(0, .)$ are used.

The silica fume reaction degree is written in a similar functional form as

$$\dot{\alpha}_s = A_{s1} (A_{s2} + \alpha_s) \langle \alpha_s^\infty - \alpha_s \rangle e^{-\frac{\eta_s \alpha_s}{\alpha_s^\infty}} e^{-\frac{E_{as}}{RT}}$$
(5)

$$\alpha_s^{\infty}(h) = \alpha_s^{\infty,sat} \gamma(h) \tag{6}$$

$$\gamma(h) = e^{h_s^0 \left(1 - \frac{1}{h}\right)} \tag{7}$$

where A_{s1} , $A_{s2} = 0.05$, η_s and h_s^0 are material parameters, α_s^{∞} is the asymptotic hydration degree, $\alpha_s^{\infty,sat}$ is the asymptotic hydration degree at saturation, and E_{as} is the silica fume reaction activation energy. The notation $\langle . \rangle$ is once again used in equation 5 to avoid an unphysical decrease of silica fume reaction degree over time.

The cement hydration degree and the silica fume reaction degree are taken into account in a single formulation, namely the total reaction degree (Di Luzio & Cusatis 2013) as

$$\alpha = \frac{\alpha_c c \tilde{Q}_c^{\infty} + \alpha_s s \tilde{Q}_s^{\infty}}{c \tilde{Q}_c^{\infty} \alpha^{\infty} + s \tilde{Q}_c^{\infty} \alpha^{\infty}}$$
(8)

where \tilde{Q}_c^{∞} and \tilde{Q}_s^{∞} are the latent heat of hydration reaction and silica fume reaction, respectively, per unit of hydrated cementitious material, c and s are respectively the cement and silica fume contents. The chemically bound water is commonly formulated as $w_n =$ $\kappa_c \alpha_c c$, where κ_c is the mass ratio of non-evaporable water at full hydration and c is the cement content. Furthermore, one can define w_e , the evaporable water, which is the resultant of two contributions, i.e., the evaporable water present in the CSH gel w_e^{gel} and the capillary water w_e^{cap} : $w_e = w_e^{gel} + w_e^{cap}$. They are formulated as

$$w_e^{gel} = g_2 \alpha \left(c\alpha_c^{\infty} + s\alpha_s^{\infty} \frac{\tilde{Q}_s^{\infty}}{\tilde{Q}_c^{\infty}} \right) \left[1 - e^{-10\alpha_c^{\infty}(g_1 - \alpha)h} \right] (9)$$

where $g_1 = 1.5$ and g_2 are material parameters. The capillary water is computed through a mass balance

$$w_0 + \Delta w = w_e^{gel} + w_e^{cap} + w_n \tag{10}$$

where w_0 is the initial water content. Δw is defined as the mass of water to be supplied to compensate chemical shrinkage and maintain saturated conditions. It can be written as

$$\Delta w = X_c \left[\left(1 - \frac{\tilde{Q}_c^{\infty}}{\tilde{Q}_s^{\infty}} \right) c\alpha_c + s\alpha_s \right]$$
(11)

where X_c is a material parameter. The equation 1 constitutes the governing equation and is discretized and solved using an implicit Crank-Nicolson method.

2.3 Multiscale identification of the HTC model parameters

Relevant outputs of the CEMHYD3D model are post-processed, and the material parameters of the HTC model listed earlier are automatically fitted. The identification procedure is performed under saturated and sealed conditions.

At saturation, i.e., h = 1, the parameters A_{c1} , η_c , $\alpha_c^{\infty,sat}$ in Equations 2 and 3 and A_{s1} , η_s , $\alpha_s^{\infty,sat}$ in Equations 5 and 6 are identified from the curves for cement hydration and the silica fume reaction degrees versus time obtained by the CEMHYD3D model. The activation energies for silica fume and cement reactions are the ones used in the micro-scale model. \hat{Q}_c^{∞} and \hat{Q}_s^{∞} in Equation 8 are identified by computing the initial masses of the different clinker phases and silica fume. In addition, one can compute the non-evaporable water by assuming that, for each cement clinker phase, the mass of chemically bound water produced is proportional to the mass of reacted phase, through constant coefficients (Molina 1992). This allows the identification of κ_c . The evaporable water in the gel is computed from the CEMHYD3D model, along with the chemical shrinkage, allowing the identification of the parameters g_2 and X_c .

 $\overline{c\tilde{Q}_{c}^{\infty}\alpha_{c}^{\infty} + s\tilde{Q}_{s}^{\infty}\alpha_{c}^{\infty}}_{\text{Paper presented at EURO-C 2018 Conference - Computational Modelling of Concrete and Concrete Structures, Bad Hofgastein, Austria.}$ (δ) The capillary water content is then evaluated from water content is then evaluated from Paper presented at EURO-C 2018 Conference - Computational Modelling of Concrete and Concrete Structures, Bad Hofgastein, Austria. February 26, 2018 - March 1, 2018.

Table 1: Identified HTC model parameters from CEMHYD3D

HTC		w/c = 0.22	w/c = 0.36	w/c = 0.48
Parameters		s/c = 0.1	s/c = 0.1	s/c = 0.1
$\overline{A_{c1}}$	[1/h]	9.67×10^{6}	5.76×10^{6}	2.78×10^{6}
η_c	[-]	4.42	5.49	6.62
$\alpha_c^{\infty,sat}$	[-]	0.38	0.55	0.83
$\tilde{E_{ac}}$	[kJ/mole]	40.00	40.00	40.00
h_c^0	[-]	0.20	0.40	0.50
A_{s1}	[1/h]	1.17×10^{14}	9.97×10^{13}	7.71×10^{13}
η_s	[-]	0.84	2.08	1.86
$\alpha_s^{\infty,sat}$	[-]	0.49	0.76	0.84
$\tilde{E_{as}}$	[kJ/mole]	83.14	83.14	83.14
h_s^0	[-]	1.50	0.55	0.40
\tilde{Q}_c^∞	[kJ/kg]	506.74	508.04	506.52
\tilde{Q}_s^∞	[kJ/kg]	780.00	780.00	780.00
κ_c	[-]	0.27	0.26	0.25
g_2	[-]	0.32	0.29	0.24
X_c	[-]	0.22	0.21	0.23

Equation 10, knowing w_0 .

Last, the CEMHYD3D model evaluates the cement hydration and silica fume reaction degrees curves in time for the sealed condition. This enables the identification of the two remaining parameters h_c^0 and h_s^0 in Equations 4 and 7.

3 PREDICTION RESULTS

The HTC-CEMHYD3D model was used to predict experimental data on self-desiccation (Persson 1996, Persson 1997, Persson 1998). Three different concrete mixes were used, with a wide range of water to cement ratios, i.e., w/c = 0.22, w/c = 0.36, and w/c = 0.48 and a constant silica fume to cement mass ratio of s/c = 0.1. Concrete specimens were sealed and the relative humidity was measured at 28, 90, and 450 days after casting. For the mix design, a low alkali cement was used, which was replicated by using a similar cement type in the HTC-CEMHYD3D model. Table 1 shows the material parameters that were identified using the CEMHYD3D model and transferred to the HTC model.

Results are shown in Figure 1. The predictions are in agreement with the experimental data and able to capture the decrease in humidity at various concrete ages. At 28 days, the absolute differences between experiments and simulations are 1.77%, 5.92%, 3.05% respectively for the mixes with w/c = 0.22, w/c = 0.36, and w/c = 0.48. The predictions improve as time goes. Indeed, for the mix with w/c = 0.22, the absolute differences are 0.70% and 0.81% respectively at 90 and 450 days. Similarly for the mix with w/c = 0.36, one can report differences of respectively 2.61% and 0.34%, whereas for the mix with w/c = 0.48, one gets 0.40% and 1.17% respectively at 90 and 450 days.



Figure 1: Prediction of self-desiccation in three different concretes. Upper and lower data points represent respectively +/-5% in humidity (absolute). Experimental data is from Persson, 1996, 1997, 1998.

4 CONCLUSIONS

A multiscale parameter-free framework to simulate hydration in concrete has been presented, including the modeling of the cement microstructure (CEMHYD3D model) and mathematical formulations modeling the macro-scale hygro-thermochemical properties of concrete, in particular selfdesiccation. The model results were also compared to experimental data. This preliminary study showed the following:

- It is possible to simulate self-desiccation using model predictions from the micro-scale without additional parameters from the experiments.
- The model is capable of accurately predicting the drop in internal humidity in time at ages of 90 days and beyond.
- The model can take into account a wide range of water to cement ratios and the inclusion of silica fume.

ACKNOWLEDGMENT 5

The work of the first and last authors was sponsored by the U.S. Army Engineer Research and Development Center (ERDC) under Contract Number W912HZ-17-C-0027. Permission to publish was granted by the director of the ERDC Geotechnical and Structures Laboratory.

REFERENCES

- Bentz, D. (2005). Cemhyd3d: A three-dimensional cement hydration and microstructure development modeling package. version 3.0.
- Bentz, D. P. (1997). Three-dimensional computer simulation of portland cement hydration and microstructure development. Journal of the American Ceramic Society 80(1), 3-21.
- Bentz, D. P. & E. J. Garboczi (1990). Digitised simulation model for microstructural development. Ceram. Trans. 16, 211-226.
- Bishnoi, S. & K. L. Scrivener (2009). µic: A new platform for modelling the hydration of cements. Cement and Concrete Research 39(4), 266-274.
- Bullard, J. W. (2007a). Approximate rate constants for nonideal diffusion and their application in a stochastic model. The Journal of Physical Chemistry A 111(11), 2084–2092.
- Bullard, J. W. (2007b). A three-dimensional microstructural model of reactions and transport in aqueous mineral systems. Modelling and Simulation in Materials Science and Engineering 15(7), 711.
- Cervera, M., J. Oliver, & T. Prato (1999). Thermo-chemomechanical model for concrete. i: Hydration and aging. Journal of engineering mechanics 125(9), 1018-1027.
- De Schutter, G. & L. Taerwe (1996). Degree of hydration-based description of mechanical properties of early age concrete. Materials and Structures 29(6), 335.
- Di Luzio, G. & G. Cusatis (2009). Hygro-thermo-chemical modeling of high performance concrete. i: Theory. Cement and

- Di Luzio, G. & G. Cusatis (2013). Solidificationmicroprestress-microplane (smm) theory for concrete at early age: Theory, validation and application. International Journal of Solids and Structures 50(6), 957-975.
- Garboczi, E. & D. Bentz (1992). Computer simulation of the diffusivity of cement-based materials. Journal of materials science 27(8), 2083-2092.
- Gawin, D., F. Pesavento, & B. A. Schrefler (2006a). Hygrothermo-chemo-mechanical modelling of concrete at early ages and beyond. part i: hydration and hygro-thermal phenomena. International Journal for Numerical Methods in Engineering 67(3), 299-331.
- Gawin, D., F. Pesavento, & B. A. Schrefler (2006b). Hygrothermo-chemo-mechanical modelling of concrete at early ages and beyond. part ii: shrinkage and creep of concrete. International Journal for Numerical Methods in Engineering 67(3), 332–363.
- Jennings, H. M. & S. K. Johnson (1986). Simulation of microstructure development during the hydration of a cement compound. Journal of the American Ceramic Society 69(11), 790-795.
- Koenders, E. & K. Van Breugel (1997). Numerical modelling of autogenous shrinkage of hardening cement paste. Cement and Concrete Research 27(10), 1489-1499.
- Molina, L. (1992). On predicting the influence of curing conditions on the degree of hydration. Cement och Betong Institututet.
- Navi, P. & C. Pignat (1996). Simulation of cement hydration and the connectivity of the capillary pore space. Advanced Cement Based Materials 4(2), 58–67.
- Pan, Y. et al. (2017). Lattice modeling of early-age behavior of structural concrete. Materials 10(3), 231.
- Persson, B. (1996). Hydration and strength of high performance concrete. Advanced Cement Based Materials 3(3), 107-123.
- Persson, B. (1997, Nov). Moisture in concrete subjected to different kinds of curing. Materials and Structures 30(9), 533-544
- Persson, B. (1998). Seven-year study on the effect of silica fume in concrete. Advanced cement based materials 7(3), 139-155.
- Rahimi-Aghdam, S., Z. P. Bažant, & M. A. Qomi (2017). Cement hydration from hours to centuries controlled by diffusion through barrier shells of csh. Journal of the Mechanics and Physics of Solids 99, 211-224.
- Thomas, J. J. et al. (2011). Modeling and simulation of cement hydration kinetics and microstructure development. Cement and Concrete Research 41(12), 1257-1278.
- Ulm, F.-J. & O. Coussy (1995). Modeling of thermochemomechanical couplings of concrete at early ages. Journal of engineering mechanics 121(7), 785–794.
- Van Breugel, K. (1995). Numerical simulation of hydration and microstructural development in hardening cement-based materials (i) theory. Cement and Concrete Research 25(2), 319-331.
- Ye, G., K. Van Breugel, & A. Fraaij (2003). Three-dimensional microstructure analysis of numerically simulated cementitious materials. Cement and Concrete Research 33(2), 215-222

Concrete composites 31(5), 301-308. Bentz, Dale; Cusatis, Gianluca; Di Luzio, Giovanni; Masoero, Enrico; Pathirage, Madura. "A multiscale framework for the prediction of concrete self-desiccation." Paper presented at EURO-C 2018 Conference - Computational Modelling of Concrete and Concrete Structures, Bad Hofgastein, Austria. February 26, 2018 - March 1, 2018.

Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature

Lisa Choe¹, Selvarajah Ramesh¹, Matthew Hoehler¹, John Gross¹

¹National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899; email: lisa.choe@nist.gov; selvarajah.ramesh@nist.gov; matthew.hoehler@nist.gov; john.gross@nist.gov

ABSTRACT

The National Institute of Standards and Technology recently expanded its large-scale structural-fire testing capabilities in the National Fire Research Laboratory. A landmark test series is being conducted on long-span steel-concrete composite floor beams designed and constructed following U.S. building codes and standards. The test program consists of five 13 m span composite beams with one of two types of shear connections: bolted/welded double-angle connections or single plate shear connections. Two of the five specimens have slab continuity at the beam ends. This paper presents the results from a recent test on the composite beam conducted at ambient temperature. The experimentally measured behavior and flexural moment capacity will be used as a baseline to compare with the composite beams still to be tested under combined mechanical and fire load. The datasets produced are providing technical information necessary to advance performance-based design of composite floor systems in steel-framed buildings subject to fire.

INTRODUCTION

The National Institute of Standards and Technology (NIST) recently commissioned the newly expanded National Fire Research Laboratory (NFRL) to enable large-scale structural-fire experiments under controlled laboratory conditions (Bundy et al., 2015). As for its first structural-fire test program, the performance of mechanically loaded composite floor beam assemblies exposed to a structurally significant fire has been studied. The resulting datasets will serve as a technical information to improve the current practice of structural fire engineering and design of steel-framed buildings.

Composite floor systems were selected for this study because of their widespread use in construction. The WTC 7 investigation report (NIST, 2008) and a recent study by Flint et al. (2013) identified potential vulnerabilities of composite floor systems in uncontrolled fires that were related to structural layout, sources of thermal restraint, and connections. However, modeling long-span composite beams exposed to a fire is challenging, especially, when they were restrained by structural steel connections and surrounding framing. The system-level fire performance is attributed to not only the high-temperature behavior of individual elements such as a steel beam, a concrete slab, shear studs, and shear connections at the beam ends, but also the combined structural-fire interactions among them.

Although there have been some experimental studies on composite beams with fire load, most of the specimens tested in U.S. [e.g., Wellman et al. (2011); Selden et al. (2015); Alfawakhri

Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

(2016)] were short span beams in the range of 4 m to 6 m. A series of stakeholder meetings led by NIST with researchers and industry experts identified that there is a significant need for data on composite floors beams with the beam length suitable for a modern office building, i.e., 10 m or longer. Hence, the focus of this test program is to study the fire performance of nominally 13 m long composite beams, which are restrained by shear connections and surrounding gravity frames, under real fire conditions.

A total of five composite test beams were designed and constructed in accordance with U.S. building codes and standards. Two different types of connections are used: (i) welded/bolted double angles and (ii) single plate connection (shear tab). Two of the five specimens have a slab continuity at the primary girders provided by rotationally restraining steel reinforcement in concrete slab extended beyond the support columns. Figure 1 shows a rendering of the test setup. For the specimens tested with fire, a compartment wall will be constructed around the specimens to confine the fire from the burners below the specimen. For the fire tests, three 1 m \times 1.5 m natural gas burners will be used to generate a fire with a total heat release rate of 4 MW. Details about specimens and test setup are presented in the companion paper titled "Design and Construction of Long-Span Composite Beam Specimens for Large Structural-Fire Tests" included in this conference proceeding.

This paper presents the results from the composite beam test at ambient temperature. The measured behavior, failure mode, and flexural moment capacity will serve as a baseline to compare with the results from subsequent fire tests on the composite beams. The datasets produced by this research is providing technical information to better understand the fire performance of long-span composite floor beams and to advance performance-based design of composite floor systems in steel framed buildings subject to structurally-significant fire.



Figure 1 Rendering of the test setup.

Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

TEST SETUP

Composite Beam and Shear connection

Design parameters of the composite beam specimen (e.g., design loads, beam spacing, concrete slab, reinforcement, and metal decking) were determined through surveys of U.S. structural engineering practitioners. Using values from the survey, the test beam was designed to represent a typical filler beam used in composite floor systems, in accordance with the current American National Standards Institute (ANSI) and American Institute of Steel Construction (AISC) ANSI/AISC 360 specification. The appropriate shear connection for gravity frames was selected using the AISC design tables. The complete design calculations were independently checked by AISC engineers to ensure the test beam was designed according to common U.S. practice. It should be noted that numerical values reported in this manuscript have two significant figures.

Figure 2a shows the cross-section of the composite beam specimen. A lightweight aggregate concrete with polypropylene fibers (PP) was cast over 7.6 cm deep fluted galvanized steel decking. The minimum specified compressive strength of the concrete was 28 MPa. The topping concrete measured 13 m by 1.8 m in plan. The steel deck units were installed with the flutes running perpendicular to the W18×35 beam made of American Society for Testing and Materials (ASTM) grade A992 steel. The thickness of concrete topping was 8.3 cm. For composite action, 19 mm diameter steel headed studs were welded at every flute spaced at 31 cm. The resultant percentage of composite action between the concrete slab and the steel beam, i.e., the total shear connection strength divided by the yield strength of the steel cross section, was 82 % with strong shear stud position. The slab was reinforced with a 6×6 W1.4×W1.4 welded wire fabric (9 mm² steel wires spaced at 15 cm in both the longitudinal and transverse directions), which was located at the mid-depth of a topping slab.

The test beam was supported by $W12 \times 106$ columns via two $L5 \times 3 \times 3/8$ angles made of ASTM A36 steel (Figure 2b). Each angle was cut to the length of 23 cm. The angle legs on the face of support columns were shop-welded to a 1.9 cm thick sacrificial connection plate; the other legs were bolted to the web of the W18x35 beam using three ASTM A325 bolts with the diameter of 1.9 cm. The spacing of bolt holes was 7.6 cm. The gap between the sacrificial plate and the end of the steel beam was 1.3 cm.

Table 1 summarizes the mechanical properties of the steel used in the construction of the beam. Strength of all steel components were measured by ASTM E8 standard (2016). The concrete for this specimen came from two batches. The mean concrete compressive strengths – determined according to ASTM C39 (2017) – were 50 ± 1.8 MPa (west half of beam) and 54 ± 3.3 MPa (east half of beam) at the time of the ambient beam tests (290 days after casting). It is noted that standard uncertainties are reported throughout this paper (cover factor = 1).

Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.



Figure 2 (a) Cross-section of the composite beam and (b) angle connection (connection dimensions in inches; 1 in. = 2.54 cm).

Matarial	ASTM Steel					
Material	Designation	E (GPa)	F_y (MPa)	F_u (MPa)	e_{u} (%)	
W18x35-Web	A992	200	370	470	19	
W18x35-Flange	A992	210	350	470	19	
Steel Angle	A36	200	380	510	17	
Structural Bolt	A325	210	900	960	6	
Shear Stud	A108	210	410	500	5	
Wire Mesh	A1064	200	730	760	1	

Table 1 Mechanical properties of steel components.

Note: E = Elastic modulus, F_y = Yield strength, F_u = Ultimate strength, e_u = Percent elongation at ultimate strength

Mechanical Loading System and Load Protocol

The test beam was subjected to a uniform floor load simulated by applying six equallyspaced point loads along the centerline of the beam (Figure 3a). The loading system consisted of three loading beams with rectangular tube section (9×7) which were hydraulically loaded at their ends using six actuators mounted below the strong floor of the test lab. The loading beams were guided such that they could only move vertically as the specimen deforms. Each loading beam was connected to the loading truss via clevis-pin assemblies. As the loading beam was loaded, each support of loading trusses transferred concentrated loads on the beam specimen. To prevent twisting of the test beam when loaded, the edges of the floor slab were laterally braced at the location of loading beams (Figure 3b). The 5×5 steel tubes were also used to provide vertical support at the east and west ends of slabs.

Figure 3c shows the applied load versus time relationship. All six actuators connected at the ends of three loading beams were programed such that they increased loads at a rate of 1.1 ± 0.053 kN/min. A total of three cycles of loading were applied to capture the full sequence of failure when the composite test beam was subjected to monotonically increasing flexural loads. The

Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

plateaus at 11 kN, 22, kN and 44 kN during Cycle 1 were pauses to make strain measurements using fiber optics. Two actuators each connected to the same loading beam were also programmed to increase the actuator stroke (displacement) at the same rate as the test beam deformed, and therefore the loading beam was horizontally leveled throughout the test duration.



Figure 3 (a) Loading system (b) lateral brace, and (c) applied load time history.

Instrumentation and Measurement Uncertainty

Figure 4 shows the instrumentation layout to measure global and local behavior of the test beam at various locations. Linear strain gages were used to measure the section behavior of the test beam and angle connections. Linear displacement transducers were employed to measure the displacement of the concrete slab at the locations of point loads, the displacement of the bottom flange of the W18×35 steel beam at midspan, the slip of the concrete slab relative to the end of the steel beam, and the gap closure at the bottom coped web. Rotation transducers were installed to measure the end rotation of the test beam. Table 2 shows the estimated total expanded uncertainty in measurements with a coverage factor of 2 as defined in Taylor and Kuyatt (1994). Further details of individual sensors are presented in Choe et al. (2017).

Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.



(b) Location of slip, gap, and rotation sensors (c) Location of strain gauges at midspan

Figure 4 Location of sensors (a) displacement sensors (b) slip, gap, and rotation sensors, and (c) strain gages at midspan.

Table 2 Measurement uncertainty of sensors				
Sensor description	Range (max)	Total Expanded Uncertainty ¹		
Actuator Load cell	365 kN	±10 %		
String potentiometer	760 mm	±15 %		
Linear position sensor	38 mm	± 2.0 %		
Inclinometer	60°	$\pm 8 \%$		
Strain gauge	50,000 με	±5 %		

TEST RESULTS

Global Behavior

Figure 5a shows a photograph of the beam after the test. Figure 5b shows the average point load versus the displacement relation measured at the bottom flange of the steel beam at midspan. Figure 5c shows the change in the vertical displacement profile of the concrete slab during the first cycle of loading, and Figure 5d shows the maximum vertical displacement measured at the peak load of each cycle. It should be noted that the displacements in Figures 5c and 5d were measured from the top of the concrete slab.

Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

¹ The uncertainty components included calibration, installation, random, and repeatability. The combined standard uncertainty was estimated by combining the individual uncertainties using root-sum-of-squares. The expended uncertainty was then computed by multiplying the combined uncertainty by a coverage factor of 2 corresponding to an approximately 95 % confidence interval.

As shown in Figure 5b, the flexural capacity of the test beam decreased with increasing number of loading cycles. During the first cycle, the vertical displacement of the beam was initially symmetric with respect to the midspan until the horizontal shear bond failure provided by headed studs was initiated at the west end (Figure 5c). After the first peak load of 63 ± 6.1 kN was achieved, the loads were manually relaxed to 40 kN, approximately. The second cycle was proceeded with the same load rate, and the specimen reached its second peak load of 61 ± 5.9 kN at which almost the west half of the beam lost the composite action. The third load cycle was resumed with the same load rate until the specimen collapsed following the complete weld fracture failure at the west angle connection. The third peak load was measured at 57 ± 4.1 kN, and the corresponding displacement at midspan was 360 mm, approximately L/35 where L is the beam span of 13 m. The values after \pm symbol indicate the standard deviation of six point loads.

The horizontal shear failure (fracture of a shear stud) initiated at the west end also resulted the asymmetrical vertical displacement about the centerline of the test beam. At the end of the third cycle of loading, as shown in Figure 5d, a plastic hinging was formed at the section where the west support of the middle load truss was located (i.e., approximately 1 m from the beam centerline). The maximum vertical displacement of concrete slab at this location was 360 mm. With the lateral braces provided on the concrete slab at the location of three loading beams and the vertical supports at the east and west slab ends, the maximum lateral displacement of the beam specimen was 2.5 mm approximately.



Figure 5 (a) Photograph of the beam after test, (b) total load versus displacement at midspan, (c) vertical displacement profile of the concrete slab, and (d) the maximum vertical displacement measured at the peak load of each cycle.

Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

End Slip and Gap Closure

Figure 6 shows (a) the slip of the concrete slab relative to the end of the steel beam and (b) the change in gap distance between the column sacrificial plate and the bottom coped web as the test beam underwent multiple cycles of flexural loading. Those relative displacements (between two planes) were measured using spring-loaded displacement sensors at the locations shown in Figure 4b. At the west end, the end slip significantly increased (to nearly 30 mm) at the peak load of the second cycle (Figure 6a), which indicated the horizontal interaction failure between the concrete slab and the steel beam. However, the end slip at the east end remained below 0.85 mm until the test beam failed.

The initial gap distance from the bottom coped web to the face of the column sacrificial plate was designed to be about 13 mm in accordance with U.S. practice (Figure 2b). As shown in Figure 6b, the gap distance varied as the test beam was subjected to increasing point loads in each cycle of loading. The gap closure at the west end was significantly greater than the east end. When the west half of the beam lost the composite action, the non-composite steel beam section was subjected to large bending. Yet, the gap distance at both ends never decreased to zero (i.e., there was no bearing of the bottom copped web on the support column).



Figure 6 (a) Slip between concrete slab and steel beam and (b) gap distance between sacrificial plate and beam web.

Section Behavior

The section behavior induced by flexural loading was measured using strain gauges installed at various cross-sections along the beam span. This paper only focuses on the strain responses in the expected plastic hinge region subjected to maximum bending moments (i.e. between two point loads around the midspan). Figures 7a and 7b show the strain in the mid-section measured at various load levels during the first cycle and the close view of strains in compression zone (the top flange of the steel beam and concrete slab), respectively. The y-axis of each plot indicates the section depth measured from the bottom flange of the W18×35 steel beam. As shown, the steel beam was in tension and the concrete slab was in compression during the first cycle, as indicating that the neutral axis was located within the top flange. The maximum compressive strain

Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

in concrete and the maximum tensile strain was 0.07 % and 0.2 % at the peak load of the first cycle.

Figures 7c and 7d show the strain distribution in the mid-section at each peak load of loading cycles and strains in the upper portion of the cross section, respectively. The neutral axis shifted downward to the steel beam section as the horizontal shear failure propagated into the midspan. At the end of the third cycle of loading, the specimen lost the horizontal shear interaction (between the concrete slab and the steel beam) which can be indicated by the presence of two neutral axes. The tensile strain was developed in the concrete slab. The top flange of the steel beam exhibited strain of 0.2 %, and the bottom flange failed by yielding with the maximum strain of 0.5 %.



Figures 7 (a) Strain at midspan during first cycle, (b) strains in compression zone during first cycle, (c) strain distribution in the mid-section at each peak load of loading cycles and, (d) strains in in compression zone during different load cycles.

Connection Behavior

The performance of the steel angle connections was measured using strain gauges as shown in Figures 8a and 8b. Due to the page limit, this paper only presents the strains in the critical locations such as the top of the angle legs welded to the sacrificial plate on support columns (Figure 8c) and the top of the other legs of the same angle that was bolted to the beam web (Figure 8d). Since the connection behavior was complex, the load cycle number is shown in the graphs. As shown, the top of the angle leg on the column side exhibited compressive strains during the first two cycles of loading as indicating prying of the attached angles. The strain reversal occurred when fracture at the weld return was initiated at about 45 kN during the third cycle. The angles at the

Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

east end developed more ductile behavior than the west end where the shear stud failure occurred. The maximum compressive strain was 0.83 % prior to strain reversal. On the other hand, the top of the angles bolted to the beam web at both east and west ends exhibited similar behavior. There was no significant increase in strains (less than 0.05 %) during the first cycle, and then tensile stresses, indicated by a positive slope in strain data, were developed during the last two cycles until weld fracture occurred.



Figure 8 (a) Strain gage layout in the west connection, (b) strain gage layout in the east connection, (c) strains in the top the angle legs welded to the sacrificial plate, and (d) strains in the top of the angle legs bolted to the beam web.

Failure Mode

The sequence of failure in the test beam was identified through a series of load cycles in which the applied point loads were increased at a rate of 1.1 kN/min. The test specimen exhibited failure at the interface between the shear studs and the concrete when subjected to a monotonically increasing bending moment about its strong axis. This failure was initiated at the west end and then propagated toward the midspan (Figure 9a), presumably due to the slightly lower concrete strength at the west end of the specimen. As shown in Figure 9b, shear studs in this region deformed in the same direction of compressive forces by concrete slab subjected to flexural loading, and the last shear stud at the west end failed by steel rupture.

Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

Once the west half of the beam lost the continuity between the beam and slab, plastic hinging formed in the critical section at the west support of the middle loading truss (about 1 m from the beam centerline). This region was subjected to the maximum bending moment and the bearing stress from the point load. In this stage, both the beam and slab had their own neutral axes (each at their center of gravity). The beam failed by yielding and a flexural crack was developed through the depth of concrete slab (Figure 9c). The legs of west angles welded to the column sacrificial plate were completely detached at the peak load of the third cycle (Figure 9d), which was the same as the limit state of angle connections (i.e., weld shear) designed using AISC design table.



Figure 9 (a) failure of composite action at the west side of specimen, (b) exposed shear studs after the test, (c) exposed concrete at 1 m left of beam centerline, and (d) west connection after the test.

SUMMARY AND CONCLUSIONS

This paper presents the results of an ambient temperature composite beam test recently conducted at NIST. The specimen was nominally 13 m long and supported by steel columns via AISC bolted/welded double-angle connections. The data produced by this test will serve as the baseline for subsequent tests on composite beams with combined mechanical and fire load, as well as to validate computational models.

Failure initiated at the west end of the beam at the interface between the shear studs and the concrete when the specimen was subjected to a symmetric bending moment about its strong axis. Subsequently, flexural plastic hinging occurred near midspan of the beam (about 1 m to the west of the beam center). The test was terminated when fracture occurred in the welds on the double-angle connection at the west end of the beam. A peak load capacity of 63 ± 6.1 kN was achieved with a corresponding midspan deflection of approximately 100 mm. The gap between the beam web and the column flange varied during loading, but never fully closed. Processing of strain data measured in the slab and the beam is ongoing.

Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

ACKNOWLEDGEMENTS

The support of numerous NIST colleagues on this work is acknowledged and greatly appreciated: Brian Story, Anthony Chakalis, Laurean DeLauter, Philipp Deardorff, Matthew Bundy, Artur Chernovsky, Doris Rinehart, Chao Zhang, Mina Seif, Qi Li, Ana Sauca, Dale Bentz, Scott Jones, Timothy Barrett, Max Peltz, Kevin McGrattan, Morgan Bruns, Randy Shields, Fahim Sadek, Joe Main, Jonathan Weigand, and Jian Jiang. The authors would like to thank Alana Guzetta of US Concrete who prepared the concrete mix design, as well as Roberto Talavera of Superior Concrete who supervised the casting. Expertise on fiber optic instrumentation was provided by Genda Chen, Yi Bao, Steve Guo, Matthew Klegseth, Branko Glisic, and John Reilly.

REFERENCES

- Alfawakhri, F., Carter, C., Berhing, R., Zeeveld, P., Hervey, & F., Woods, L. (2016). "The Effects of Load Intensity and Restraint on the Fire Resitance of Steel and Composite Beams", proc of the 9th International Conf. Structures in Fire, Princeton, NJ.
- ASTM International. (2017) "Standard Test Method for Compressive Strength of Cylindrical Concrete Specimens", ASTM. C39/C39M - 17b, ASTM International, West Conshohocken, PA, 2017
- ASTM International. (2016) "Standard Test Methods for Tension Testing of Metallic Materials", ASTM.E8/E8M-16a, International. West Conshohocken. ASTM PA. 2016, doi:10.1520/E0008 E0008M-16A
- Bundy, M., Hamins, A., Gross, J., Grosshandler, W., and Choe, L. (2016). "Structural fire experimental capabilities at the NIST National Fire Research Laboratory," Fire Technology, 52(4), 959-966.
- Choe, L., Ramesh, S., Hoehler, M., Seif, M., Bundy, M., Zhang, C., and Gross, J. (2017). "National Fire Research Laboratory Commissioning Project: Testing Steel Beams under Localized Fire Exposure", NIST TN 1977, DOI: https://doi.org/10.6028/NIST.TN.1977
- Flint, G., Lamont, S., Lane, B., Sarrazin, H., Lim, L., Rini, D., and Roben, C. (2013). "Recent lessons learned in structural fire engineering for composite steel structures", Fire Technology, 49(3), 767-792.
- NIST (2008). Final Report on the Collapse of World Trade Center Building 7, Federal Building and Fire Safety Investigation of the World Trade Center Disaster (NIST NCSTAR 1A)
- Selden, K. L., Fischer, E. C., & Varma, A. H. (2015). "Experimental investigation of composite beams with shear connections subjected to fire loading". Journal of Structural Engineering, 142(2), 04015118.
- Taylor, B. N., and Kuyatt, C. E. (1994). "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results." NIST Technical Note TN-1297, Gaithersburg, MD.

Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

Wellman, E. I., Varma, A. H., Fike, R., and Kodur, V. (2011). "Experimental evaluation of thin composite floor assemblies under fire loading. Journal of Structural Engineering", 137(9), 1002-1016.

Choe, Lisa; Gross, John; Hoehler, Matthew; Ramesh, Selvarajah. "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

Design and Construction of Long-Span Composite Beam Specimens for Large Structural-Fire Tests

Selvarajah Ramesh, Lisa Choe¹, Matthew Hoehler, William Grosshandler, John Gross

National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899; email: selvarajah.ramesh@nist.gov; lisa.choe@nist.gov; matthew.hoehler@nist.gov; william.grosshandler@nist.gov; john.gross@nist.gov

Abstract

A series of experiments is being conducted on five 12.8 m long composite floor beams at the newly-commissioned National Fire Research Laboratory at the National Institute of Standards and Technology. The test beams were designed according to current U.S. building codes with input and design-review provided by panel of expert practitioners. Each specimen consists of a W18x35 beam and a 1.83 m wide lightweight concrete slab cast over profiled steel deck units. Composite action between the steel beam and concrete slab is achieved using 19 mm diameter steel shear connectors spaced at 305 mm on center. One of the five specimens was tested at ambient temperature to measure its behavior and flexural strength, and to verify the purpose-built mechanical loading system. The remaining four specimens will be tested under simultaneously applied mechanical loading and an enclosure fire using a natural gas fuel delivery system. The beam-to-column connections and rotational restraint of the slab ends over the support girders were varied among those four specimens.

Introduction

Long-span steel-concrete composite floor beams are widely used in steel-frame buildings because they are a cost-effective means of spanning large open spaces. Predicting the fire behavior of composite beams, however, is challenging due to the many influencing parameters, which include the gravity connections, force-slip behavior provided by shear connectors, metal decking profiles, concrete characteristics, and steel framing of adjacent bays. For instance, in multi-bay buildings, composite floor systems are continuous over primary girders. The steel reinforcements, such as reinforcing bars or welded wire fabric, are typically placed in concrete slabs at those locations to minimize surface cracking due to a negative bending moment from prescribed floor loads. Also, thermally-induced restraints in the floor assemblies, when exposed to fire, would vary depending upon the combination of stiffness of surrounding frames and beam-to-connection types. Most of the furnace fire tests conducted in the past, however, focused on the fire-resistant ratings of unrestrained composite floor assemblies with limited sizes, and generated very limited information to be used for evaluating the effects of thermally induced restraints. The most recent test data in U.S., e.g. Selden et al. (2015), Wellman et al. (2011), and Guo et al. (2010), are available for composite beam specimens with lengths ranging from 4 m to 6 m.

Recently, the National Institute of Standards and Technology (NIST) hosted three stakeholder workshops [Almand et al. (2004); Almand (2013); Yang et al. (2015)] to prioritize the needs of structural-fire experimental research that could be conducted in the newly expanded National Fire Research Laboratory (NFRL). Based on the stakeholder's recommendations made

¹ Corresponding Author

Choe, Lisa; Gross, John; Grosshandler, William; Hoehler, Matthew; Ramesh, Selvarajah. "Design and Construction of Long-Span Composite Beam Specimens for Large Structural-Fire Tests." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

from those workshops, composite floor systems were selected for the NFRL's first structural-fire test program because of their widespread use in building construction and because of modeling challenges in such systems under a building fire. The proposed NFRL test programs consist of two phases: long-span steel-concrete composite beam tests (Phase 1) and steel-concrete composite floor tests (Phase 2). Data and technical information developed by the NFRL test programs would be of "great practical import and a major impact on design methods" (Almand, 2013) for composite floor system under structurally significant fire conditions. Specifics of the test frame configuration, test fires, and details of the test matrices were proposed through a series of stakeholder meetings, hosted by NIST, with national and international researchers and industry experts in the field of structural fire engineering. This paper presents the experimental design and construction of the NFRL long-span composite beam test series. Testing is currently in progress.

A total of five 12.8 m long composite test beams are being tested under combined mechanically applied gravity loads and a structurally significant fire provided by the NFRL natural gas fuel delivery system. Specimens were designed to represent composite floor beams, commonly found in mid-rise office buildings, in accordance with U.S. design codes and standards. One specimen was recently tested at ambient temperature. The test results are presented in the authors' companion paper titled "Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature" included in this proceeding. Other four specimens were designed to achieve a 2 hour fire rating as prescribed in building codes and will be exposed to a structurally significant fire with the maximum heat release rate of 4 MW. Thermal load is supplied using three purpose-built 1 m \times 1.5 m natural gas burners. Mechanical load is applied at six loading points along the centerline of the specimen using hydraulic actuators. As for test variables, two types of beam connections were considered, including (a) bolted/welded double-angles and (b) single-plate shear connections. Slab continuity conditions are provided in two of the five specimens by rotationally restraining the steel reinforcements in concrete slab at the location of support girders.

Design Basis of Composite Beam Specimens

Specimens were designed to represent composite floor beams in a 12.8 m long bay of a mid-rise steel frame building. For structural design, it was assumed that a steel beam was spaced at 3.05 m. A steel beam was made composite with a lightweight concrete slab casted on 76 mm deep profiled steel deck units via 19 mm headed steel shear studs. The spacing of shear studs was 305 mm on center (i.e., every flute of deck units). A 2 hour fire rating of the composite floor assemblies was considered as prescribed in the building codes. The factored (design) floor loads were determined by the gravity load combinations prescribed in ASCE/SEI 7-10 (ASCE, 2010) and using inputs provided by the panel of expert practitioners such as construction live loads. The size and proportion of the prototype composite beam which is assumed to be framing into a 12.8 m by 6.1 m bay and the beam-to-column shear connections were designed in accordance with ANSI/AISC 360-16 (AISC, 2016), AISC Design Guide 3 (West et al., 2003), and SDI C-2011 (SDI, 2011). For Phase 1 test program, the test beams consisted of a 12.8 m long steel beam and a 12.8 m by 1.83 m slab.

Design Loads

Choe, Lisa; Gross, John; Grosshandler, William; Hoehler, Matthew; Ramesh, Selvarajah. "Design and Construction of Long-Span Composite Beam Specimens for Large Structural-Fire Tests." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

The unit weight of a lightweight, lightly reinforced concrete slab used in the test was 18.9 kN/m^3 . The calculated self-weight of the concrete slab casted on profiled steel decking was 2.20 kN/m^2 . The self-weight of the steel beam was 0.51 kN/m. Other ASCE/SEI 7-10 design loads, including a construction live load of 0.96 kN/m^2 , super imposed dead load of 0.48 kN/m^2 , and a live load of 3.35 kN/m^2 , were used to design the composite test beam. The calculated moment and shear demand at ambient temperature was 548 kN-m and 171 kN, respectively. For extreme fire events, the moment and shear demand was reduced to 326 kN-m and 99.4 kN, respectively, in accordance with ASCE/SEI 7-10 load combination ($1.2 \times$ dead load + $0.5 \times$ live load).

Design of Composite Beam

Figure 1 illustrates a cross-section of the test beam. A 20-gauge galvanized 7.62 cm deep fluted steel deck (with dimensions similar to Vulcraft 3VLI20) with 27.6 MPa lightweight concrete was used. The thickness of the concrete slab above the steel deck was 8.26 cm to provide a 2 hour fire rating without additional fire protection coating on the deck. The ribs of the steel deck were oriented perpendicular to the floor beams and spaced at 30.5 cm. A 6x6 W1.4xW1.4 welded wire fabric was placed in the mid height of the concrete above the steel deck as the minimum required shrinkage and temperature control reinforcement. For two specimens with slab continuity, four no. 4 reinforcing bars were placed at the spacing of 45.7 cm on center at the beam ends, which were primarily designed for a crack control of the slab over the support girders. They were placed at the top of the welded wire fabric and their embedded length from the slab end was 76.2 cm.

The W18x35 beam steel beam was designed to support the ASCE 7 construction loads, including the self-weight of a steel beam, a wet concrete slab and construction live loads, and to meet the serviceability requirement (displacement limit) for a non-composite section. The composite action between the beam and slab was provided using 19 mm diameter steel headed studs spaced at 305 mm on center. The total shear connection strength (composite action) was 82 % of the yield strength of the steel beam cross section at ambient temperature. The calculated flexural capacity of the composite section (with the AISC strength reduction factor of 0.9) per ANSI/AISC 360-16 was 695 kN-m which is about 1.27 times the moment demand at ambient temperature. The governing failure mode was fracture of steel shear studs.



Figure 1. Cross-section of the composite beam specimen.

Design of Shear Connections

Choe, Lisa; Gross, John; Grosshandler, William; Hoehler, Matthew; Ramesh, Selvarajah. "Design and Construction of Long-Span Composite Beam Specimens for Large Structural-Fire Tests." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018

Figure 2 shows the shear connections for the test beams which were designed using either double angles or shear tab in accordance with ANSI/AISC 360-16. For double angle connection (Figure 2a), $L5 \times 3 \times 3/8$ angles made of Gr. 36 steel were used. The angle legs were bolted to the beam web using three 19 mm diameter A325 bolts with the spacing of 76 mm. The top bolt was located at 114 mm from the top of the steel beam. The bolt holes in the web were drilled at 76 mm from the column face. The bottom flange of the steel beam was coped. The other legs of angles were shop-welded to a sacrificial plate on columns. The size of fillet weld was 8 mm. The design of the double angle connection was governed by the weld strength in shear, and the corresponding calculated shear capacity was 267 kN (including the AISC strength reduction factor of 0.75).

For the shear tab connection (Figure 2b), a 11 mm thick plate (Gr. 36) was bolted to the web of the steel beam and welded to a sacrificial plate of the support column. The same bolts were placed at 76 mm spacing as used in the angle connection. The location of bolt holes was also remained unchanged. The size of fillet weld was 8 mm fillet weld. The calculated shear capacity was 242 kN (including the AISC strength reduction factor of 0.75).



Figure 2. Shear connections at the end of specimens (a) double angle and (b) single plate. (weld dimension in inches; 1 in. = 2.54 cm)

Passive Fire Protection

The steel beam and connections were coated with sprayed fire resistive materials (SFRM) to meet the 2 hour fire rating as prescribed in the buildings code. In the connection regions, SFRM was sprayed to achieve the 3 hour fire rating which was required for primary members (i.e. columns). Since the thickness of a topping concrete was 8.26 cm (3.25 in.) for the 2 hour fire rating, no additional fire protection was necessary at underside of steel decking. The SFRM product used in this test program was the Southwest Type 5MD, a cementitious gypsum based material manufactured by Carboline. The required thicknesses of this product for the W18x35 steel beam and the connections was 16 mm and 27 mm, respectively, in accordance with the Underwriter Laboratory (UL) directory.

Test Matrix

Table 1 shows the test matrix for this test program. Test variables included shear connection type (i.e., bolted/welded double-angles versus single-plate shear connections) that are

Choe, Lisa; Gross, John; Grosshandler, William; Hoehler, Matthew; Ramesh, Selvarajah. "Design and Construction of Long-Span Composite Beam Specimens for Large Structural-Fire Tests." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018

extensively used in the U.S. construction practice and the presence or absence of slab continuity at the beam ends as shown in Figure 3. The testing is currently ongoing. Specimen 1 was recently tested at ambient temperature to measure the behavior and strength of the test beam as a baseline information that will be used to compare with the fire test results. For the ambient temperature test, the test beam was subjected to a monotonically increasing moment until failure. The maximum moment capacity (M_{max}) was measured at 695 kN-m ± 2 % where the value after ± symbol is the standard uncertainty with coverage factor of 1. Specimens 2 through 5 will be tested under combined mechanical and fire loads. For those specimens, the applied load is equivalent to a bending moment of 326 kN-m that was computed using the ASCE 7 load combination for fire condition (i.e., $1.2 \times \text{dead load} + 0.5 \times \text{live load}$). This gives the applied load ratio approximately equal to $0.5 M_{max}$. All four specimens will be exposed to a structurally significant fire with the maximum heat release rate of 4 MW until incipient failure occurs.

Table 1. Test matrix for the laboratory experiments					
Spaaiman	End connection	Slab Continuity	Mechanical	Eine looding	
specimen	End connection	Provided	loading		
Sussimon 1	double angle	No	Load to failure	Ambient	
Specimen I	double-angle	INO	(M _{max})	Test	
Specimen 2	double-angle	No	$0.5 M_{max}$	4 MW fire	
Specimen 3	double-angle	Yes	$0.5 M_{max}$	4 MW fire	
Specimen 4	Single plate	No	$0.5 M_{max}$	4 MW fire	
Specimen 5	Single plate	Yes	$0.5 M_{max}$	4 MW fire	



Figure 3. Slab end support conditions (a) specimen with slab continuity and (b) specimen without slab continuity.

Construction of Specimens and Test Setup

Materials

The measured and code specified minimum material properties of the steel beam, shear studs, wire mesh, reinforcing bar, and the connection materials such as shear tab, angle, and bolts are listed in Table 2. The elastic modulus (E), yield strength at 0.2 % offset ($\sigma_{y0.2}$), ultimate strength (σ_u), and the percent elongation at the ultimate strength (e_u) were measured from tensile coupon

Choe, Lisa; Gross, John; Grosshandler, William; Hoehler, Matthew; Ramesh, Selvarajah. "Design and Construction of Long-Span Composite Beam Specimens for Large Structural-Fire Tests." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018

tests that were conducted in accordance with the relevant ASTM standards (such as ASTM E8, 2016).

The concrete mix design is provided in Table 3. The guiding principle behind the concrete mix was to have a hardened concrete with mechanical properties typical of what is used in current practice, but with a low propensity for fire-induced spalling; which would add undesired variability to the experiments. To reduce the likelihood of spalling, 2.37 kg/m³ of monofilament polypropylene microfibers (FRC MONO-150) were used in the mix. To further reduce the chance of fire-induced spalling, expanded slate lightweight aggregate with a low water-retention characteristics and high desorption was selected (Pour-Ghaz, 2012) to help expedite the reduction of moisture in the slabs during curing. A large slump was required because sensitive optical fibers were embedded in the concrete to measure strain and temperature in the slab that limited the use of mechanical vibration during casting.

Fabrication of specimen

Figure 4 shows a photograph of the steel beam with metal deck, rebar chair, and shear studs. The flutes of metal deck units were oriented perpendicular to the steel beam and attached to the steel beam using power actuated fasteners. The shear studs that were 13.2 cm in original length were welded to the steel beam through the holes in the metal deck. The height of the shear studs after installation was 12.7 cm. As shown in the figure, shear studs were off centered by 2.5 cm from the center of the valleys of the metal deck toward each end of the specimen (in the 'strong direction' for shear). 3.8 cm high bar chairs were placed on top of the metal deck and 6x6 W1.4xW1.4 welded wire mesh was placed above the reinforcing bar chairs.

The concrete was batched at a local ready-mix concrete plant and trucked to NIST for casting. Because approximately 2.8 m³ of concrete was required per specimen and the trucks held 7.6 m³, two trucks (batches) were required to cast the five specimens. Two of the four specimens to be tested with fire loading were cast from each batch of concrete, with a mix of concrete from the two batches (half the length cast from Batch 1 and half from Batch 2) being used to cast the ambient temperature specimen. Although the mix design was the same for both batches, a larger amount of high-range water reducer (BASF Glenium 7920) was needed in Batch 2 to achieve the target slump; which influence the hardened concrete strength and had consequences for the ambient specimen discussed in a separate paper. Mechanical vibration of the concrete was only performed around edge embed plate and on the slab formwork. The concrete was screed out, but no floating or 'hard troweling' was performed.

	_	Min. sp	ecified	Me	asured materi	al properties	5
Material	Material Spec.	σ _y (Mpa)	σ _u (Mpa)	E (GPa)	σ _{y0.2} (Mpa)	σ _u (Mpa)	eu
Steel beam-Flange	ASTM A992	345	450	210	350	470	19%
Steel beam-Web	ASTM A992	345	450	200	370	470	19%
Shear stud	ASTM A29	350	450	210	410	510	5%
Wire mesh	ASTM A185	450	520	200	730	760	1%
No.4 rebar	ASTM A615 Gr 60	410	620	190	470	710	11%
Shear tab plate	ASTM A36	250	400	210	350	490	20%
Double angle	ASTM A36	250	400	200	380	510	17%
19 mm dia. Bolt	ASTM A325	630	830	210	900	960	6%

Table 2. Measured and minimum specified material properties.

Choe, Lisa; Gross, John; Grosshandler, William; Hoehler, Matthew; Ramesh, Selvarajah. "Design and Construction of Long-Span Composite Beam Specimens for Large Structural-Fire Tests." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018

		w/cm: 0.46	Slump: 8.5±1.0 inch			
		Surface Saturated				
	Material	Dry (SSD), lb	Volume, ft ³			
Cement:	ASTM C-150: Type I/II Lehigh	520	2.65			
Fly Ash:	ASTM C-618: Seperation Technologies Class F	130	0.88			
Aggregate:	ASTM C-33: Carolina Stalite LTWT	890	9.51			
Sand:	ASTM C-33: Howlin Sand	1370	8.41			
Air:	2.5%		0.67			
Water:	ASTM C-1602; ASTM C-1603	300	4.81			
Admixture:	See Below	10	0.07			
	Total	3220	27.00			
	Unit Weight (pcf)	119.2				
	Calculated Equilibrium Dry Density (pcf)	113.6				
Admixtures						
FRC MON	O-150-4 lb/cy					
BASF Glen	BASF Glenium 7920-3 \pm 3 oz/cwt					
BASF Pozz	BASF Pozzolith 322N - 4 ± 2 oz/cwt					
BASF DEL	BASF DELVO Stabilizer- 2 ± 2 oz/cwt					
BASF Rhee	TEC Z-60 - 4 ± 2 oz/cwt					

Table 3. Concrete mix design.

The specimens were cast indoors in a test hall. Immediately after casting, the specimens were covered with wet burlap and then plastic to maintain a wet concrete surface. The burlap was re-wet, as required, for the first 7-days of curing, after which the plastic and burlap was removed. Within the next 7 days, the four specimens to be fire tested were moved to a large underground curing room where the temperature was maintained at 30 °C \pm 2 °C (standard uncertainty) and the target relative humidity of the air was 50 %. Due to problems with the humidity control in the curing room, the achieved relative humidity of the air was around 30 % \pm 10 %. The ambient specimen was cured in the testing hall where the temperature was 22 °C \pm 2 °C (standard uncertainty) and the relative humidity of the air was not measured (typically less than 30 % during the first 3 months after casting; winter). The mean 28-day concrete cylinder strengths were 42.1 \pm 0.9 MPa and 48.7 \pm 1.6 MPa for Batch 1 and Batch 2 concretes, respectively.

After the fire test specimens were cured for about 6 months, the specimens were moved to the laboratory floor and the fire resistive material (SFRM) was sprayed to the steel beam and the connections of the specimens. Figure 5 shows the photographs of specimens after the SFRM was sprayed. The thickness of the SFRM was measured at various locations. Either the SFRM was scraped off or more SFRM was sprayed to maintain a uniform thickness of 16 mm with a tolerance of $\pm 25\%$ of the required thickness as prescribed in building codes.



Figure 4. Placement of metal deck, shear stud, wire mesh.



Figure 5. Specimens after spraying fire resistive materials (a) front view (b) side view.

Test Setup

The test setup is shown in Figures 6(a) and (b). The composite beam specimens were attached to the support columns at the ends using either a double angle connection or a single plate shear connection. The top of the slab, top of the steel beam, and the bottom of the steel beam were at elevations of 3.93 m, 3.77 m, and 3.32 m from the strong floor, respectively. Mechanical load was applied to the specimen at six load points simulating a uniform floor load. Three loading beams were used to apply the load as shown in Fig 7. The hollow steel section (HSS) loading beams were guided such that they can only move down vertically as the specimen deforms. Each loading beam was pulled down by two actuators at its ends. Actuators were placed underneath the basement and high-strength bars were used to transfer the forces from the actuators to the loading beam. As shown in Figures 6 and 7, a loading truss was used to transfer the force from each loading beam to two load points on the specimen.

The support columns were braced using a brace module (Figure 6) to provide lateral stiffness simulating the stiffness provided by the surrounding steel framing to the column in a building. The columns and the brace modules were anchored to the strong floor using post-tensioned high strength bars to prevent slip. The length of the steel beams was 12.3 m. The center-to-center distance between the bolt lines was 12.2 m. The steel beams that had double-angle connections at the end were coped at the lower flange. The flanges of the steel beams with single plate shear connection were not coped. The slab extended to the centerlines of the support columns and was supported on a HSS 5x5x1/2 girder to simulate typical construction conditions. For the specimens with slab continuity (Specimens 3 and 5), the wire mesh and the no.4 crack control reinforcing bars were extended beyond the slab edge and anchored as shown in Figure 3.

The longitudinal sides of the slab were laterally braced at the locations of the loading beams to simulate the restraints provided by the adjacent concrete slab. Figures 8(a) and 8(b) show a schematic and a photograph of the lateral-torsional restraint used in this test, respectively. The actuators that pulled down the loading beam were programmed such that the loading beam is horizontally leveled while it vertically moves up or down along with the specimen during the test. The axial displacement of the loading beam is restrained at its ends using laterally braced columns. Frictionless bearings were used between the HSS tube attached to the specimen and the lateral-torsional restraining system to minimize any force transfer to the edge of the slab.

Choe, Lisa; Gross, John; Grosshandler, William; Hoehler, Matthew; Ramesh, Selvarajah. "Design and Construction of Long-Span Composite Beam Specimens for Large Structural-Fire Tests." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

For the specimens tested under fire conditions, a compartment (total area of 110 m²) constructed surrounding the specimens with a ventilation area of 4.93 m² is used to confine the fire below the composite beam specimen. Figure 9 shows the layout of the compartment wall and the locations of the burners. The enclosure walls were constructed with stiffened sheet steel (18 gauge) protected with 50 mm thick ceramic blanket. Thermal load will be applied using three 1 m \times 1.5 m natural gas burners. A heat release rate of 1.33 MW will be maintained for each burner for 3.5 hours followed by a controlled cooling for 30 minutes. The layout of the compartment wall, the locations of the burners, and the heat release rate of the fire applied during the test were designed to develop a structurally significant fire during the fire test. Computational fluid dynamics (CFD) models were developed using the NIST Fire Dynamics Simulator (FDS) software to estimate the required heat release rate of the fire. Expected temperature time histories of the upper hot gas layer and the lower flange developed using the CFD models are shown in Figure 10. As shown, the lower flange is expected to reach a maximum temperature of 990 °C in about 3 hours 10 minutes of fire exposure.



Figure 6. (a) A 3D schematic and (b) photograph of the structural test setup. Fire compartment walls and ventilation opening not shown.



Figure 7. Mechanical loading system.

Choe, Lisa; Gross, John; Grosshandler, William; Hoehler, Matthew; Ramesh, Selvarajah. "Design and Construction of Long-Span Composite Beam Specimens for Large Structural-Fire Tests." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.



Figure 8. Loading truss and lateral brace setup (a) A 3D schematic and (b) a photograph.



Figure 9. Locations of burners and the layout of the compartment wall (a) front view and (b) plan view. (units in cm)



Figure 10. Expected temperature time history during the tests.

Conclusion

This paper presented the design and construction of the specimens as well as the experimental program used for the large structural-fire tests conducted at the National Fire Research Laboratory at NIST on five 12.8 m long-span composite floor beams. Test variables include the connection type at the beam ends and the slab continuity over the support girders. One specimen is tested at ambient temperature and four specimens are tested under combined structural and fire loads. The test data obtained from this research study can be used to evaluate the performance of long-span composite floor beams and to better understand the behavior of composite floor assembly under fire.

Disclaimer

Certain commercial equipment, instruments, or materials are identified in this report to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Acknowledgements

The support of numerous NIST colleagues on this work is acknowledged and greatly appreciated: Brian Story, Anthony Chakalis, Laurean DeLauter, Philipp Deardorff, Matthew Bundy, Artur Chernovsky, Doris Rinehart, Chao Zhang, Mina Seif, Qi Li, Ana Sauca, Dale Bentz, Scott Jones, Timothy Barrett, Max Peltz, Kevin McGrattan, Morgan Bruns, Randy Shields, Fahim Sadek, Joe Main, Jonathan Weigand, and Jian Jiang. The authors would like to thank Alana Guzetta of US Concrete who prepared the concrete mix design, as well as Roberto Talavera of Superior Concrete who supervised the casting. Expertise on fiber optic instrumentation was provided by Genda Chen, Yi Bao, Steve Guo, Matthew Klegseth, Branko Glisic, and John Reilly.

Choe, Lisa; Gross, John; Grosshandler, William; Hoehler, Matthew; Ramesh, Selvarajah. "Design and Construction of Long-Span Composite Beam Specimens for Large Structural-Fire Tests." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

References

- AISC (2010). Steel Construction Manual, 14th edition, American Institute of Steel Construction (AISC), Chicago, IL
- Almand, K. H., Phan, L. T., McAllister, T. P., Starnes, M. A., & Gross, J. L. (2004). NIST-SFPE Workshop for Development of a National R&D Roadmap for Structural Fire Safety Design and Retrofit of Structures: Proceedings. NIST Interagency/Internal Report (NISTIR)-7133
- Almand, K. H. (2013). Structural Fire Resistance Experimental Research: Priority Needs of US Industry. Springer Science & Business Media
- American Society of Civil Engineers. (2010). *Minimum design loads for buildings and other structures*. ASCE 7-2016, Amer Society of Civil Engineers.
- ANSI/ SDI (2012). "C-2011 Standard for composite steel floor deck slabs", American National Standards Institute/ Steel Deck Institute
- ASTM International. (2016) "Standard Test Methods for Tension Testing of Metallic Materials", *ASTM.E8/E8M-16a*, ASTM International, West Conshohocken, PA, 2016, doi:10.1520/E0008_E0008M-16A
- Guo, S., Bailey, C., G. (2011). Experimental behavior of composite slabs during the heating and cooling fire stages. *Eng. Struct.*, 33 (2) (2011), pp. 563-571.
- Mohammad Pour-Ghaz; Javier Castro; Eileen J. Kladivko; and Jason Weiss: "Characterizing Lightweight Aggregate Desorption at High Relative Humidities Using a Pressure Plate Apparatus." ASCE Journal of Materials in Civil Engineering, Volume 24 Issue 8, August 2012, DOI: 10.1061/(ASCE)MT.1943-5533.0000422.
- NIST (2008). Federal Building and Fire Safety Investigation of the World Trade Center Disaster: Final Report on the Collapse of World Trade Center Building 7 (NIST NCSTAR 1A). National Institute of Standards and Technology, Gaithersburg, Maryland.
- NIST (2015). International R&D Roadmap for Fire Resistance of Structures Summary of NIST/CIB Workshop (NIST Special Publication 1188). National Institute of Standards and Technology, Gaithersburg, Maryland.
- Selden, K., L., Fischer, E., Varma, A., H. (2015). Experimental investigation of composite beams with shear connections subjected to fire loading. J Struct Eng ASCE Rest VA. doi:10.1061/(ASCE)ST.1943-541X.0001381
- Wellman, E. I., Varma, A. H., Fike, R., & Kodur, V. (2011). Experimental evaluation of thin composite floor assemblies under fire loading. *Journal of Structural Engineering*, 137(9), 1002-1016. doi: 10.1061/(ASCE)ST.1943-541X.0000451
- West, M., Fisher, J., Griffis L., G. (2003). "AISC Steel Design Guide 3 (Second Edition):

Choe, Lisa; Gross, John; Grosshandler, William; Hoehler, Matthew; Ramesh, Selvarajah. "Design and Construction of Long-Span Composite Beam Specimens for Large Structural-Fire Tests." Paper presented at Structures Congress 2018, Fort Worth, TX, United States. April 19, 2018 - April 21, 2018.

Serviceability Design Considerations for Steel Buildings". American Institute of Steel Construction (AISC), Chicago, IL

Yang, J. C., Bundy, M., Gross, J., Hamins, A., Sadek, F., & Raghunathan, A. (2015). International R&D roadmap for fire resistance of structures, summary of NIST. In CIB workshop (NIST SP 1188), National Institute of Standards and Technology.



System Lifecycle Handler — Spinning a Digital Thread for Manufacturing

Manas Bajaj, PhD Intercax 47 Perimeter Center East, Suite 410 Atlanta, GA 30338, USA +1-404-592-6897 manas.bajaj@intercax.com

Thomas Hedberg, Jr. Systems Integration Division, Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA +1-301-975-4247 thomas.hedberg@nist.gov

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States. Approved for public release; distribution is unlimited.

Abstract. Transforming the manufacturing economy from paper-based information flows to a seamless digital thread across geographically distributed supply chains has the potential to reduce cycle time by 75% and save manufacturers \$30 billion annually. The "Digital Thread for Smart Manufacturing" project at NIST¹ is developing methods and protocols for completing a digital thread running through design, manufacturing, and product support processes. In this paper, we present a proof-of-concept System Lifecycle Handler (SLH) software environment being developed for the digital thread initiative. The SLH provides services to build, manage, query, and visualize the digital thread by connecting heterogeneous artifacts ranging from requirements and system architecture to PLM/CAD/CAM and simulation models to machines and sensor data streams. The SLH software environment leverages the Syndeia platform, and exposes its capabilities via a web dashboard and standard REST/HTTP API.

Motivation: Model-Based, Smart Manufacturing

A recent economic study released by the National Institute of Standards and Technology (NIST) estimates that a cyber-physical infrastructure enabled by linked-data and systems-thinking would save United States (U.S.) manufacturing \$100 Billion annually (Anderson, 2016). Manufacturing accounts for one-third of the U.S. economy. The total manufacturing value chain is estimated to be \$5.5 Trillion (MAPI Foundation, 2017). The value chain starts at the mining and processing of raw materials, moves through the production systems, and ends with wholesaling, transportation, retail distribution, and servicing activities. \$3.60 of value is added to the economy for every one dollar of manufacturing value (Meckstroth, 2017). In addition, Meckstroth estimates that 3.4 full-time equivalent (FTE) non-manufacturing sector jobs are created for every one FTE manufacturing job. All of this points to final manufacturing demand accounting for 34% of U.S. Gross Domestic Product (GDP), with 59% of manufacturing demand attributed to exports (Giffi et al., 2017).

Between 1998 and 2015, manufacturing productivity grew three times faster than the service economy (MAPI Foundation, 2015). While manufacturing exhibited growth and success, significant opportunity remains. One study found that simply transitioning from paper-based processes to (digital) model-based processes, for design through production, would achieve an approximately 75% reduction in cycle-time (Hedberg Jr et al., 2016). Further, enhanced sensing and monitoring,

Bajaj, Manas; Hedberg, Thomas. "System Lifecycle Handler – Spinning a Digital Thread for Manufacturing." Paper presented at 28th Annual INCOSE International Symposium, Washington, DC, United States. July 7, 2018 - July 12, 2018.

¹ https://www.nist.gov/progr<u>ams-projects/digital-thread-smart-manufacturing</u>

seamless transmission of digital information, and advances in analyzing data and trends would save manufacturers \$30 Billion annually (Anderson, 2016).

But industry is also approaching the fundamental limits of what its tools and processes can manage. Data, system, and viewpoint interoperability remains a challenge for industry (Trainer et al., 2016, Ruemler et al., 2016, Feng et al., 2017, Hedberg Jr et al., 2017b, Regli et al., 2016). Industry needs connected systems and linked-data federated across enterprises. Point-to-Point interoperability (e.g., file-based data translation) is not enough. Industry must stop thinking about data interoperability through mapping exercises and instead focus on domain and interface interoperability.

However, domain interoperability (e.g., design to manufacturing) requires a normalized method for accessing and contextualizing data at different points of the product lifecycle. Often the focus of interoperability has been confined to the formats in which the data is stored and not the semantics. Focusing on the information for the "thing" being represented in the data would help industry keep more focus on solving problems for the thing than focusing on communication and data exchange. Further, actors in industry must also consider the interfaces, outputs and inputs, on the boundaries of their domains. Standard interfaces between domains must be developed and understood to support efficient flow of required information through the product lifecycle. This effective communication of information brings with it an almost \$8 Billion return-on-investment annual opportunity (Anderson, 2016).

We recognize that context varies based on the phase of the lifecycle (e.g., design, manufacturing, quality). In addition, context varies based on the level of interaction with data (e.g., systems, operations, enterprises) (Regli et al., 2016). Thus, geometry and manufacturing specification is not enough to define products – behavioral and contextual definitions are required too. Furthermore, all three aspects must be generated, documented, and communicated using an agile and dynamic method.

The digital thread concept shows promise for supporting industry's needs. This paper outlines the digital thread concept, describes a methodology for linking and tracing data throughout the product lifecycle using the digital thread, and presents a prototype implementation of the methodology to enable the digital thread.

Digital Thread for Manufacturing

The digital thread is an integrated information flow that connects all the phases of the product lifecycle using accepted authoritative data sources, e.g. requirements, system architecture, technical data package (US Department of Defense), three-dimension (3D) computer-aided design (CAD) models (Kraft, 2016, Hedberg Jr et al., 2016, Wardhani and Xu, 2016), and project tasks. The product lifecycle is a complex heterogeneous system (or system-of-systems depending on how one draws the system boundaries). The aim of digital thread is to develop an integration framework that brings all phases and systems of a product lifecycle together to enable efficient and effective measurements of the lifecycle in support of data-driven methods. Specific interests relate to knowledge building, decision support, requirements management, and control. A major goal for enabling the digital thread is linking universally heterogeneous information systems and data sets across the various domains (e.g., design, manufacturing, quality) of the product lifecycle dynamic ways without requiring one-to-one data mapping.

Hedberg Jr et al. (2017a) proposes a Lifecycle Information Framework and Technology (LIFT) idea to support effective implementation of the digital thread. The framework is comprised of three levels: (1) product lifecycle data, (2) data certification and traceability, and (3) data-driven applications. In general, the LIFT ideas strives to link data across the information silos, while building trust through traceability, for driving applications with data. Hedberg Jr et al. (2017a) also introduces an example schematic (Figure 1) of technology for linking and integrating systems. The proposed technology

would utilize agent-based adapters connected to the typical client support systems (e.g., product data management, enterprise resource planning, manufacturing execution, quality management) used in industry. The adapters would support micro-services (e.g., query, data retrieval, control) that would manage the curation, discoverability, retrieval, and observation of data from across the federated product lifecycle. The adapters connect the various client support systems through a master handle system that acts as a data "traffic cop" to aggregate all related data to a particular thing and present unified results back to user in the client support system with which he/she is most comfortable working (Hedberg Jr et al., 2017a).

Our work builds on the framework and technology presented by Hedberg Jr et al. (2017a). We consider the master handle system to be a type of "in-out bus" for data to flow and link. This manufacturing in-out bus (Figure 2) would enable linking people, machines, federated data, things, and systems together – thus, supporting a digital-thread infrastructure that brings full lifecycle cyber-physical connections to reality. Links and relationships between artifacts would be generated and tracked in near-real-time using the agent-based adapters. Generating a graph of all data and links across the product lifecycle for a product is assumed to be a large unstructure data set. However, by tracing the links and relationships while users interact with the data, structure could be inferred through observation and lifecycle-wide information models could be generated dynamically. The digital thread is inspired by the ideas of Digital Twin (Glaessgen and Stargel, 2012) and Semantic Web² in the context of manufacturing.



Figure 1. Example technology schematic for partially enabling the digital thread with linked data through the use of agent-based adapters, applications programming interfaces, and a master data handling system [from (Hedberg Jr et al., 2017a)].

² <u>https://www.w3.org/standards/semanticweb/</u>


Figure 2. Supported artifacts of a manufacturing in-out bus-like handle system.

System Lifecycle Handler for Digital Thread

The *System Lifecycle Handler* (SLH) software platform realizes the LIFT framework and provides services and APIs to build, manage, update, visualize, and query the digital thread. In this section, we present the functional requirements for the System Lifecycle Handler system, followed by a description of its conceptual architecture. The System Lifecycle Handler builds on the Syndeia software platform developed by Intercax. We also present an overview of the Syndeia platform, specifically focusing on the key services used by the System Lifecycle Handler.

Requirements

In this section, we present the functional requirements for the System Lifecycle Handler (SLH) system.

Identifying artifacts and relationships in the digital thread – The SLH system should be able to identify various types of artifacts and their inter-relationships in the digital thread. This includes both virtual artifacts such as models, model elements, data, documents, and software, as well as physical artifacts such as organizations, machines, tools, factories, operators (human and robotic), sensor devices, raw materials, and physical parts/assemblies. The SLH system should also be able to navigate relationships between the artifacts, such as the relationships between the design model of a part, the software used for authoring that model, the machine instructions used for manufacturing the part, the physical part produced by the machine, and the quality inspection report on that part. Without the relationships, we would end up with silos of artifacts and not a continuous digital thread.

Unique global identifier system (GID) for the digital thread – The SLH system should provide a unique, global identifier system (GID) for addressing and searching all artifacts and their inter-relationships in the digital thread. This is a challenging task because the information about the artifacts participating in the digital thread originates from multiple repositories, databases, and software environments—requirements managed in requirement management systems (e.g., Jama or DOORS), system architecture models in SysML modeling tools (e.g. MagicDraw, Rhapsody), part structure BOM and CAD models managed in a PLM system (e.g., Teamcenter, Windchill), and manufacturing machine instructions and data as MTConnect streams. Each type of repository/database provides its own identification system that is local to artifacts and relationships managed in it. However, when we build a digital thread by federating artifacts from multiple repositories, we need a global identifier system that can be used to address any artifact or inter-relationship throughout the product lifecycle.

Multiple-step resolver system for global identifiers – The digital thread has to work in a world of computer firewalls, network security, and multi-layer authentication. The global identifier for an

artifact in the digital thread may not be a single URI, as in a generic Linked Data³ approach, but an ordered set of addresses (URIs or other identifier types) that have to be resolved recursively to navigate through multiple layers of namespaces, firewalls, and authentication servers. The digital thread includes data that is originating from multiple sources, e.g. flat files (spreadsheets and documents), computer models, real-time data streams, and hardware. The identifier system for the data may range from cells in a spreadsheet to unique string-based identifiers for a part in a computer model, to URIs provided by a REST/HTTP service.

Consider the example shown in Figure 3 where the design model of a part managed in a PLM system needs to be assigned to a specific machine in the factory that will make this part. For simplicity, we will assume that the same organization is designing and manufacturing the part. In a globally distributed supply chain, the challenge presented here will be compounded. The abbreviation A(x, base) is used to represent the address of an artifact x in the context of the *base* artifact. The address can be a URI or some form of an identifier that can be resolved. At the highest level, an organization artifact (Org) may have a gateway server available on the internet (world wide web) for all incoming requests, denoted as A(Org, www) in the figure. Next, the gateway servers for the various divisions in the organization are generally not reachable directly from the open internet due to firewalls, but reachable from the organization's gateway server. A(DesignDiv,Org) and A(ManufDiv,Org) are the addresses of the design division and manufacturing division of the organization in context of the organization. Similarly, A(PLM.DesignDiv) is the address of the PLM server



Figure 3. Multi-level addresses for locating artifacts across enterprise layers

reachable from the design division, and A(P,PLM) is the address of the part P in the context of the PLM server. Hence, the global identifier and address for part P is an ordered set of addresses: $\{A(Org, www), A(DesignDiv, Org), A(PLM, DesignDiv), A(P, PLM)\}$. To reach part P, the SLH should provide a resolver system that recursively traverses the chain of addresses, authenticating the request at each base artifact to reach the next artifact. A similar resolution process has to be followed to reach machine M. Once the part P and machine M can be reached uniquely in this manner, we can establish a relationship that part P is made by machine M. A collection of these relationships spins the digital thread.

Common handler system for global identifiers – The SLH should provide a common handler system in which we can input the global id of an artifact and lookup its meta-data, e.g. the name, type, address, and any other attributes of the artifact that are published openly. If additional details for the artifact are requested, the handler system can invoke the resolver system (as described above) to fetch the complete artifact. This handler system is similar to the Distributed Object Identifier (DOI) handler system available at http://dx.doi.org/ where one can provide the DOI number for any document to lookup its meta-data and obtain the URI to the original source of the document on the internet. One would still need to be authenticated to fetch/download the document. For the SLH system, we will be using *mfg.io* as the domain name for handle server.

³ See the "References" section.

Search and query artifacts and their versions in the digital thread – The SLH system should provide a common platform to search, query, and generate/register new artifacts. This includes, for example, connecting to a requirements management system and querying requirements, their relationships, and attributes; or connecting to a PLM system and iterating over the multi-level BOM for a hardware assembly and querying mass properties (e.g. bounding box) for parts; or connecting to a machine and analyzing real-time data streams as a part is being made.

Building a digital thread by establishing connections between artifact – The SLH system shall provide capabilities to generate/register artifacts in the digital thread and to link them using connections. This includes, for example, generating design models from requirements (design synthesis), or generating simulation models and manufacturing process plans from design models, or registering new machines and machine configurations on a factory floor. Further, the SLH system should enable automated generation of connections between artifacts when one is generated from the other e.g. connections between design and analysis models are automatically generated when analysis models are generated from design models. Manual creation of connections between artifacts can be laborious.

Figure 4 shows a representative example of a digital thread where requirements in a requirements management system are connected to mechanical design model elements in a CAD system. The latter is then connected to CAM models, which are then connected to quality inspection tests. We can see both intra-model connections (solid arrows) between artifacts originating in the same repository or system, and inter-model connections (dashed red lines) between artifacts in different repositories.



Figure 4. A representative example of a digital thread for manufacturing

Traceability, impact analysis, and continuous validation and integration of the digital thread – The greatest impact of the digital thread is in the continuous analyses that can be performed on it. At the first level, we have basic traceability when one can traverse the digital thread using the intra- and inter-model connections, starting with any artifact. However, as shown in Figure 4, a greater capability is to use graph pattern matching and graph traversals to assess the upstream and downstream impact of changes in any artifact [Bajaj, Backhaus, et al., 2017]. For example, computing the downstream impact of changes in a requirement, or querying upstream requirements and analyses done on a part when it fails during operation.

Digital thread as the information flow highway –The inter-model connections in the digital thread can provide highways for seamless flow of information in multiple directions. For example, if design parameters of a part in a CAD model are connected to machine instructions for manufacturing that part, then changes in the design parameters can be synchronized to machine instructions, or conversely tolerances achievable by machines can be communicated to designers working with part geometry. The flow of information across the inter-model connections may not just be primitive data but complex model structures, such as synchronizing simulation models (e.g. Simulink model) connected to system architecture models (e.g. SysML internal block structure).

Visualization and graph queries on the digital thread – The SLH system should also provide capabilities to visualize the digital thread and perform graph queries to answer questions.

Exposing the digital thread as an API – The SLH system should expose the digital thread as a service with a RESTful API so that higher-level analysis and verification applications can be built and deployed for design, manufacturing, and operation teams.

Handle System (mfg.io) Web Browser Digital Thread Dashboard *Core Services* (*Syndeia platform*) REST (HTTP) API API Connection Service Graph Service i

SLH Software Architecture

Figure 5 illustrates a high-level conceptual architecture of the System Lifecycle Handler (SLH) that

satisfies the functional requirements presented above and implements the LIFT framework.

Figure 5. High-level conceptual architecture of the System Lifecycle Handler

In its current state, the SLH system includes three major components:

Handle System – We have deployed a Handle.Net server at <u>https://hdl.mfg.io/</u>. Using the global ID for any artifact in the digital thread, the Handle System can provide metadata about that artifact, such as (a) name of the artifact, (b) type of the artifact and the schema, (c) author of the artifact, (d) location of the artifact, and (e) other attributes. Figure 6 illustrates the metadata for the design of a heat sink for avionics printed circuit board, as provided by the Handle System accessed via a web browser. Each artifact in the digital thread graph has a URL which the resolver can use to fetch the actual artifact after necessary authentication.

Digital Thread Dashboard – This is a web dashboard that provides the ability to browse, query, and visualize the digital thread or any of its sub-graphs. This dashboard is currently being built using the Syndeia platform and will provide views similar those presented in Figure 7 for accessing and browsing various repositories and databases, Figure 8 for generating artifacts and creating

connections between artifacts in a digital thread, Figure 9 for verifying the consistency of connected artifacts, Figure 10 and Figure 11 for visualizing the digital thread graph and sub-graphs, and Figure 12 for executing graph queries on the digital thread. The Digital Thread Dashboard invokes the core services of the System Lifecycle Handler system provided by the Syndeia platform.

\leftarrow	C Secure	https://hdl.mfg.io/20.	500.11993/d2mi.cad.827-9999-904.stp?noredirect	☆
		+ ®		
Пс	inule.ive	L		_
Hand	lle Values for: 20.5	500.11993/d2mi.cad.82	7-9999-904.stp	
Index	Туре	Timestamp	Data	_
1	URL	2017-08-31 03:29:032	https://smstestbed.nist.gov/tdp/d2mi/CAD/827-9999-904.stp	
2	ТҮРЕ	2017-08-31 03:29:032	cyber.data.model.design	
3	SCHEMA	2017-08-31 03:29:032	http://schema.org/ProductModel	
4	DATE_CREATE	2017-08-31 03:29:032	2 2017-07-06	_
5	ATTRIBUTE	2017-08-31 03:52:132	<pre>2 { "@context": "http://schema.org", "@graph": [{ "@id": "#model", "additionalType": "http://www.productontology.org/id/heat_sink", "description": "A heat sink for an aerospace avionics printed circuit board", "gin13": "6921407390089", "name": "Aluminum Heatsink" }, { "@type": "DigitalDocument", "fileFormat": "application/step", "about": "Design model for a heat sink for an aerospace avionics printed circu "author": "https://hdl.mfg.io/20.500.11993/nist.tdh1", "additionalType": "http://www.productontology.org/id/heat_sink", "hasDigitalDocumentPermissionType", "permissionType": "http://schema.org/WritePermission", "grantee": { "@type": "DigitalDocumentPermissionType", "gentee": { "@type": "Thtps://hdl.mfg.io/20.500.11993/nist.tdh1" } }, {</pre>	it bc

Figure 6. Metadata for any artifact in a digital thread available via the handle server (hdl.mfg.io)

Syndeia Platform Services – The System Lifecycle Handler leverages the core services and the API of the Syndeia platform. Syndeia[®] is a software environment for integrated model-based engineering. It provides core services for weaving a Total System Model graph of a product/system by combining models and data originating in a variety of software repositories and tools, such as SysML tools, PLM and ALM environments, databases, requirements and project management systems, databases, and simulation tools. Syndeia leverages open standards such as STEP, REST/HTTP, and OSLC.

The four main services of Syndeia that are being leveraged for the System Lifecycle Handler are presented below. These services are available via a graphical user interface and a standard REST/HTTP API.

(a) Repository Service provides the capability to connect to enterprise data repositories, such as PLM systems, ALM systems, requirements management systems, databases, and simulation tools/environments. Users can search, browse, and query artifacts in various repositories using a common platform.



Figure 7: Repository view in the Syndeia Dashboard

- (b) *Artifact Service* provides the capability to search and query the attributes, structure, and interfaces of all artifacts in the digital thread, from CAD models to machine tools, as shown for PLM systems (Windchill, Teamcenter) and a MySQL database in Figure 7.
- (c) Connection Service provides the capability to create inter-model relationships between artifacts in various repositories or perform model transformations to generate artifacts and connections, as shown in Figure 8. This service also provides the ability to compare and synchronize connected artifacts, as shown in Figure 9 below.



Figure 8: Drag and drop interface to generate models and create inter-model connections

	🙀 Repository Manager 🔁 Connection Manager 🔡 Connection Browser 🐻 Connection Summary 📲 Comparison Result 🔧 Settings					
0,	Q- Type here to filter connections Clear Export to Excel					
Co	nn ID 🔻	Source 🧅 🔻	Target 👻	Latest Target 🔹	Comment 👻 🖽	
8	😽 e3f03	Unmanned Aerial Vehicle	000464/C;2-Unmanned Aerial Vehicle	000464/C;2-Unmanned Aerial Vehicle	The block Unmanned Aerial Vehic 🔺	
	-	wimax module : WiMax Module	wimax module : 000472/A;1-WiMa	wimax module : 000472/A;1-WiMax M	Part property wimax module and	
		visual camera : Visual Camera	visual camera : 000475/A;1-Visual	visual camera : 000475/A;1-Visual Ca	Part property visual camera and	
		trackers : Sensor			Part property trackers has no co	
		thermal camera : Thermal Camera	thermal camera : 000476/A;1-Ther	thermal camera : 000476/A;1-Therma	Part property thermal camera an 😑	
		software : Software System	software : 000487/B;1-Software S	software : 000487/B;1-Software Syst	Part property software and part	
		payload controller : Payload Controller	payload controller : 000470/A;1-P	payload controller : 000470/A;1-Payl	Part property payload controller	
		modem : Spread Spectrum Radio M	modem : 000474/A; 1-Spread Spec	modem : 000474/A;1-Spread Spectru	Part property modem and part o	
		ir detector : Wide Angle IR Detector	ir detector : 000477/A;1-Wide Ang	ir detector : 000477/A; 1-Wide Angle I	Part property ir detector and pa	
		gps : GPS	gps : 000466/A;1-GPS	gps : 000466/A;1-GPS	Part property gps and part occu	
		gprs module : GPRS UMTS Module	gprs module : 000473/A; 1-GPRS U	gprs module : 000473/A; 1-GPRS UMT	Part property gprs module and p	
		flight controller : Flight Controller	flight controller : 000469/A;1-Fligh	flight controller : 000469/A; 1-Flight C	Part property flight controller an 👻	

Figure 9: Compare interface for comparing the connected artifacts and producing a report

(d) Graph Service provides the capability to visualize, query, and traverse the digital thread, as shown in Figure 10, Figure 11, and Figure 12. This service leverages production-strength graph databases (e.g. Neo4j) to manage the digital thread graph and query/traverse it using graph pattern matching languages (e.g. Neo4j Cypher). The services also provides various forms of graph visualization. For example, Figure 10 shows all the inter-model connections between the models participating in a digital thread at a given time.



Figure 10: Visualization of inter-model connections in the digital thread at a given point in time

Figure 11 illustrates the result of progressively querying the inter- and intra-model relationships in the digital thread starting with the SysML block for a Hybrid Powertrain of an automotive vehicle (LHS node). Expanding the intra- and inter-model relationships shows connections to issues in JIRA, simulation models in Simulink, parts in Teamcenter, and control system software in GitHub, all of which are related to the Hybrid Powertrain at various degrees-of-separation.



Figure 11: Traversing the digital thread graph starting with an artifact

Figure 12 illustrates graph queries performed on the digital thread. The LHS of the figure shows the result of a pattern matching query where we are providing two artifacts (nodes) and checking if a path exists between those artifacts using intra- and inter-model connections. The two artifacts are: (1) Electrical System part managed in Windchill PLM system (red node), and (2) UAV system represented as a SysML block (yellow node). The response of the query is the complete path using intra- and inter-model relationships. The RHS of the figure is the response of a similar graph pattern matching query to assess if changing a requirement in DOORS-NG (grey node) will impact the hardware design of the UAV platform assembly managed in Windchill (red node).



Figure 12: Graph queries on the digital thread graph

All Syndeia services presented above are exposed using standard REST/ HTTP API. Figure 13 shows an example of a REST GET request to get all repositories participating in the digital thread. The result of the GET request is shown on the RHS of the figure.

	History Collections	▶ Rep	ositorie	es - All						Examples (0) 🔻
All	Me Team	GE	t v	http://localhos	Params	Send 🗡	Save	~	Status: 200 OK Time: 164 ms	
POST	Signup	Auth	Heade	ers (2) Body	Pre-req. Test	S		Code	Body Cookies Headers (3) Test Results	
POST	SignIn		Key		Value	Bulk Edit	Preset	s 🔻		
GET	Repositories - All	~	Conten	t-Type	application/jso	n				
GET	Repositories - By type		X-Auth-	Token	eyJ0eXAiOiJKV1	QiLCJhbGciOlJIUzI	1NiJ9.eyJz	dWI	2 - {	
GET	Repository - By gid		New ke						3 "id": "12b171a1-c0b2-11e7-bc	c99-11f811851e1d",
GET	Repository - By id								5 "createdDate": 1509725670842	2,
GET	Containers - All								6 "modifiedBy": "12b171a0-c0b2	2-11e7-bc99-11f811851e1d"
GET	Containers - Syndeia								7 "modifiedDate": 150972567084	42,
GET	Artifacts - All								8 "key": "", 9 "aid": "lama @ Intercax".	
GET	Artifacts - By id								10 "name": "Jama @ Intercax",	
GET	Relations - All								11 "description": "", 12 "attributes": {}.	
GET	Relations - By id								13 "type": "Jama repository",	and and and
GET	Relations - By type								15 "authentication": {},	amacloua.com",
GET	Relations - By source id								16 "isDisabled": false	
GET	Relations - By target id								18 - {	
GET	Relations - By target repo id								19 "id": "37878401-c0ae-11e7-bc	c99-11f811851e1d",
GET	Relations by Container id								21 "createdDate": 1509724014656	5,
POST	Create relation								22 "modifiedBy": "37878400-c0ae	e-11e7-bc99-11f811851e1d"
POST	Create artifact								23 "modifiedDate": 150972401465	56,
POST	Create repository								24 "key": "", 25 "gid": "DOORS 602",	
	Syndeia Cloud Remote								26 "name": "DOORS 602",	
	13 requests								28 "attributes": {},	

Figure 13: Syndeia services in the System Lifecycle Handler are exposed using a REST/HTTP API (Example GET request shown in Postman REST client)

Summary

In this paper, we have presented the motivation and foundations for a digital thread for manufacturing. The work presented in this paper is being performed under a grant project of the NIST Digital Thread initiative. The paper presents the *System Lifecycle Handler* software environment for spinning, managing, visualizing, and querying a digital thread for products/systems. We have presented the key functional requirements and the high-level conceptual architecture of the System Lifecycle Handler. Further, we have also presented the current state of the development of the System Lifecycle Handler, including the handle server as a common metadata registry for all artifacts in the digital thread; a web dashboard; and core services to connect to data repositories in the digital thread, query and search artifacts, create and synchronize connections between the artifacts, and query and visualize the digital thread. These core services build upon the Syndeia platform, and are exposed via REST/HTTP API endpoints.

References

- Anderson, G. 2016. The Economic Impact of Technology Infrastructure for Smart Manufacturing. Gaithersburg MD: National Institute of Standards and Technology.
- Bajaj, M., Backhaus, J., Walden, T., Schreiber, C., Waikar, M., Zwemer, D. "Graph-Based Digital Blueprint for Model Based Engineering of Complex Systems". 27th Annual INCOSE International Symposium, Adelaide (Australia), Jul 15-20, 2017.
- Bajaj, M., Cole, B., Zwemer, D. "Architecture to Geometry Integrating System Models with Mechanical Design". AIAA Space 2016 Conference, Long Beach, CA, USA, Sep 13-16, 2016. https://goo.gl/S5LkT4
- Bajaj, M., Zwemer, D., Yntema, R., Phung, A., Kumar, A., Dwivedi, A., Waikar, M. "MBSE++ -Foundations for Extended Model-Based Systems Engineering Across System Lifecycle". 26th Annual INCOSE International Symposium (IS 2016) Edinburgh, Scotland, UK, July 18-21, 2016. <u>https://goo.gl/qpFaOT</u>
- CATIA (Dassault Systèmes) http://www.3ds.com/products-services/catia, as accessed on Nov 9, 2017

Creo (PTC) - http://www.ptc.com/cad/creo, as accessed on Nov 9, 2017

Dropbox - https://www.dropbox.com/, as accessed on Nov 9, 2017

Egnyte - https://www.egnyte.com, as accessed on Nov 9, 2017

Enterprise Architect (Sparx Systems) - <u>http://www.sparxsystems.com/products/ea/</u>, as accessed on Nov 9, 2017

Enovia (Dassault Systèmes) - <u>http://www.3ds.com/products-services/enovia</u>, as accessed on Nov 9, 2017

- FENG, S. C., BERNSTEIN, W. Z., HEDBERG, J. T. & BARNARD FEENEY, A. 2017. Toward Knowledge Management for Smart Manufacturing. *Journal of Computing and Information Science in Engineering*, 17, 031016-031016-9.
- Fisher, A., Nolan, M., Friedenthal, S., Loeffler, M., Sampson, M., Bajaj, M., VanZandt, L., Hovey, K., Palmer, J. and Hart, L. (2014), "Model Lifecycle Management for MBSE". INCOSE International Symposium, 24: 207–229. doi: 10.1002/j.2334-5837.2014.tb03145.x (PDF available at http://goo.gl/ZWywBH)
- GIFFI, C., RODRIGUEZ, M. D. & MONDAL, S. 2017. A look ahead: How modern manufacturers can create positive perceptions with the US public. Washington DC: The Manufacturing Institute.
- FMI http://fmi-standard.org/, as accessed on Nov 9, 2017
- Git https://git-scm.com, as accessed on Nov 9, 2017
- GitHub https://github.com/, as accessed on Nov 9, 2017
- Glaessgen, E and Stargel, D., "*The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles*", 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Structures, Structural Dynamics, and Materials and Co-located Conferences, 23-26 April 2012, Honolulu, Hawaii, <u>https://doi.org/10.2514/6.2012-1818</u>

Google Drive - https://www.google.com/drive/, as accessed on Nov 9, 2017

- Handle.net (CNRI) http://handle.net/, as accessed on Nov 9, 2017
- Hedberg Jr, T., Barnard Feeney, A., Helu, M. & Camelio, J. A. 2017a. "Towards a Lifecycle Information Framework and Technology in Manufacturing". Journal of Computing and Information Science in Engineering, 17, 021010-021010-13.
- Hedberg Jr, T. D., Hartman, N. W., Rosche, P. & Fischer, K. 2017b. "Identified research directions for using manufacturing knowledge earlier in the product life cycle". International Journal of Production Research, 55, 819-827.
- Hedberg Jr, T. D., Lubell, J., Fischer, L., Maggiano, L. & Barnard Feeney, A. 2016. "Testing the Digital Thread in Support of Model-Based Manufacturing and Inspection". Journal of Computing and Information Science in Engineering, 16, 1-10.
- ISO 10303 (STEP) http://goo.gl/qH7Rdw, as accessed on Nov 9, 2017
- JIRA (Atlassian) https://www.atlassian.com/software/jira/, accessed on Nov 9, 2017
- JSON http://www.json.org/, accessed on Nov 9, 2017

KRAFT, E. M. 2016. "The Air Force Digital Thread/Digital Twin - Life Cycle Integration and Use of Computational and Experimental Knowledge". 54th AIAA Aerospace Sciences Meeting. American Institute of Aeronautics and Astronautics.

Linked Data - http://linkeddata.org/, as accessed on Nov 9, 2017

MagicDraw (No Magic) - http://www.nomagic.com/products/magicdraw.html, as accessed on Nov 9, 2017

MAPI Foundation. 2015, Accessed: 2017-08-21. "Facts About Modern Manufacturing" [Online]. Arlington VA: MAPI Foundation, http://www.webcitation.org/6t5JUs9ye

- MAPI Foundation. 2017, Accessed: 2017-08-21. *Myth-Busting American Manufacturing* [Online]. Arlington VA: MAPI Foundation. <u>http://www.webcitation.org/6t5loctUk</u>
- Meckstroth, D. J. 2017, Accessed: 2017-08-21. A New Model for Manufacturing's Multiplier Effect [Online]. Arlington VA: MAPI Foundation. http://www.webcitation.org/6t5lw6KtP
- MATLAB/Simulink (MathWorks) http://in.mathworks.com/products/simulink/, as accessed on Nov 9, 2017

Mathematica (Wolfram Research) - http://www.wolfram.com/mathematica/, as accessed on Nov 9, 2017

Melody (Intercax) - http://intercax.com/products/melody/, as accessed on Nov 9, 2017

MTConnect - http://www.mtconnect.org/, as accessed on Nov 9, 2017

Neo4j (Neo Technology) - https://neo4j.com/, as accessed on Nov 9, 2017

- NX (Siemens) http://www.plm.automation.siemens.com/en_us/products/nx/, as accessed on Nov 9, 2017
- OSLC http://open-services.net/, as accessed on Nov 9, 2017
- ParaMagic® (Intercax) http://intercax.com/products/paramagic/, as accessed on Nov 9, 2017
- ParaSolver (Intercax) http://intercax.com/products/parasolver/, as accessed on Nov 9, 2017
- PLCS https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=plcs, as accessed on Nov 9, 2017
- Postman https://www.getpostman.com/, as accessed on Nov 9, 2017
- PTC Integrity Modeler (formerly Artisan Studio from Atego) -

http://www.ptc.com/model-based-systems-engineering/integrity-modeler, as accessed on Nov 9, 2017

- QIF http://qifstandards.org/, as accessed on Nov 9, 2017
- RDF http://www.w3.org/RDF/, as accessed on Nov 9, 2017
- REGLI, W., ROSSIGNAC, J., SHAPIRO, V. & SRINIVASAN, V. 2016. The new frontiers in computational modeling of material structures. *Computer-Aided Design*, 77, 73-85.
- ReqIF http://www.omg.org/spec/ReqIF/, as accessed on Nov 9, 2017
- REST http://www.w3.org/2001/sw/wiki/REST, as accessed on Nov 9, 2017
- Rhapsody (IBM) http://goo.gl/qMXwn6, as accessed on Nov 9, 2017
- Ruemler, S. P., Zimmerman, K. E., Hartman, N. W., Hedberg, J. T. & Barnard Feeny, A. 2016. Promoting Model-Based Definition to Establish a Complete Product Definition. *Journal of Manufacturing Science and Engineering*, 139, 051008-051008.
- Semantic Web https://www.w3.org/standards/semanticweb/, as accessed on Nov 9, 2017
- SOAP http://www.w3.org/TR/soap/, as accessed on Nov 9, 2017
- Solvea (Intercax) http://intercax.com/products/solvea/, as accessed on Nov 9, 2017
- Subversion (Apache) https://subversion.apache.org, as accessed on Nov 9, 2017
- Syndeia (Intercax) http://intercax.com/products/syndeia/, as accessed on Nov 9, 2017
- Systems Modeling Language (SysML, OMG) <u>http://www.omgsysml.org/</u>, as accessed on Nov 9, 2017 Teamcenter (Siemens PLM) - <u>http://goo.gl/bv7iEB</u>, as accessed on Nov 9, 2017
- Trainer, A., Hedberg Jr, T., Barnard Feeney, A., Fischer, K. & Rosche, P. "Gaps Analysis of Integrating Product Design, Manufacturing, and Quality Data in the Supply Chain using Model-Based Definition". 2016 Manufacturing Science and Engineering Conference, 2016 Blacksburg VA. {American Society of Mechanical Engineers}.
- UPDM http://www.omg.org/spec/UPDM/, as accessed on Nov 9, 2017
- US DEPARTMENT OF DEFENSE 2013. Standard Practice: Technical Data Packages.
- Wardhani, R. & Xu, X. "Model-based manufacturing based on STEP AP242". 12th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA), 29-31 Aug. 2016 2016 Auckland, New Zealand. 1-5.
- Windchill (PTC) <u>http://www.ptc.com/product-lifecycle-management/windchill</u>, as accessed on Nov 9, 2017
- WSDL https://www.w3.org/TR/wsdl, as accessed on Nov 9, 2017

Biography



Manas Bajaj, PhD is the Co-Founder and Chief Systems Officer at Intercax. He has led multiple government and corporate sponsored R&D projects over last 10 years, including SBIR Phase 1 & 2 awards. He has led the development of several commercial software applications, including the *Syndeia* platform referenced in this paper. Dr. Bajaj is also leading the development of the *System Lifecycle Handler* software framework at Intercax, in collaboration with NIST. Dr. Bajaj

earned his PhD (2008) and MS (2003) in Mechanical Engineering from Georgia Tech, and BTech (2001) from Indian Institute of Technology (IIT), Kharagpur, India. He has been actively involved in the development of the OMG SysML standard and the ISO STEP standards, and is a Content Developer for the OCSMP certification program. Dr. Bajaj is the author of numerous technical papers and articles. He is a co-developer of a widely popular SysML and MBSE training program with over 4500 participants since 2008. He enjoys playing tennis and cricket for leisure.



Thomas Hedberg, Jr. is a member of the Systems Engineering group in the Systems Integration Division (SID) of the Engineering Laboratory (EL) at the National Institute of Standards and Technology (NIST). Mr. Hedberg is the Project Manager of the Digital Thread for Smart Manufacturing project in the Smart Manufacturing Operations Planning and Control program and the Co-Leader of the Smart Manufacturing Systems Test Bed. Mr. Hedberg is also a Voting Member of the American Society of Mechanical Engineers (ASME) Y14.37, Y14.41, and Y14.41.1 subcommittees from the ASME Y14 suite of standards. Mr. Hedberg has

a strong interest in the areas of model-based enterprise, smart manufacturing, product-lifecycle management (PLM), data/information flow in the lifecycle, product-data quality, product-data visualization, and long-term data archival and retrieval (LOTAR).

Mr. Hedberg is a registered Professional Engineer (PE) in the State of Arizona and the State of Maryland. He is a lifetime senior member of the American Institute of Aeronautics and Astronautics (AIAA) and an active member in the American Society of Mechanical Engineers (ASME), American Association for the Advancement of Science (AAAS), International Council on Systems Engineering (INCOSE), SME, and the National Society of Professional Engineers (NSPE). When not conducting research, Mr. Hedberg enjoys playing ice hockey, general aviation, and cycling.

Conflict-driven Hybrid Observer-based Anomaly Detection

Zheng Wang¹, Farshad Harirchi², Dhananjay Anand³, CheeYee Tang³, James Moyne¹, Dawn Tilbury¹

Abstract— This paper presents an anomaly detection method using a hybrid observer – which consists of a discrete state observer and a continuous state observer. We focus our attention on anomalies caused by intelligent attacks, which may bypass existing anomaly detection methods because neither the event sequence nor the observed residuals appear to be anomalous. Based on the relation between the continuous and discrete variables, we define three conflict types and give the conditions under which the detection of the anomalies is guaranteed. We call this method conflict-driven anomaly detection. The effectiveness of this method is demonstrated mathematically and illustrated on a Train-Gate (TG) system.

I. INTRODUCTION

Cyber-Physical Systems (CPS) are systems that are shaped by a combination of computing devices, communication networks, and physical processes [1]. The integration of these systems into our every-day life is inevitable. The performance and functionality of many critical infrastructures such as power, traffic and health-care networks and smart cities rely on the advances on CPS. A fault or an attack on one of these critical systems, may affect a large portion of society with serious and lethal consequences. As such, the safety and reliability of CPS becomes more and more crucial every day. Fault, attack and anomaly detection mechanisms play a vital role in providing such reliability and safety to CPS. In this paper, we propose an anomaly detection approach that provides formal detection guarantees for an extended class of anomalies in CPS. Similar to [2], we refer to any occurrence that is different from what is standard, normal, or expected as anomaly. In this paper, we utilize the rich dynamical behavior of mixed continuous and discrete (i.e., hybrid) systems [3] as our modeling framework to describe CPS. Even though the design and implementation of anomaly detection methods is significantly more challenging on hybrid models, we leverage these models, because of their advantage in better representing the real-world CPS.

Our motivational example is a Train-Gate (TG) system, consisting of a train and a gate with a road crossing the track, as shown in Fig.1. It is an abstracted example that captures one of the important characteristics of a railway system which is railway level crossing control system. The TG system is a hybrid system. The train with an internal controller for the train speed is the continuous system. An external controller changes the reference train speed based

³ Software and Systems Divison, National Institute of Standards and Technology, Gaithersburg, MD, USA *dhananjay.anand*, *cheeyee.tang@nist.gov* on the measured train position such that the train passes the gate at a lower speed. The gate is a discrete system, which is raised or lowered by a controller using two presence sensors located on both sides of the road. If sensor 1 detects the train, the gate must be lowered down to stop traffic on the road. If sensor 2 detects the train, the gate must be raised up to allow traffic on the road. Two monitors are used to detect anomalies. One monitor detects anomalies in the continuous train system, which uses the continuous system model and compares the measured variables with the estimated ones. The other monitor detects anomalies in the discrete gate system, which uses the discrete system model and compares the expected discrete event sequence with observed one. If an anomaly is detected from either of these monitors, some actions should be taken to mitigate its impact.





However, an attacker can launch an attack to cause an anomaly bypassing both monitors. For example, an anomalous ramp signal could be added to the measured train position without increasing the difference between the measured and the estimated variables. The drifted measured position can make the train pass the gate with a high speed, causing insufficient time to lower the gate. A driver may pass the gate, causing an accident.

In order to detect this type of anomaly, we propose a higher level monitor to augment the previous two monitors. This new monitor uses a hybrid model of the system, and estimates both the continuous and the discrete variables. For the anomaly of a ramp signal injection on the train position, although the continuous system is anomalous, the discrete system is normal. If sensor 1 detects the train but the estimated train position indicates that the train is far away from sensor 1, a "conflict" between the continuous and the discrete variables occurs. This new monitor expands the types of anomalies that can be detected by checking the occurrence of conflicts, called conflict-driven method. Both mathematical demonstrations and simulation results illustrate the effectiveness of the conflict-driven method.

II. BACKGROUND AND CONTRIBUTIONS

Various model-based anomaly detection methods have been developed for both continuous systems and discrete systems [4], [5]. Even though discrete model-based anomaly detection methods are computationally efficient [6], they cannot provide sufficient resolution of continuous degradations

¹ Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA *zhengwa, moyne, tilbury@umich.edu*

² Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA *harirchi@umich.edu*

for hybrid systems [7]. Continuous model-based methods are impractical for the diagnosis of hybrid systems with a large number of discrete states because multiple continuous models need to run in parallel, each model corresponding to one discrete state [8].

As most CPS are hybrid, consisting of both continuous dynamics and discrete behavior, hybrid model-based approaches are promising in anomaly detection. Hybrid model-based anomaly detection includes set membership-based methods [9] and observer-based methods [10]. Given a data trajectory, set membership-based methods check whether the trajectory can be generated by the model. Even though these methods provide necessary and sufficient conditions in some cases for anomaly detection, they are computationally demanding, as they require costly set calculations or mixed integer programming. The set membership-based methods are also utilized in active fault diagnosis, where the goal is to design a minimal excitation that guarantees the detection of anomalous behavior [11], [12], [13].

Observer-based methods assume the continuous component of the hybrid model is observable under both normal and anomalous operations. For most observer-based methods, a residual, which is the difference between the estimated output and the actual output, is analyzed to determine the occurrence of an anomaly. State estimation problem is directly related to observer-based methods. Among various hybrid state estimation methods, a hybrid observer is better for real-time computation since it can reduce the computational complexity [10]. A hybrid observer consists of two components: a discrete state observer identifying what is the current discrete state, and a continuous state observer estimating the continuous state [10], [14]. With the hybrid observer framework, various traditional residual-based anomaly detection methods can be applied for hybrid systems, including different residual generation methods, such as the dedicated and generalized observer scheme [15], [16], and some residual evaluation methods, such as adaptive threshold [17].

A. Contributions

Even though the residual-based methods are efficient, intuitive and easy to implement, they can easily be tricked by a smart attacker or by sensor faults that make the continuous system unobservable, causing anomalies. An example of such class of anomalies is described in Section I. In this paper, we propose a conflict-driven anomaly detection approach with three conflict types defined based on the relation between the discrete and the continuous variables of the hybrid systems and in addition to faults that can be detected by traditional observer-based and residual-based methods, it is capable of providing guarantees on the detection of attacks and faults that are undetectable using these methods.

III. PROBLEM FORMULATION & SOLUTION

In this section, we describe the modeling framework that we consider and the anomaly types that are of interest. Also, a review of utilized hybrid observer is given.

A. Notation

Let $\|\cdot\|$ denote ∞ -norm, $\tilde{\cdot}$ denote estimated variables, \cup denote disjoint union, and $\Box \sigma$ denote the ball of center 0 and of radius σ . In addition, $\mathbf{x} \in \mathbb{R}^n$ represents a vector, where its i^{th} element is indicated by $\mathbf{x}^{(i)}$. $\mathbf{A} \in \mathbb{R}^{n \times m}$ represents a matrix. The linear span of a set of vectors is denoted by $span(\cdot)$. For a set $X \subset \mathbb{R}^n$, we denote its closure, interior, and boundary by \overline{X} , X^o and ∂X respectively. Clearly, $\partial X = \overline{X} \setminus X^o$. The volume of the closed set \overline{X} is denoted by $Vol(\overline{X})$.

B. Modeling Framework

1) Hybrid Model: A hybrid system can be modeled as a hybrid automaton $\mathcal{H} = (\mathcal{X}, \mathcal{U}, \mathcal{Y}, Init, field, E, \phi, \eta)$, where each element is defined as

- $\mathcal{X} = Q \times X$: a set of discrete and continuous states
- $\mathcal{U} = \Psi \times U$: a set of discrete and continuous inputs
- $\mathcal{Y} = \Omega \times Y$: a set of discrete and continuous outputs
- *Init* = $(q(t_0), \mathbf{x}(t_0)) \in \mathcal{X}$: an initial state
- *field* : $\mathcal{X} \times \mathcal{U} \to X$: a time invariant vector field
- $E = \Psi \cup \Omega$: a set of discrete events
- $\phi: Q \times \Psi \to Q$: a set of discrete transitions
- η : X × U → Y: an output map consisting of a discrete output map ζ and a continuous output equation h ζ: Q×Ψ → Ω: a discrete output map
 - $h: \mathbf{y}(t) = \mathbf{C}_{a}\mathbf{x}(t) + \mathbf{v}(t)$: a continuous output equation

The hybrid models considered in this paper capture both nominal system model with a set of nominal discrete states Q_n and anomaly models with a set of anomalous discrete states Q_f . The set of all discrete states is defined as Q = $Q_n \cup Q_f$. The nominal hybrid system \mathcal{H}_n can be derived by removing Q_f and the events and transitions connecting Q_f . The initial state *Init*, which is a combination of initial discrete state $q(t_0) \in Q_n$ and initial continuous state $\mathbf{x}(t_0)$, is not required to be known.

For each discrete state $q \in Q$, we consider continuous dynamics that can be represented by a Linear Time Invariant (LTI) model, subject to process and measurement noise.

$$field: \mathbf{x}(t+1) = \mathbf{A}_q \mathbf{x}(t) + \mathbf{B}_q \mathbf{u}(t) + \mathbf{w}(t),$$

$$h: \mathbf{y}(t) = \mathbf{C}_q \mathbf{x}(t) + \mathbf{v}(t),$$
(1)

where $\mathbf{A}_q \in \mathbb{R}^{n \times n}$, $\mathbf{B}_q \in \mathbb{R}^{n \times n_u}$, $\mathbf{C}_q \in \mathbb{R}^{n_y \times n}$ are system matrices, $\mathbf{x} \in X \subset \mathbb{R}^n$, $\mathbf{u} \in U \subset \mathbb{R}^{n_u}$ and $\mathbf{y} \in Y \subseteq \mathbb{R}^{n_y}$ are continuous states, inputs and outputs, respectively. The process and measurement noise are represented by $\mathbf{w} \sim \mathcal{N}(0, \mathbf{W})$ and $\mathbf{v} \sim \mathcal{N}(0, \mathbf{V})$, respectively, where $\|\mathbf{w}\| \le w$ and $\|\mathbf{v}\| \le v$. Each entry of the process and measurement noise has its bound, i.e., $|\mathbf{w}^{(i)}| \le w_i$ and $|\mathbf{v}^{(i)}| \le v_i$. The continuous dynamical models of the system in anomalous discrete states are not required to be known. To simplify the notation, we assume:

Assumption 1: The output matrix C_q is an identity matrix in all discrete states, i.e., $\forall q \in Q, C_q = I$. We can easily extend our work to general C matrix assuming

the continuous system is observable. Discrete events *E* can be partitioned into observable events E_a and unobservable events E_{ua} , i.e., $E = E_a \cup E_{ua}$. Only

 E_o and unobservable events E_{uo} , i.e., $E = E_o \cup E_{uo}$. Only observable events can be detected by an observer. We denote the set of observable input events as Ψ_o and a set of

unobservable input events as Ψ_{uo} . Obviously, all of the output events are observable.

The *i*th discrete event occurs at time t_i . The continuous evolutions occur in time $t \in [t_{i-1} + 1, t_i], \forall i = 1, 2, ...$ In reality, discrete events may occur between two adjacent sample times. We assume

Assumption 2: The occurrence of the discrete events can be captured at sample times. At most one input event occurs within one sampling period. An output event occurs simultaneously with an input event.

Note that the discrete state is changed one time step after a discrete input event occurs, that is $\phi(q(t_i), \psi) = q'(t_i + 1)$, where $q(t_i), q'(t_i + 1) \in Q$. To each discrete transition $\phi(q, \psi) = q'$, we associate a guard:

$$G(q,q',\boldsymbol{\psi}) = \{\mathbf{x} : s_G \mathbf{x}^{(i_G)} \ge c_G\},\tag{2}$$

where c_G is a constant value and s_G is either -1 or 1. A guard is a closed half-space divided by the hyperplane

$$\mathcal{P}(q,q',\boldsymbol{\psi}) = \{\mathbf{x} : \mathbf{x}^{(i_G)} = s_G c_G\}.$$
(3)

A guard $G(q,q',\psi)$ indicates that the transition ψ will take place if and only if the i_G^{th} state variable of $s_G \mathbf{x}$ is no smaller than c_G in discrete state q.

To each discrete state $q \in Q$, we associate an invariant:

$$Inv_q = \{\mathbf{x} : \forall i = 1, ..., n, \underline{\beta}_i \le \mathbf{x}^{(i)} \le \overline{\beta}_i, \} \subseteq X,$$
(4)

where $\underline{\beta}_i$ and β_i are constant values. An invariant is a hyperrectangle with bounded intervals on each continuous state variable. An invariant Inv_q indicates that the system can remain in the discrete state q if and only if the continuous state $\mathbf{x} \in Inv_q \setminus \bigcup_j G(q, q_j, \psi_j)$.

Our definitions of guard $G(q,q',\psi)$ and invariant Inv_q indicate that c_G is between the lower and upper bounds of the state variable $\mathbf{x}^{(i_G)}$ of the invariant Inv_q , i.e., $\underline{\beta}_{i_G} \le c_G \le \overline{\beta}_{i_G}$. We define a neighbor hyperplane of guard $G(q,q',\psi)$ as

Definition 1: (Neighbor hyperplane of guard $G(q, q', \psi)$) is one of the hyperplanes forming the boundary of the invariant ∂Inv_q , which is defined as follows:

$$\mathcal{L}(q,q',\boldsymbol{\psi}) = \{ \mathbf{x} \in X : |\mathbf{x}^{(i_G)} - \mathbf{x}'^{(i_G)}| = \min(c_G - \underline{\beta}_{i_G}, \beta_{i_G} - c_G) \\ \wedge \mathbf{x}^{(i_G)} \in \{ \underline{\beta}_{i_G}, \overline{\beta}_{i_G} \}, \mathbf{x}' \in \mathcal{P}(q,q',\boldsymbol{\psi}) \}.$$
(5)

An example of neighbor hyperplane of $G(q,q',\psi)$ is shown in Fig.2. To simplify notation, we denote $c_{\mathcal{L}}$ as the value of $\mathbf{x}^{(i_G)}$, where $\mathbf{x} \in \mathcal{L}(q,q',\psi)$. If $\mathcal{P}(q,q',\psi)$ forms one of the hyperplanes of ∂Inv_q , then $\mathcal{L}(q,q',\psi) = \mathcal{P}(q,q',\psi)$. Otherwise, $\mathcal{L}(q,q',\psi) \cap \mathcal{P}(q,q',\psi) = \emptyset$. Discontinuities may exist in continuous variables due to discrete transitions in general hybrid systems. However, in our hybrid system formalism, no discontinuities exist in the continuous variables. This is imposed without any reset maps.

The hybrid observer used in this paper is proposed in [14], which is designed based on the Finite State Machine (FSM) associated with the nominal hybrid model. The FSM M_n is derived by removing all of the continuous dynamics in H_n , and is represented by tuple $(Q, \Psi, \Omega, q(t_0), E, \phi, \zeta)$. In order to get a unique estimate of the discrete state with the hybrid observer after finite observable events, we assume Assumption 3: The FSM M_n is current-state observable. Current-state observable is defined in [14].

Definition 2: (Current-State Observable) A FSM is current-state observable if there exists an integer k such that for any unknown initial discrete state, the discrete state at i can be determined from the observed input/output event pairs sequence up to i, i.e., $i \ge k$.

Note that one input/output event pair is considered as one input event to the hybrid observer. Thus, after the k^{th} input/output event pair occurs, the hybrid observer can give a unique estimated discrete state. The necessary and sufficient condition of current state observability is given in [14].

2) Nominal Discrete States: We partition the invariants of the nominal discrete states into an intermediate region and several normal operating regions. The intermediate region \mathcal{R}_{in} is the union of all the intersections between the invariants of any two nominal discrete states

$$\mathcal{R}_{in} = \{ \mathbf{x} \in X : \forall q, q' \in Q_n, q \neq q' \land \mathbf{x} \in Inv_q \cap Inv_{q'} \}.$$
(6)

For discrete state $q \in Q_n$, we define a normal operating region as the set of continuous states that are in the invariant but not the intermediate region,

$$\mathcal{R}_{no,q} = \overline{Inv_q} \setminus \overline{\mathcal{R}_{in}}.$$
(7)

We pose an assumption on the hybrid model to restrict state-space abstraction method. This assumption helps select the appropriate hybrid model of the system with which the conflict-driven method can provide detection guarantees.

Assumption 4: The intermediate region is bounded by the hyperplane $\mathcal{P}(q,q',\psi)$ corresponding to the guard $G(q,q',\psi)$ and the neighbor hyperplane of $G(q,q',\psi)$ and ∂Inv_q in each discrete state.

$$\mathcal{R}_{in} \subset \bigcup_{q \in \mathcal{Q}_n} \{ \mathbf{x} \in Inv_q : \forall q_j \in \mathcal{Q}_n : \exists \mathcal{P}(q, q_j, \psi_j), \\ \min(c_G, c_{\mathcal{L}}) \leq \mathbf{x}^{(i_G)} \leq \max(c_G, c_{\mathcal{L}}) \}.$$
(8)

The visualization of this assumption on a 2-dimensional system is shown in Fig.2. Assumption 4 indicates $\mathcal{P}(q,q',\psi)$ is one of the hyperplanes forming $\partial Inv_{q'}$.



Fig. 2: Normal operating and intermediate regions

The basic principle of the conflict-driven method is to check at each time step, whether or not the sets of continuous states, which are calculated based on the estimated continuous state, intersect with the invariant of the estimated discrete state. The sets of continuous states include an initial set with the estimated continuous state as the center, and a forward reachable set which is the set of all continuous states that can be reached along trajectories starting in the initial set. Reachable set calculation requires following assumptions.

Assumption 5: The continuous system in each nominal discrete state is open-loop stable or marginally stable, i.e., $|\lambda(\mathbf{A}_q)| \leq 1$, where $\lambda(\mathbf{A}_q)$ are the eigenvalues of \mathbf{A}_q .

Assumption 6: The continuous input signal is bounded, and the bound is known, i.e., $||\mathbf{u}|| \le \mu$.

A great deal of attention has been given to algorithms and software developed for analysis of hybrid systems. To date, the most efficient way to compute the reachable set is to use zonotopes [18]. A zonotope is a Minkowski sum of a finite set of line segments, defined as

Definition 3: (Zonotope Z) is a set such that:

$$Z = (\mathbf{x}_{\mathbf{c}}, < \mathbf{g}_{1}, ..., \mathbf{g}_{p} >)$$

= { $\mathbf{x} \in \mathbb{R}^{n} : \mathbf{x} = \mathbf{x}_{\mathbf{c}} + \sum_{i=1}^{i=p} b_{i} \mathbf{g}_{i}, -1 \le b_{i} \le 1$ }, $p \ge n$, (9)

where $\mathbf{x}_{\mathbf{c}}, \mathbf{g}_{\mathbf{i}} \in \mathbb{R}^{n}$ are the center and generators, respectively. Both *p* and *n* determine the maximum number of vertices and facets.

C. Hybrid Observer

Given the nominal hybrid model \mathcal{H}_n , we can design a hybrid observer to estimate both the discrete state and the continuous state of the system using the method in [14]. The hybrid observer \mathcal{O} consists of a discrete state observer \mathcal{D} and a continuous state observer \mathcal{C} , as shown in Fig.3. The discrete state observer receives discrete input/output event pair (ψ, ω) and gives \tilde{q} . The estimated discrete state \tilde{q} contains a set of estimated discrete states before the occurrence of the k^{th} observable input/output event pair. After the occurrence of the k^{th} observable input/output event pair, \tilde{q} , which contains a unique estimate, is passed to the corresponding continuous state observer. Then the continuous state observer gives an estimated continuous state $\tilde{\mathbf{x}}$ using the continuous input \mathbf{u} and output \mathbf{y} .

The discrete state observer is represented by a FSM which is a tuple $\mathcal{D} = (\tilde{Q}, E_{\mathcal{D}}, -, Q_n, E_{\mathcal{D}}, \tilde{\phi}, -)$, where $E_{\mathcal{D}} = (\Psi, \Omega)$ is the set of discrete input/output event pairs of \mathcal{M}_n . The discrete state observer is tracking the set of possible discrete states that the system can be in. Therefore, no discrete output events or discrete map are defined for discrete state observer.

The construction of \mathcal{D} starts from $\tilde{q}(t_0)$: with unknown initial discrete state of \mathcal{M}_n , $\tilde{q}(t_0) = Q_n$. Then for each discrete state $\tilde{q} \in \tilde{Q}$, we identify the input/output event pairs (ψ, ω) , that label all the transitions out of any state q' in \tilde{q} . These events are called active event set of \tilde{q} . For each pair (ψ, ω) in the active event set, we identify $q \in Q_n$ that can be reached from $q' \in \tilde{q}$, and these states return as a new \tilde{q} in \tilde{Q} . This transition is added to $\tilde{\phi}$ satisfying

$$\tilde{\phi} \coloneqq \{ q \in Q_n : \exists q' \in \tilde{q}, q \in \phi(q', \psi) \land \omega = \zeta(q', \psi) \}.$$
(10)

Repeat this step until no new \tilde{q} and $\tilde{\phi}$ can be added to \mathcal{D} .

To reduce the effect of system noise on state estimation, we use a Kalman filter as the continuous state observer, with Kalman gain $\mathbf{K}_{q(t)}$. It is well known that the Kalman gain will converge in a few steps in practice if the system is observable [19]. We can use the steady state Kalman gain given in [19], with which the eigenvalues of $(\mathbf{A}_q - \mathbf{K}_q \mathbf{A}_q)$ are stable. Note that we have different Kalman gains for different discrete states. Let us define

Definition 4: (Dwell time Δt) is the minimum time to guarantee the convergence of the estimation error.

Dwell time Δt should satisfy the condition in section 3.2 in [14], Then we assume:

Assumption 7: The time gap between any two consecutive transitions is greater than dwell time, i.e., $t_i - (t_{i-1} + 1) > \Delta t$. With bounded noise, we design Kalman filter such that the estimation error $\mathbf{x}_{\mathbf{e}}(t) = \mathbf{x}(t) - \tilde{\mathbf{x}}(t)$ is bounded when the Kalman filter reaches its steady state, i.e., $\exists t_{ss}, \|\mathbf{x}_{\mathbf{e}}(t)\| \le \theta, t > t_{ss}$. The residual \mathbf{r} of the system is defined as the difference between the measure output and the estimated output,

$$\mathbf{r}(t) = \mathbf{y}(t) - \tilde{\mathbf{x}}(t). \tag{11}$$

In the nominal discrete states, the residual $\mathbf{r}(t)$, $t > t_{ss}$ is bounded by $\theta + v$ because of bounded estimation error and noise. If $\|\mathbf{r}(t)\| > \theta + v$, $t > t_{ss}$, then the system is in an anomalous discrete state.

D. Anomalous Discrete States

An anomaly $f \in \Psi_{uo}$ is defined as an unobservable input event that transits the system from a nominal discrete state $q_n \in Q_n$ to an anomalous discrete state $q_f \in Q_f$. Arguably, the multiplicative anomalies can be represented by additive anomaly models (e.g., Section 3.5 in [4]). Thus, we restrict our attention to additive anomaly models as follows.

$$\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{v}(t) + \mathbf{\Gamma} \boldsymbol{\gamma}(t), \qquad (12)$$

where $\mathbf{\Gamma} \in \mathbb{R}^{n \times n}$ is a diagonal matrix with binary variables. The *i*th diagonal variable is 1 if and only if the *i*th output is added with an anomalous signal $\boldsymbol{\gamma}(t) \in \mathbb{R}^n$. Then the residual in anomalous discrete states is changed to

$$\mathbf{r}(t) = \mathbf{x}_{\mathbf{e}}(t) + \mathbf{v}(t) + \mathbf{\Gamma} \boldsymbol{\gamma}(t).$$
(13)

The conflict-driven method is guaranteed to detect the anomalies that are not consistent with the continuous dynamics of the system, i.e, the anomalies that make the residual greater than threshold $\theta + v$. This is because of leveraging continuous state observer that is described in the previous subsection. Additionally, the proposed method extends the types of anomalies that can be detected compared to the methods mentioned in Section II.

Perfectly attackable systems are defined by Mo, et al. in [19] as continuous systems for which anomalies caused by certain attacks can remain undetected, i.e., the residual will not increase. One of the conditions for a continuous system to be perfectly attackable is that the state matrix \mathbf{A}_{q_f} has at least one unstable or marginally stable eigenvalue. If the continuous system only has stable eigenvalues, anomalies on the system will increase the residual. The smart attacks that cannot be detected in perfectly attackable systems are called False Data Injection Attack (FDIA) as defined and demonstrated in [19]. One of the conditions of FDIA is that the eigenvector $\boldsymbol{\xi}$ corresponding to an unstable or marginally stable eigenvalue of \mathbf{A}_{q_f} is in the span of Γ , i.e., $\boldsymbol{\xi} \in span(\Gamma)$. If $\boldsymbol{\xi} \notin span(\boldsymbol{\Gamma})$, the anomaly will increase the residual and will be detected by the Kalman filter implemented as the continuous state observer in conflict-driven method.

As mentioned before, detecting FDIA type anomalies is challenging, as their effect cannot be observed in the value of residual. In addition to anomalies that can be detected by checking the residual, our main contribution is to also guarantee the detection of this particular type of anomalies, if they satisfy certain conditions (explained in Section IV). Let us define Type- C_u anomalies for the hybrid systems as:

Definition 5: (Type- C_u anomaly) is an anomaly that is caused by False Data Injection Attack. If an anomaly occurs at time t_f , it satisfies the following two conditions.

1) The input-output sequence generated from the anomalous discrete state satisfies the continuous dynamics of the nominal discrete states for $t \ge t_f$, that is, the residual does not grow larger than the threshold $\theta + v$.

2) The occurrence of the anomaly results in:

for $t \ge t_f$, if $q \in Q_f \implies ||\mathbf{x}_{\mathbf{e}}(t)|| > \theta$.

Our objective is to extend the detection guarantees to the class of Type- C_u anomalies. In order to establish the goal, we also assume that:

Assumption 8: An anomaly occurs after the continuous state observer enters its steady state, i.e., $t_f \ge t_{ss}$.

IV. CONFLICT-DRIVEN ANOMALY DETECTION METHOD

In the conflict-driven method, we define three conflict types. This method checks the occurrence of the conflicts to detect anomalies. The work flow diagram is shown in Fig.3. Note that this method is used after the hybrid observer is in the steady state, i.e., $t \ge t_{ss}$.



Fig. 3: Conflict-driven anomaly detection method

The conflict-driven method has three steps:

1) Calculate an initial set $X_I(t)$: An initial set $X_I(t)$ is constructed as a zonotope based on $\tilde{\mathbf{x}}(t)$ and $\mathbf{r}(t)$, as $X_I(t) =$ $(\tilde{\mathbf{x}}(t), \langle \mathbf{g_1}, ..., \mathbf{g_n} \rangle)$. The *i*th generator $\mathbf{g_i}^{(i)} = |\mathbf{r}^{(i)}(t)| + v_i$. Other entries of vector \mathbf{g}_{i} are zero. Based on Equation (11), we have $|\mathbf{x}_{\mathbf{e}}^{(i)}(t)| \leq |\mathbf{r}^{(i)}(t)| + v_i$ in nominal discrete states. Thus, we can ensure $\mathbf{x}(t) \in X_{I}(t)$ when the system is in nominal discrete states. The initial set is changing at each time step because of the changes in the estimated continuous state and the residual.

2) Calculate the reachable set $R_{\delta_{\tilde{q}(t)}}(X_I(t))$: The $\delta_{\tilde{q}(t)}$ time-step forward reachable set $R_{\delta_{\tilde{q}(t)}}(X_I(t))$ is calculated starting from $X_I(t)$ constructed in Step 1. It satisfies

$$R_{\delta_{\tilde{q}(t)}}(X_{I}(t)) \subseteq \mathbf{A}_{\tilde{q}(t)}^{o_{\tilde{q}(t)}} X_{I}(t) + \Box \sigma_{\tilde{q}(t)}$$
(14)

where $\sigma_{\bar{q}(t)} = \frac{1 - \|\mathbf{A}_{\bar{q}(t)}\|^{\delta_{\bar{q}(t)}}}{1 - \|\mathbf{A}_{\bar{q}(t)}\|} (\|\mathbf{B}_{\bar{q}(t)}\| \|\mu + w)$. For more details about reachable set calculation using zonotopes, refer to [18].

3) Check conflicts: We define three conflict types in this paper, as shown in Fig.4:

Conflict A. The volume of the initial set is larger than the bound, i.e., $Vol(X_I(t)) > \prod_{i=1}^n (2\theta + 4v_i)$

Conflict B. The initial set has no intersection with the invariant of the estimated discrete state $(X_I(t) \cap Inv_{\tilde{q}(t)} = \emptyset)$

Conflict C. The $\delta_{\tilde{q}(t)}$ time steps forward reachable set has no intersection with the invariant of the estimated discrete state, i.e., $R_{\delta_{\tilde{q}(t)}}(X_I(t)) \cap Inv_{\tilde{q}(t)} = \emptyset$. If one of these conflicts occurs, the system is in an

anomalous state.



Note that for Step 2, we do not consider the discrete behavior in reachability analysis. The reachable set could be completely outside the invariant if $\delta_{\tilde{q}(t)}$ is too large, causing false alarms. To avoid false alarms and provide detection guarantees, we determine the time steps δ_q for each discrete state with given nominal hybrid model \mathcal{H}_n of the system according to the following steps:

1) In Inv_q , starting from the intersection of the hyperplane corresponding to the *i*th guard $G(q,q_i,\psi_i)$ as defined by (3) and Inv_q , we find the minimum time steps $\delta_{q,i}$ which satisfies

$$R_{\delta_{q,i}+1}(\mathcal{P}(q,q_i,\psi_i) \cap Inv_q) \cap \mathcal{L}(q,q_i,\psi_i) \neq \emptyset$$
(15)

Note that $\delta_{q,i}$ may be different for different guards in the same discrete state. The reason we use $\delta_{q,i} + 1$ is that the continuous system is a discrete-time model and we want to ensure the $\delta_{q,i}$ time-step forward reachable set, starting from any possible real continuous state when a transition occurs, has intersection with Inv_q in nominal discrete states.

2) Let $\delta_q = \min_i(\delta_{q,i})$. If the distance between $\mathcal{P}(q, q_i, \psi_i)$ and $\mathcal{L}(q, q_i, \psi_i)$ is small, δ_q may be 0. Then we only need to check Conflicts A and B in discrete state q.

Following proposition and theorem demonstrate the effectiveness of the conflict-driven method. We give some intuitions first. Proposition 1 gives the upper bound for the volume of the initial set. Based on Assumption 1, in a nominal discrete state, the estimation error of the continuous state, as well as the residual, should converge. Therefore, an upper bound exists for the volume of the initial set $Vol(X_I(t)), t > t_{ss}$, as demonstrated in Proposition 1. The increase of $Vol(X_I(t))$ indicates the increase of the residual. Conflict A can detect anomalies that increase the residual. Since the main contribution of this paper is focusing on the

Anand, Dhananjay; Moyne, James; Tang, CheeYee; Tilbury, Dawn; Wang, Zheng. "Conflict-driven Hybrid Observer-based Anomaly Detection." Paper presented at American Control Conference, Milwaukee, WI, United States. July 2, 2018 - July 6, 2018.

detection of Type- C_u anomaly which does not increase the residual, finding the lower bound of the anomalous signal which causes conflict A and the conditions under which a residual-based method is equivalent to checking Conflict A are part of our future work.

Proposition 1: Given a nominal hybrid automaton \mathcal{H}_n and a hybrid observer \mathcal{O} with bounded estimation error in steady state, i.e., $\forall t > t_{ss}$, $\|\mathbf{x}_{\mathbf{e}}(t)\| \le \theta$, the volume of the initial set is also bounded, i.e., $Vol(X_I(t)) \le \prod_{i=1}^n (2\theta + 4v_i)$.

Proof: In steady state, $\forall t > t_{ss}$,

$$Vol(X_{I}(t)) = \prod_{i=1}^{n} (2(|\mathbf{r}^{(i)}(t)| + v_{i})) \le \prod_{i=1}^{n} (2(||\mathbf{x}_{e}(t)|| + 2v_{i})) \le \prod_{i=1}^{n} (2||\mathbf{x}_{e}(t)|| + 4v_{i}) \le \prod_{i=1}^{n} (2\theta + 4v_{i})$$

As discussed before, Type- C_u anomaly affects the continuous outputs of the system, but can remain undetectable by residual-based methods and unobserved by discrete state observer. In order to detect this type of anomaly, we leverage the estimated states from both continuous and discrete observers, and take advantage of observation of a discrete event. This enables us to employ the contradictions among estimated continuous and discrete states and the model parameters such as guards and invariants to detect these challenging anomalies. These contradictions are formalized in Conflicts B and C. In what follows, we set the stage to present the main contribution of this paper, which is Theorem 1. This theorem provides sufficient conditions on the lower bound of the anomalous signal, under which the conflictdriven method is guaranteed to detect Type- C_{μ} anomalies. Towards this goal, we first find the lower bound of the estimation error that creates one of Conflicts B or C, and then relate this bound to the lower bound on the anomalous signal according to (13).

Let us assume that a Type- C_u anomaly occurs at time t_f which causes a large estimation error on the i_G^{th} state variable, i.e., $|\mathbf{x}_e^{(i_G)}| > \theta$, and a discrete event ψ occurs at time t_e which associates a guard with condition on the i_G^{th} state variable, i.e., $\{\mathbf{x} \in Inv_q : s_G \mathbf{x}^{(i_G)} \ge c_G\}$. Without loss of generality, we assume that the projection of $\mathcal{R}_{no,q}^{o}$ onto $\mathbf{x}_{e}^{(i_{G})}$ is bounded above by c_{G} , i.e., $\mathbf{H}_{i_{G}}\mathcal{R}_{no,q}^{o} \leq c_{G}$ (because $s_{G} = 1$), where $\mathbf{H}_{i_{G}} \in$ \mathbb{R}^n is the projection row vector with the i_G^{th} entry "1" and "0" elsewhere. The procedure for the case where $\mathbf{H}_{i_G} \mathcal{R}_{no,q}^o \ge -c_G$ (because $s_G = -1$) is identical. When this event occurs, we can only have two possibilities for the estimated state at time t_e , either $\tilde{\mathbf{x}}(t_e) \in \mathcal{R}_{no,q}^o$, or $\tilde{\mathbf{x}}(t_e) \in \mathcal{R}_{in}^o \cap Inv_q$. Based on our definitions of guard, invariant, neighbor hyperplane of the guard, and Assumption 4, along the i_G^{th} state variable, the upper bound of Inv_q is $c_{\mathcal{L}}$ and the lower bound of $Inv_{q'}$ is c_G . For brevity in notation and as in this section we mainly consider $G(q,q',\psi)$, we refer to it as G.

First, consider the case where $\tilde{\mathbf{x}}(t_e) \in \mathcal{R}_{no,q}^o$, that is, when the real continuous state satisfies the guard, the estimated state is in the normal operating region of discrete state q. The goal is to find the lower bound of the estimation error along the i_G^{th} state variable, such that:

• The initial set $X_I(t_e+1)$ has no intersection with $Inv_{q'}$. We denote such minimum estimation error corresponding to *G* by z_G^* . To find z_G^* , it suffices to find the minimum *z* such that for all $\tilde{\mathbf{x}}(t_e+1)$ the upper bound of $X_I(t_e+1)$ is smaller than the lower bound of $Inv_{q'}$ along the t_G^{th} state variable,

$$\mathbf{H}_{i_G} \tilde{\mathbf{x}} (t_e + 1) + \boldsymbol{\theta} + 2v < c_G.$$
⁽¹⁶⁾

Note that at time t_e , the continuous state of the system along the i_G^{th} state variable is greater than or equal to c_G , and smaller than the maximum value of the one time step forward reachable set from $\mathcal{P}(q,q',\psi) \cap Inv_q$ along the i_G^{th} state variable, i.e., $c_G \leq \mathbf{H}_{i_G} \mathbf{x}(t_e) < \varepsilon$, where $\varepsilon = \max(\mathbf{H}_{i_G} R_1(\mathcal{P}(q,q',\psi) \cap Inv_q))$. After the occurrence of event ψ , the state equation of the anomalous discrete state is changed to $(\mathbf{A}_{q'}, \mathbf{B}_{q'})$ and the estimated discrete state is changed to q' at time $t_e + 1$. Then the set of all possible continuous states at time $t_e + 1$ can be represented by:

$$\forall \mathbf{x}(t_e) \in Inv_q, c_G \leq \mathbf{H}_{i_G} \mathbf{x}(t_e) < \varepsilon,$$

$$\mathbf{x}(t_e+1) \in R_1(\mathbf{x}(t_e)) \subseteq \mathbf{A}_{q'} \mathbf{x}(t_e) + \Box \sigma_{q'},$$
(17)

where $\sigma_{q'} = \|\mathbf{B}_{q'}\| \mu + w$.

Now, we can pose the problem of finding z_G^* as a robust optimization problem.

$$z_{G}^{*} = \min_{z} \quad z$$

s. t. $z \ge 0, z \ge \mathbf{H}_{i_{G}} \mathbf{A}_{q'} \mathbf{x} + \boldsymbol{\sigma}_{q'} + \boldsymbol{\theta} + 2v - c_{G} \quad (18)$
 $\forall \mathbf{x} \in Inv_{q'}, c_{G} \le \mathbf{H}_{i_{G}} \mathbf{x} \le \varepsilon,$

By utilizing methods from robust optimization literature, e.g., [20], we can convert (18) to a linear programming problem as follows:

$$z_{G}^{*} = \min_{\mathbf{J}, z} \quad z$$

s. t.
$$\begin{bmatrix} 1\\1 \end{bmatrix} z - \begin{bmatrix} \mathbf{J}^{\mathsf{T}} \boldsymbol{\rho}_{\mathbf{I}} \\ 0 \end{bmatrix} \ge \begin{bmatrix} \boldsymbol{\sigma}_{q'} + \boldsymbol{\theta} + 2v - c_{G} \\ 0 \end{bmatrix} \quad (19)$$
$$\boldsymbol{\Lambda}^{\mathsf{T}} \mathbf{J} \ge (\mathbf{H}_{i_{G}} \mathbf{A}_{q'})^{\mathsf{T}}, \mathbf{J} \ge \mathbf{0}$$

where $\mathbf{0} \in \mathbb{R}^{2n \times 1}$ is a zero vector. **x** is in a polytopic uncertain set, i.e., $\mathbf{A}\mathbf{x} \leq \boldsymbol{\rho}_{\mathbf{1}}$ for problem (18), where $\mathbf{A} \in \mathbb{R}^{2n \times n}$, $\boldsymbol{\rho}_{\mathbf{1}} \in \mathbb{R}^{2n \times 1}$ and $\mathbf{J} \in \mathbb{R}^{2n \times 1}$ is a variable of the optimization problem.

For the second possibility, i.e., $\tilde{\mathbf{x}}(t_e) \in \mathcal{R}_{in}^o \cap Inv_q$, we are seeking the lower bound of the estimation error along the i_G^{th} state variable such that it satisfies the following:

• The reachable set for δ_q time steps from any point within the initial set $X_I(t_e)$ of the estimated continuous state has no intersection with Inv_q .

Considering the worst case that the continuous state is the furthest to the upper bound of ∂Inv_q along the i_G^{th} state variable, i.e., $\mathbf{H}_{i_G}\mathbf{x}(t_e) = c_G$, our objective can be equivalently changed to find the minimum distance between c_G and $\mathbf{H}_{i_G}\tilde{\mathbf{x}}(t_e)$. We denote this minimum distance by d_G^* . Define $d = |\mathbf{H}_{i_G}\tilde{\mathbf{x}}(t_e) - c_G|$ as the distance between $\mathcal{P}(q,q',\psi)$ and the estimated state along the i_G^{th} state variable. With this definition, the initial set at time t_e can be represented as $X_I(t_e) = \{\mathbf{x} : \mathbf{H}_{i_G}\mathbf{x} \in [c_G + d - \theta - 2v, c_G + d + \theta + 2v]\}$. Starting from this initial set $X_I(t_e)$, the projection of the reachable set for δ_q time steps forward onto the i_G^{th} state variable becomes $\mathbf{H}_{i_G}\mathbf{A}_q^{\delta_q}\mathbf{x} \pm \sigma_q$, $\forall \mathbf{x} \in X_I(t_e)$, where $\sigma_q = \frac{1-||\mathbf{A}_q||^{\delta_q}}{1-||\mathbf{A}_q||}(||\mathbf{B}_q||\mu+w)$. If $\mathbf{H}_{i_G}\mathbf{A}_q^{\delta_q}\mathbf{x} - \sigma_q > c_{\mathcal{L}}$, $\forall \mathbf{x} \in X_I(t_e)$, then it is guaranteed that the δ_q time-step forward reachable set starting from this

initial set $X_I(t_e)$ has no intersection with the invariant Inv_q . We can pose the problem of finding d_G^* as the following robust optimization problem.

$$d_{G}^{*} = \min_{d} \quad d$$

s. t. $d \ge 0, \mathbf{H}_{i_{G}} \mathbf{A}_{q}^{\delta_{q}} \mathbf{x} - \boldsymbol{\sigma}_{q} \ge c_{\mathcal{L}} \qquad (20)$
 $\forall \mathbf{x} \in Inv_{q}, \mathbf{x} \in X_{I}(t_{e}).$

With a change of variables and by employing the robust optimization techniques [20], we can write an equivalent problem to (20) as a linear program.

$$d_{G}^{*} = \min_{\mathbf{D}, \mathbf{J}} \quad \mathbf{H}_{i_{G}} \mathbf{D}$$

s. t.
$$\begin{bmatrix} \mathbf{H}_{i_{G}} \mathbf{A}_{q}^{\delta_{q}} \\ \mathbf{H}_{i_{G}} \end{bmatrix} \mathbf{D} - \begin{bmatrix} \mathbf{J}^{\mathsf{T}} \boldsymbol{\rho}_{2} \\ 0 \end{bmatrix} \ge \begin{bmatrix} \sigma_{q} + c_{\mathcal{L}} \\ 0 \end{bmatrix} \qquad (21)$$
$$\mathbf{\Lambda}^{\mathsf{T}} \mathbf{J} \ge - (\mathbf{H}_{i_{G}} \mathbf{A}_{q}^{\delta_{q}})^{\mathsf{T}}, \mathbf{J} \ge \mathbf{0}, \mathbf{D} \ge \mathbf{0}$$

where **0** is a zero vector with proper dimension, and $\mathbf{D} \in \mathbb{R}^n$ is a vector with the i_G^{th} entry *d* and other entries "0". **x** is in a polytopic uncertain set, i.e., $\mathbf{A}\mathbf{x} \leq \boldsymbol{\rho}_2$, where $\boldsymbol{\rho}_2 \in \mathbb{R}^{2n \times 1}$ and $\mathbf{J} \in \mathbb{R}^{2n \times 1}$ is the dual variable.

Now that we have introduced z_G^* and d_G^* , we can present the main result of the paper.

Theorem 1: Given a nominal hybrid automaton \mathcal{H}_n . Assume a Type- C_u anomaly f occurs at time t_f . If an event $\psi \in \Psi_o$ occurs at $t_e > t_f$, which is supposed to transit the system from discrete state q to q', and the guard $G(q, q', \psi)$ is a condition on the real continuous state which is affected by the anomaly f, i.e., $G(q, q', \psi) : s_G \mathbf{x}^{(i_G)} \ge c_G$ and $|\mathbf{x}_{\mathbf{e}}^{(i_G)}| \ge \theta$, then the conflict-driven method is guaranteed to detect the anomaly, if the anomaly satisfies:

$$\|\mathbf{\Gamma}\boldsymbol{\gamma}(t)\| > \max(z_a^* + \theta + 2\nu, d_a^* + \theta + 2\nu), \tag{22}$$

where $z_q^* = \max_{q'} z_G^*$ and $d_q^* = \max_{q'} d_G^*$ can be derived by solving the robust optimization problems (18) and (20), respectively for all possible q'.

Proof: The solution z_G^* is the lower bound of the estimation error which ensures $X_I(t_e + 1) \cap Inv_{q'} = \emptyset$, i.e. Conflict B. The values of z_G^* varies from one guard to another. Therefore, by considering z_a^* , we guarantee that at the discrete state q, regardless of guard, Conflict B occurs, if $\|\mathbf{\Gamma}\boldsymbol{\gamma}(t)\| > z_q^* + \theta + 2v$. On the other hand, the solution d_G^* is the lower bound of the estimation error, which ensures $R_{\delta_a}(X_I(t_e)) \cap Inv_q = \emptyset$, i.e., Conflict C. The values of d_G^* varies for different guards, hence, we similarly take the maximum of these values for all possible q', which is d_a^* . Based on the relation between the estimation error and anomalous signal in (13), it is guaranteed that if $\|\Gamma \gamma(t)\| > d_a^* + \theta + 2v$, regardless of guard, Conflict C occurs. By combining the two conditions obtained on the magnitude of anomalous signal for the two possibilities, we can conclude that the proposed conflict-driven method provides detection guarantees on the detection of anomalous signals that satisfy condition (22), regardless of where the estimated state is located in the Inv_q at the time of event. This concludes the proof.

V. SIMULATION RESULT

In this section, we revisit the TG system. We present the nominal hybrid model of the TG system and compare the

1. Run, up (A, B) $lnv_1 = \{x \in X : 0 \le x_p \le 46\}$	$\begin{array}{c} 2. \operatorname{Run}, \operatorname{down}(\boldsymbol{A}, \boldsymbol{B}) \\ \hline \\ c_{up}/s_2 \end{array} \xrightarrow{2. \operatorname{Run}, \operatorname{down}(\boldsymbol{A}, \boldsymbol{B})} Inv_2 = \{\boldsymbol{x} \in X: 45 \leq x_p \leq 76 \\ \land 0 \leq x_v \leq 0.4\} \end{array}$
3. Run, up (A , B) $Inv_3 = \{x \in X: 75 \le x_n \le 80\}$	

Guard: $G(1,2, c_{down})$: $x_p \ge 45$, $G(2,3, c_{up})$: $x_p \ge 75$ Fig. 5: Hybrid automaton \mathcal{H}_n of the TG system.

conflict-driven method with a residual-based method under a Type- C_u anomaly.

The graphic representation of the nominal hybrid model \mathcal{H}_n of the TG system is shown in Fig.5. The train automaton has one discrete state "run". The gate automaton has two discrete states: "up" and "down" (The time of raising and lowering the gate is ignored for simplicity). Although the automata product results in two discrete states, we additionally partition discrete state "run, up" to two discrete states to ensure hyperrectangle invariants as defined in (4). The discrete transitions between discrete states are determined by discrete input events c_{up} and c_{down} , where c_{up} means "raise the gate" and c_{down} means "lower the gate". When sensor 1 detects the train and emits discrete output event s_1 , the gate controller sends out c_{down} . When sensor 2 detects the train and emits discrete output event s_2 , the gate controller sends out c_{up} . For each transition, we associate a guard. The invariants of the discrete states and the guards are given in Fig.5. The continuous state of the TG system is $\mathbf{x} = \begin{bmatrix} x_p & x_v \end{bmatrix}^{\mathsf{T}}$, where x_p, x_v are the train position and the train speed, respectively. The continuous output of the TG system is $\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{v}(t)$. If the train is within 16*m* of the gate, the reference speed is 0.2m/s. Otherwise, it is 1m/s. The desired operation is that the train speed is no faster than 0.4m/s when the train is within 12m of the gate. The TG system is current state observable. Based on Assumption 8, we will only focus on the observer's steady state 1,2 .

The intersections of the invariants give the intermediate region \mathcal{R}_{in} as $\mathcal{R}_{in} = \{ \forall \mathbf{x} \in X : (45 \le x_p \le 46 \lor 75 \le x_p \le 76) \land 0 \le x_v \le 0.4 \}$. Then we can determine the normal operating regions of the three discrete states, as shown in Fig.6,

$$\mathcal{R}_{no,1} = \{ \forall \mathbf{x} \in X : 0 \le x_p < 45 \lor (45 \le x_p \le 46 \land x_v > 0.4) \}, \\ \mathcal{R}_{no,2} = \{ \forall \mathbf{x} \in X : 46 < x_p < 75 \land 0 \le x_v \le 0.4 \}, \\ \mathcal{R}_{no,3} = \{ \forall \mathbf{x} \in X : 76 < x_p \le 80 \lor (75 \le x_p \le 76 \land x_v > 0.4) \}.$$
(23)

The neighbor hyperplane of each guard is then:

$$\mathcal{L}(1,2,c_{down}): x_p = 46, \quad \mathcal{L}(2,1,c_{up}): x_p = 76,$$
 (24)

With the invariants, guards and neighbor hyperplanes, we can determine the time step for reachability analysis of each discrete state, which is $\delta_1 = \delta_2 = 9, \delta_3 = 0.$

¹More parameters: track length: 80*m*; gate, sensor 1, sensor 2 locations: 60*m*, 45*m*, 75*m*; sampling period: 0.1*s*; Upper bounds of noise: v = 0.1, w = 0.01 (Units depend on the state variable with larger noise); estimation error upper bound in the observer's steady state in nominal discrete states: $\theta = 0.05$ (The unit depends on the state variable with larger estimation error at sample times).

 2 In reality, the train track intersects with multiple roads at different locations. The discrete state observer gives a unique estimated discrete state after passing the first road. We only focus on the track segment when the observer is in steady state.



Fig. 6: The normal operating regions and the intermediate region of the TG system.

In these three discrete states, state matrices (\mathbf{A}, \mathbf{B}) are the same. The eigenvalues of \mathbf{A} are 1 and 0.95. The eigenvector $\boldsymbol{\xi}$ corresponding to the marginally stable eigenvalue is $\begin{bmatrix} 1 & 0 \end{bmatrix}^{\mathsf{T}}$. The non-zero element of $\boldsymbol{\xi}$ corresponds to the measured train position. A Type- C_u anomaly scenario is a ramp anomalous signal with slope 0.02m/s added to the measured train position. The anomaly starts at 0*s* and runs until the end of the simulation, which makes the system violate its desired operation at 180.8*s* with position 71.55*m* and speed 0.41*m/s*.

The comparison of the detection performance of the conflict-driven method and a residual-based method under the anomaly mentioned above is shown in Fig.7. The threshold of the residual-based method is $\theta + v = 0.15$ (The unit depends on θ). The residual-based method fails to detect the anomaly because the residual does not increase. The conflict-driven method detects this anomaly at time 48.2*s* when Conflict *C* occurs. The estimated discrete state is 1, but the reachable set $R_{\delta_1}(X_I(482)) \cap Inv_1 = \emptyset$. At 48.2*s*, the norm of the anomalous signal is 0.96*m*, which is lower than the lower bound 0.98*m* calculated by solving robust optimization problems. That means the conflict-driven method may detect the anomalies with norm lower than the lower bound which we can provide detection guarantees.



Fig. 7: Simulation result under the Type- C_u anomaly: (Top) Residual; (Bottom) The occurrence of Conflict *C*.

VI. CONCLUSION AND FUTURE WORK

In this paper, we propose a conflict-driven method, which uses the discrete and continuous variables and the hybrid model of the system, to provide detection guarantees for anomalies that are undetectable with traditional residualbased methods in addition to anomalies that can be detected with these methods. We define three different conflict types. If any one of the conflicts occurs, the anomaly is detected. Both mathematical demonstration and simulation result illustrate the effectiveness of the conflict-driven method.

More work needs to be done about the conflict-driven method. One future work is to improve the hybrid observer design such that we can apply the conflict driven method to more general hybrid systems with reset maps. One potential solution is to use the Convergence Ratio method in [16], which calculates the estimation error of the continuous state with two continuous state observers. Other future work includes the analysis of the conflict-driven method in detecting anomalies that affect both the continuous and discrete variables of the system.

ACKNOWLEDGMENT

This work was supported by the National Institute of Standards and Technology under Award No.70NANB16H205. We thank Isaac Spiegel, a Ph.D. student at University of Michigan for his discussion on hybrid system definition.

REFERENCES

- A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *CPSSW*, 2009, p. 5.
- [2] F. Lopez, M. Saez, Y. Shao, E. Balta, J. Moyne, M. Mao, K. Barton, and D. Tilbury, "Categorization of anomalies in smart manufacturing systems to support the selection of detection mechanisms," in *IEEE CASE*, 2017.
- [3] K. Wan, D. Hughes, K. L. Man, and T. Krilavičius, "Composition challenges and approaches for cyber physical systems," in *IEEE NESEA*, 2010, pp. 1–7.
- [4] S. Ding, Model-based fault diagnosis techniques: design schemes, algorithms, and tools. Springer Science & Business Media, 2008.
- [5] M. Sayed-Mouchaweh, Discrete Event Systems: Diagnosis and Diagnosability. Springer Science & Business Media, 2014.
- [6] M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, and D. C. Teneketzis, "Failure diagnosis using discrete-event models," *IEEE Trans. on control systems technology*, vol. 4, no. 2, pp. 105–124, 1996.
- [7] C. H. Goodrich and J. Kurien, "Continuous measurements and quantitative constraints: Challenge problems for discrete modeling techniques," 2001.
- [8] F. Zhao, X. Koutsoukos, H. Haussecker, J. Reich, and P. Cheung, "Monitoring and fault diagnosis of hybrid systems," *IEEE Trans. on Systems, Man, and Cybernetics-B*, vol. 35, no. 6, pp. 1225–1240, 2005.
- [9] F. Harirchi, S. Z. Yong, E. Jacobsen, and N. Ozay, "Active model discrimination with applications to fraud detection in smart buildings," in *IFAC WC*, *Toulouse, France*, 2017.
- [10] M. W. Hofbaur and B. C. Williams, "Hybrid estimation of complex systems," *IEEE Trans. on Systems, Man, and Cybernetics-B*, vol. 34, no. 5, pp. 2178–2191, 2004.
- [11] S. L. Campbell and R. Nikoukhah, Auxiliary signal design for failure detection. Princeton University Press, 2015.
- [12] R. Nikoukhah and S. L. Campbell, "Auxiliary signal design for active failure detection in uncertain linear systems with a priori information," *Automatica*, vol. 42, no. 2, pp. 219–228, 2006.
- [13] F. Harirchi, Z. Luo, and N. Ozay, "Model (in) validation and fault detection for systems with polynomial state-space models," in ACC, 2016, pp. 1017–1023.
- [14] A. Balluchi, L. Benvenuti, M. D. Di Benedetto, and A. L. Sangiovanni-Vincentelli, "Design of observers for hybrid systems," in *HSCC*. Springer, 2002, pp. 76–89.
- [15] R. N. Clark, "Instrument fault detection," *IEEE Trans. on Aerospace and Electronic Systems*, no. 3, pp. 456–465, 1978.
- [16] Z. Wang, D. Anand, J. Moyne, and D. Tilbury, "Improved sensor fault detection, isolation, and mitigation using multiple observers approach," *Systems Science & Control Engineering*, vol. 5, no. 1, pp. 70–96, 2017.
- [17] A. Emami-Naeini, M. M. Akhter, and S. M. Rock, "Effect of model uncertainty on failure detection: the threshold selector," *IEEE Trans.* on Automatic Control, vol. 33, no. 12, pp. 1106–1115, 1988.
- [18] A. Girard, "Reachability of uncertain linear systems using zonotopes," in *HSCC*, vol. 5. Springer, 2005, pp. 291–305.
 [19] Y. Mo and B. Sinopoli, "False data injection attacks in control
- [19] Y. Mo and B. Sinopoli, "False data injection attacks in control systems," in *1st Workshop on SCS*, 2010.
- [20] D. Bertsimas and F. J. de Ruiter, "Duality in two-stage adaptive linear optimization: Faster computation and stronger bounds," *INFORMS Journal on Computing*, vol. 28, no. 3, pp. 500–511, 2016.

Modeling Performance Measurement of Mobile Manipulators

Roger Bostelman^{1,2}, Sebti Foufou³, Tsai Hong¹

¹National Institute of Standards and Technology, Gaithersburg, MD 20899, USA ²Le2i, Université de Bourgogne, BP 47870, 21078 Dijon, France ³ CSE Dept., College of Engineering, PO. Box 2713, Qatar University, Doha, Qatar roger.bostelman@nist.gov

Abstract- Mobile manipulators are being marketed around the world as single manufacturer systems (i.e., robot arm and vehicle manufactured by the same organization) and as independent robot arms, from a different manufacturer and integrated with automatic guided vehicles (AGVs) or mobile robots. Test methods for measuring safety and performance of either of these types of mobile manipulators have yet to be developed and therefore, potential users cannot compare one system to another to match to tasks. Similarly, the control of these systems can vary greatly from AGV control to more autonomous mobile robot control and further, to robot arm control methods. Systems Modeling Language (SysML)¹ is a general-purpose modeling language for systems engineering applications that supports the specification, analysis, design, verification, and validation of simple through complex systems, such as mobile manipulators. This paper uses SysML to describe a method using an artifact for performance measurement of mobile manipulators performing assembly tasks. Uncertainty propagation, a key component in understanding the effects of mobile manipulator constraints, is also modeled and described.

I. INTRODUCTION

Future smart manufacturing systems will include more complex coordination of systems, such as mobile manipulators (i.e., robot arms mounted on mobile bases). Mobile manipulators expand the fixed robot automation of the past into more flexible and capable robots. "Mobile manipulators offer high mobility and manipulability. An ideal utilization of the motion redundancy in the mobile manipulator is to perform assembly tasks on a moving vehicle body while tracking." [1] Mobile base (e.g., automatic guided vehicle (AGV) or mobile robot) and onboard manipulator functionality and performance specifications should be provided by the manufacturer so that the user can match the system to the task, such as assembly or welding tasks. However, to date there are no standard performance measurement methods published so that all mobile manipulator manufacturers can provide similar performance data to the potential system user. Additionally, mobile manipulator control must be known so that users can rapidly and cost effectively program the system to perform as expected. As any quick search on the internet can provide,

there are many programming languages available today and used to control robots, for example: Robot Operating System (ROS), LISP, Assembly, MATLAB, C#/.NET, Java, Python, and C/C++ to name a few. Just as there is a need for a standard performance measurement test method for measuring mobile manipulators, there should also be a standard robot control, as well as modeling, language for use within the test method allowing standard representation of the system under test.

Model Based Systems Engineering (MBSE) provides a simplified representation of a system. Specifically, Systems Modeling Language (SysML) is a graphical modeling language that supports the "specification, analysis, design, verification, and validation of systems that include hardware, software, data, personnel, procedures and facilities." [2] SysML provides four essential tools, also called pillars: Structure (with definition and use), Behavior (with interaction state machines, and activity/function), Requirements, and Parametrics (with equations and units). Rahman, et al [3] say that using SysML can enable the creation of reusable software modules for programming the robot to allow platform independent design and reduced development time. Additionally, Rahman, et al also suggest that SysML is uniquely suited for both accurately modeling increasingly complex and physical robotics systems, as well as the standardization of such an approach useful across many different industries.

Measurement of the robot's Cartesian pose, which is combined with the mobile base's pose, is relatively complex where the system can include nine or more degrees of freedom. To simplify measurement, an artifact was designed to allow various geometric patterns to be traced by the robot wielding a tool point sensor to sense the dimensional points along the patterns. The artifact was designed and manufactured to include a flat surface with embedded geometric patterns to trace allowing for different mobile manipulator performance measurement scenarios. The scenarios include: A) static: the AGV stops while the robot accesses all points within its work volume, B) indexed: the AGV initially stops while the robot accesses most points within its work volume, informs the AGV to increment to a new point, and to stop while the robot accesses the remaining points, and C) dynamic: both the AGV and robot simultaneously move while the robot accesses all points.

Bostelman, Roger; Foufou, Sebti; Hong, Tsai. "Modeling Performance Measurement of Mobile Manipulators." Paper presented at 7th Annual IEEE Int. Conf. on CYBER Technology in Automation, Control, and Intelligent Systems (IEEE-CYBER 2017), Honolulu, HI, United States. July 31, 2017 - August 4, 2017.

¹ Commercial equipment, software, and materials are identified in order to adequately specify certain procedures. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials, equipment, or software are necessarily the best available for the purpose.

This paper applies SysML modeling to the performance measurement of mobile manipulators. The SysML models are



Fig. 1 - (a) NIST mobile manipulator and artifact with left inset showing the Bisect search path using the laser retroreflector. (b) SysML package diagram showing the mobile manipulator and measurement artifact structures.

verified through review of the systems being used in this iteration, including the mobile base (AGV), manipulator (robot arm), and an artifact measurement system. Previous experiments have occurred [5] that also verify the mobile manipulator performance measurement concepts modeled in this paper. The paper first considers the mobile base and onboard manipulator systems in block definition and internal block diagrams showing their interconnections. Similarly, the artifact measurement system used to measure performance is then modeled and described. Following is a discussion on the uncertainty propagation for the performance measurement of mobile manipulators, as well as a description of an example use case model within a production facility.

II. MOBILE MANIPULATOR SYSTEM

An example mobile manipulator system used as basis for this paper is shown in Fig. 1 (a). The system is used for developing mobile manipulator performance test methods and for uncertainty measurements under the National Institute of Standards and Technology (NIST) Robotic Systems for Smart Manufacturing Program [6]. This program provides "the measurement science needed to enable all manufacturers, including small and medium ones, to characterize and understand the performance of robotics systems within their enterprises." The mobile manipulator shown in Fig. 1 (a) provides a collaborative robot measurement platform who's position and orientation (pose) of the mobile base relies on reflectors mounted on the surrounding walls or within the AGV world. The robot arm or manipulator is mounted on the AGV top-front. An artifact that was developed at NIST, called the reconfigurable mobile manipulator artifact (RMMA), and was used for measuring the mobile manipulator as a novel and relatively low cost method.

Fig. 1 (b) shows a SysML package model of the described systems. This high-level model, to be detailed in the following sections, provides an overview of the systems that make up the mobile manipulator and RMMA. The RMMA is shown to include two main components: a bisect-fiducial block that is used for registration of the mobile manipulator to the RMMA and a fiducial-reflector block that can include one or many to make up patterns used to test the mobile manipulator.

The mobile base is an AGV manufactured with the industry's pseudo-standard controller and software, shown on the left side of Fig. 2, as parts to the AGV controller-offboard block. Within the AGV controller-offboard lists the software used to control the vehicle and residing in this offboard computer. This example vehicle has many of the same components found in autonomous industrial vehicles with navigation sensors (Nav sensor) that may or may not require facility reflectors. If the vehicle uses simultaneous localization and mapping (SLAM), features of the facility would be shown in place of the Facility reflectors block. Steer and Drive motors and amplifiers, and batteries are also typical. In the experimental case provided in this paper, there is also an offboard manipulator controller (Manipulator controlleroffboard) which may not be typical of industry as all manipulator control may be onboard the vehicle.

The manipulator block definition diagram, shown on the right side of Fig. 2, has a similar component layout as for the AGV with motors/amplifiers, encoders, and an onboard and offboard controller. Additionally, an end-of-arm tool (EOAT) is shown that includes a laser retro-reflector tool. The Manipulator controller-offboard provides the connection to the onboard AGV controller (CVC600) and lists associated software parts (part is a SysML term for subcomponent).

However, it is essential that the independent AGV and onboard manipulator controllers communicate their poses (position and orientation), in this case relative to the AGV world (facility reflectors). There has been a lot of research in centralized and decentralized offboard robot-to-robot communication [7] and combined controller communication [8], although there is little discussion of combined, yet independently controlled, mobile base and manipulator control

Bostelman, Roger; Foufou, Sebti; Hong, Tsai. "Modeling Performance Measurement of Mobile Manipulators." it. Conf. on CYBER Technology in Automation, Control, and Intelligent Systems (IEEE-CYBER 2017), Honolulu, HI, United States. July 31, 2017 - August 4, 2017. Paper presented at 7th Annual IEEE Int.

communication methods in the literature. The manipulator also can also share power from batteries as shown in the figure. mm step sizes will take much longer to find the reflector center verses a 2 mm step size, although with much less accuracy.



Fig. 2 - SysML internal block diagram of the mobile manipulator subcomponents. The AGV and Robot arm each have independent onboard components with only power (battery) linking them together. Offboard controllers also list the programs (parts) that control their associated systems (e.g., AGV).

The manipulator internal block diagram shown in Fig. 2 also includes an additional constraint of tool positioning along with the base mounting constraint (Manip Base constraint). This part constraint describes the mounting uncertainty that can occur when the manipulator is mounted to the mobile base. And, although the AGV is linked to the Robot arm part due to the onboard manipulator mount, the AGV includes its additional constraint of pose uncertainty. These will be further detailed in section IV. Uncertainty Propagation.

Fig. 3 shows an internal block diagram of the software algorithms that control the manipulator during performance measurements. The indexing test (B) described in section I is modeled. On the lower right are one hardware part (CVC600) and two software parts from the AGV (System Manager Run and CWay). The manipulator is dependent upon the System Manager Run program informing the manipulator of the AGV pose when parked at the RMMA. The manipulator performs intermediate motions to a manipulator base pose that causes the manipulator to approach the Bisect Control (see Fig. 1 (a) left inset) and Search Control registration points on the RMMA in the same way. This ensures the manipulator will not attempt to pass the EOAT through the robot base or perform other selfdestructive motions.

Dependent upon the stakeholder (e.g., test requestor) selection of performance measurement type, either the bisect or search method, or both are performed. Step sizes for the Bisect Control are left variable allowing the operator to choose the time for the manipulator to bisect to find the large reflector center and/or the accuracy of the center. For example, a 0.25



Figure 3 - SysML internal block diagrams of the manipulator control software components.

Similarly, when using only the Search Control for registration to the RMMA, a very small step size provides relatively higher accuracy. However, a step size of half the diameter of fiducial reflector was determined an ideal step size to maximize accuracy measurement and minimize search time. For example, when 1 mm fiducial diameters were used, the 0.5 mm

Bostelman, Roger; Foufou, Sebti; Hong, Tsai. "Modeling Performance Measurement of Mobile Manipulators." it. Conf. on CYBER Technology in Automation, Control, and Intelligent Systems (IEEE-CYBER 2017), Honolulu, HI, United States. July 31, 2017 - August 4, 2017. Paper presented at 7th Annual IEEE Int.

step size was used. Based on the AGV location, either the circle or square pattern is then traversed. Stowing the manipulator was programmed to occur when the pattern was completed or when performing Bisect Control or Search Control that did not produce appropriate results within a chosen time period. For example, if the Bisect Control did not initially result in a reflector detect or if the Search Control took more than 200 steps, the manipulator Stow function was executed and the AGV System Manager Run program was alerted that the AGV could move. A smaller number of steps could also be used.

III. PERFORMANCE MEASUREMENT ARTIFACT

Metrology methods for measuring performance of mobile manipulators, with technologies used to access parts or assemblies in manufacturing processes, listed in [9], include: physical contact using a touch probe, cameras detecting fiducials, laser interferometry, theodolites, coordinate measuring arms, path comparison, trilateration, polar coordinate measuring, triangulation, optical tracking, inertial measuring, Cartesian coordinate, and path drawing.

This section provides an alternative to these methods by using the RMMA with non-contact measurement using a laser retroreflector. This method could potentially prove cost



Figure 4 - SysML internal block diagram showing the reconfigurable mobile manipulator artifact (RMMA) structure.

effective while providing the desired maximum uncertainty for mobile manipulator performance measurement. The RMMA is

expected to be used within a standard test method to measure the performance of static and mobile manipulators. By comparison, it is estimated that the use of the RMMA could be 20 times lower cost than the use of, for example, an optical tracking system. The RMMA is a metal plate with fiducial mount points at precise locations. The RMMA, shown in Fig. 1 beside the mobile manipulator, could also be made using additive manufacturing and estimated to further reduce costs by another order of magnitude. Reflective fiducials are to be detected using a laser retroreflector detector, carried by the manipulator as the EOAT, passing through a collimator attached to the RMMA. A 305 mm (12 in) diameter circle pattern and a 457 mm (18 in) square pattern of fiducials are machined with 0.025 mm (0.001 in) tolerance into the RMMA. Other components are also part of the RMMA where all components are modeled in a SysML internal block diagram shown in Fig. 4.

Beginning at the laser retroreflector (Fig. 4, bottom-left), a positioning constraint is applied to the EOAT provided by the robot manufacturer specification. Moving up the left of the model, the collimator has a 13 mm inside diameter limiting the EOAT angle relative to the RMMA where fiducial detection can occur. The collimator is attached to two different types of fiducial reducers ('fid-refl-reducer-fixed' with a fixed reflector diameter of 2 mm or greater, depending on the EOAT uncertainty chosen, and a 'fid-refl-reducer' with a variable reflector diameter of 1 mm or greater that uses an optical aperture to minimize diameter to the center of the reflector). Both of the fiducial reducers are above 10 mm square fiducialreflectors and attached to the RMMA through surface connectors into circle and square patterns embedded in the machined surface of the RMMA.

Since the mobile manipulator may or may not already be registered to the RMMA, a means is needed to allow this registration. The Fig. 4-left modeled parts can be used for mobile manipulator registration with the RMMA using search methods where the fiducial locations are previously taught. A second set of parts is also modeled in Fig. 4 - right showing the laser retroreflector being used to detect 42 mm diameter reflectors (bisect-refl-reducer) for an alternative mobile manipulator-to-RMMA registration method. The 42 mm diameter was chosen so that the EOAT would always detect these reflectors and a control method, called bisect and shown in the inset of Fig. 1 (a), could be used for the registration process. In other words, as the mobile manipulator indexes from one pose to the next, the vehicle pose combined with the robot arm pose would always detect the large reflector. Offthe-shelf 50 mm x 80 mm rectangular reflectors (bisectreflectors) were covered by the square-to-round 42 mm opening bisect-refl-reducer and mounted to the RMMA using surface connectors at initially taught manipulator locations (registration-laser) on the machined surface.

IV. UNCERTAINTY PROPAGATION

Bostelman, Roger; Foufou, Sebti; Hong, Tsai. "Modeling Performance Measurement of Mobile Manipulators." it. Conf. on CYBER Technology in Automation, Control, and Intelligent Systems (IEEE-CYBER 2017), Honolulu, HI, United States. July 31, 2017 - August 4, 2017. Paper presented at 7th Annual IEEE Int.

As a preliminary notion, the world within which a mobile base, such as an AGV, should be measured and provided to the vehicle controller as reference. The vehicle pose will only be as accurate as its reference. Therefore, AGV calibration is essential to enable higher accuracy and repeatability for the mobile manipulator which references the robot base pose to its mobility system or, in this case, AGV.

A typical method of measuring reflector locations in the world is to use a metrology system, such as a surveyor's tool (i.e., approximately 1.5 mm uncertainty over 1.5 km [10]) or a laser tracker (i.e., approximately 18 µm uncertainty over 12 m [11]). The authors chose the laser tracker so that the AGV reference to the world would be relatively more accurate. An onboard, spinning, navigation laser range and azimuth sensor then provides pose information to the vehicle controller. One issue (i.e., first major uncertainty point (AGV)) with the AGV control is that it uses the measured pose with respect to the world (facility reflectors) and the AGV control reference location is at floor level, at the vehicle centroid (i.e., beneath the vehicle). As such, this location is very difficult to use as a measurement reference.

The robot arm is mounted on a machined breadboard with 51 mm spaced, threaded holes and the robot arm is mounted to the breadboard with a machined interface plate. There is some uncertainty as to how accurately the breadboard is mounted with respect to the AGV reference point and causing a second uncertainty point (Manip Base). The third uncertainty point (EOAT) is the relative accuracy of end-of-arm-tool pose of the carried laser that the robot arm is capable of providing. The uncertainty propagation can therefore be modeled as World which combines the constraints (AGV, Manip Base, and EOAT constraints) and can then be modeled in a block definition diagram, as shown in Fig. 5, which allows each of their constraint parameters to be clearly displayed. Also, the parameters for each of the constraints and interconnects that produce the uncertainty propagation are shown in the figure which can be described in the matrix equation:

$$_{W}P_{E} = _{W}H_{A} * _{A}H_{M} * _{M}P_{E}$$

$$\tag{1}$$

where: P represents points, H represents rotation and translation vectors or sets of 3 x 3 homogeneous equations, W = World, E = EOAT, A = AGV, and M = Manipulator. A SysML parametric diagram, not included here, can then be used to further display the equations within a model.

V. USE CASE

Up to this point, the mobile manipulator system and the measurement system have been modeled, including the uncertainty propagation that can occur from performance measurements. SysML models are therefore needed to show how this information would be useful when applying the mobile manipulator performance measurement concept. Four



Figure 5 - SysML (a) block definition diagram showing the constraints that lead to uncertainty propagation for a mobile manipulator.

models (activity, sequence, state, and use case) are useful to show all aspects of the production case. For this paper and to only demonstrate the modeling concept, a use case diagram is provided.

Fig. 6 shows a SysML use case diagram modeling the process that represents a production facility where a mobile manipulator is used normally from "Execute MM operations with production facility" to "Continue normal production operations". Additionally, and in parallel to normal operation, is the mobile manipulation measurement task. The model shows that the mobile manipulator can also be sent from the production area to "Execute mobile manipulator performance measurement test" within a calibration area "MM (mobile manipulator) System and Measurement Systems package". The mobile manipulator is adjusted upon calibration (violet task), and then returned to "Continue normal production operations". To be thorough, the addition of the three actors (with blue heads) were also needed to perform tests during the author's experimental research. The use of the RMMA (yellow tasks) is dependent upon the stakeholder's requirements for mobile manipulator accuracy and cost. In the research use case, there is of course, no return of the system back into production. Also, it is expected that the 'Adjust MM (mobile manipulator) parameters based on performance tests' task would be performed automatically. However, in a test case, maintenance staff or researchers would log the data and suggest that it passes or fails the performance measurement test.

VI. CONCLUSIONS

Mobile manipulators are relatively complex tools that are now capable of performing manufacturing assembly tasks. Experiments at NIST and referenced in this paper suggest that their uncertainty has been shown to be within 1 mm. The complexity of these systems are exemplified in their

Bostelman, Roger; Foufou, Sebti; Hong, Tsai. "Modeling Performance Measurement of Mobile Manipulators." it. Conf. on CYBER Technology in Automation, Control, and Intelligent Systems (IEEE-CYBER 2017), Honolulu, HI, United States. July 31, 2017 - August 4, 2017. Paper presented at 7th Annual IEEE Int.



Figure 6 - SysML use case diagram of the RMMA (yellow tasks) used to measure performance of a mobile manipulator as may be found in a production facility during operation. The addition on the three blue head actors were required during research and are not required for a production case.

subcomponents (e.g., controller, navigation sensor, amplifiers, wheel/joint encoders, etc.) and the number of degrees of freedom to be controlled. SysML provides a useful method to model, not only the internal subcomponents that make up mobile manipulators but also, the control algorithms.

A test method to measure the performance of mobile manipulators is also critical for users to match the capability of these systems to assembly and other tasks. However, no safety or performance tests are currently standardized for these systems. A novel artifact, called the reconfigurable mobile manipulator artifact (RMMA), has been designed and used at NIST through experimentation to prove the performance of mobile manipulators for assembly tasks. As with the control of these systems, their performance measurement can also be modeled using SysML. The outcome of the models can then show system or component constraints, uncertainty propagation, and use cases in a simple and standardized way.

REFERENCES

- [1] Bradley Hamner, Seth Koterba, Jane Shi, Reid Simmons, Sanjiv Singh, 'Mobile Robotic Dynamic Tracking for Assembly Tasks," IROS 2009.
- [2] Object Management Group Systems Modeling Language (OMG SysML) Tutorial. www.omgsysml.org/INCOSE-OMGSysML-Tutorial-Final-090901.pdf, September, 2009, accessed December 19, 2016.
- [3] Mohd Azizi Abdul Rahman, Katsuhiro Mayama, Takahiro Takasu, Akira Yasuda, and Makoto Mizukawa. "Model-driven development of intelligent mobile robot using systems modeling language (SysML).' ' In Mobile Robots - Control Architectures, Bio-Interfacing, Navigation, Multi Robot Motion Planning and Operator Training. InTech, 2011.
- [4] Bostelman, R., Eastman, R., Hong, T., Enein, O. A., Legowik, S., & Fourfou, S., "Comparison of Registration Methods for Mobile Manipulators", In 19th International Conference on CLAWAR, September 2016.
- [5] Bostelman, R., Eastman, R., Hong, T., Enein, O. A., Legowik, S., & Fourfou, S., "Comparison of Registration Methods for Mobile Manipulators", In 19th International Conference on CLAWAR September 2016.
- [6] National Institute of Standards and Technology (NIST) Robotic Systems Manufacturing Program, for Smart www.nist.gov/programs-

projects/robotic-systems-smart-manufacturing-program, accessed February 8, 2017.

- Hao Su, "Cooperative Control of Payload Transport by Mobile [7] Manipulator Collectives, Master of Science Thesis, Graduate School of the State University of New York, Buffalo, June 2008.
- [8] Keiji Nagatani and Shin'ichi Yuta, "An Autonomous Mobile Manipulator with a Functionally Distributed Control System", Proceedings of the International Conference on Automation (ICAUTO-95) India December 12-14 1995
- Bostelman R, Hong T, Marvel J., "Survey of Research for Performance [9] Measurement of Mobile Manipulators", Journal of Research of the National Institute of Standards and Technology, Volume 121 (2016), March 2015
- [10] Leica Flexline, TS02/06/09. (2008). Leica Geosystems. Accessed 10 February 2017
- Bostelman R, Hong T, Cheok G., "Navigation performance evaluation for automatic guided vehicles", IEEE International Conf. on Technologies for Practical Robot Applications (TePRA), May 11, 2015.

Bostelman, Roger; Foufou, Sebti; Hong, Tsai. "Modeling Performance Measurement of Mobile Manipulators." it. Conf. on CYBER Technology in Automation, Control, and Intelligent Systems (IEEE-CYBER 2017), Honolulu, HI, United States. July 31, 2017 - August 4, 2017. Paper presented at 7th Annual IEEE Int.

Proceedings of the ASME 2018 International Manufacturing Science and Engineering Conference, MŠEC 2018 MSEC 2018 June 18-22, 2018, College Station, USA

MSEC2018-6492

DEVELOPING MAINTENANCE KEY PERFORMANCE INDICATORS FROM MAINTENANCE WORK ORDER DATA

Michael P. Brundage* **KC Morris** National Institute of Standards and Technology Gaithersburg, MD, USA

Thurston Sexton Sascha Moccozet Michael Hoffman National Institute of Standards and Technology Gaithersburg, MD, USA

ABSTRACT

Maintenance management for manufacturing is a crucial activity for improving productivity within a facility. Within this process, maintenance work orders (MWOs) are used when tracking and solving any maintenance-related issue. The MWOs often capture the problem, the solution, at what machine the problem occurred, who solved the problem, when the problem occurred, and other information. These MWOs are manually written by maintenance technicians, entered into a database, or recorded directly into maintenance management software. Technicians often describe or record information informally - or do not record it at all — leading to inconsistencies and/or inaccuracies in the data. This paper outlines maintenance key performance indicators (KPIs), developed using MWOs, that show why consistent and accurate data collection is important for maintenance decision making. The maintenance data, or "elements," and their corresponding KPIs are derived from MWOs from real manufacturers (large manufacturers and small and medium enterprises). While all elements or KPIs are not recorded by every manufacturer, the guideline provided here outlines the elements necessary to calculate specific KPIs. These examples are developed to aid in common maintenance decisions.

INTRODUCTION

As manufacturers face increasing global competition, methods to increase productivity with less resource consumption are needed. Properly executed maintenance management procedures are one method to increase productivity through increased machine life and availability, and through reduced machine failures. These maintenance management procedures are often performed on an ad-hoc basis with little influence from previous historical data. This paper provides key performance indicators (KPIs) to guide maintenance management procedures using historical maintenance work order (MWO) data.

The standard IEC 62264-3 defines maintenance operations management as "the collection of activities which coordinate, direct and track the functions that maintain the equipment, tools and related assets to ensure their availability for manufacturing and ensure scheduling for reactive, periodic, preventative, or proactive maintenance" [1]. Maintenance operations management involves providing maintenance responses to equipment problems, scheduling and performing maintenance based on time or cycles of a machine or part, providing condition based maintenance, or optimizing resource operating performance and efficiency. Within IEC 62264-3, a number of steps are defined for performing maintenance operations management. One step is the development and management of key performance indicators related to maintenance. Key performance indicators are defined in ISO 22400-1 as a "quantifiable level of achieving a critical objective", while the elements of a KPI are defined as "relevant measurements for use in the formula of a key performance indicator" [2]. Maintenance KPIs aid maintenance decisions, such as "what machine to send a technician?", "who to send to solve a problem?", "what is the most likely cause of the problem?",

Copyright © 2018 by ASME

1

^{*}Corresponding Author: mpb1@nist.gov

Brundage, Michael: Hoffman, Michael: Moccozet, Sascha: Morris, Katherine: Sexton, Thurston,

[&]quot;Developing Mainteance Average Static States, Morris, Nathenne, Sexton, Thurston. "Developing Maintenance Key Performance Indicators from Maintenance Work Order Data." Paper presented at 46th SME North American Manufacturing Research Conference (NAMRC) and the ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18, 2018 - June 22, 2018.

"what are common problems throughout the facility?", "how is a machine performing?".

Currently, no standardized set of maintenance KPIs exists, but existing research explores maintenance KPIs and how they relate to higher level organizational goals. Multiple maintenance KPI studies identify the relationship between maintenance performance measures and higher-level corporate and manufacturing objectives [3, 4]. Muchiri et al. claim that maintenance KPIs should be developed so they are aligned with higher-level functions in the organization ensuring that achieving maintenance performance objectives will support long-term strategic goals [3]. Wireman offers strategies for measuring maintenance performance and emphasizes the importance of keeping corporate objectives in mind [4]. Horenbeek et al. researched maintenance KPI selection determining which performance measures have the greatest impact on high-level objectives [5].

The use of computerized maintenance management systems (CMMS) has also been investigated thoroughly for the storage and retrieval of maintenance KPIs [6]. Although used in many applications, the effectiveness of a CMMS is often limited by its implementation. Many systems collect and analyze maintenance data, but further development of decision support capabilities of such systems is needed [7]. Another limitation of existing systems is the restriction of proprietary CMMS software packages. These systems do not use standard KPIs, making it difficult to compare performance across different CMMSs [8].

Often, the CMMS data has an unstructured, natural language component that can lead to inconsistencies in data analysis. No substantial work connecting raw data collection in the form of maintenance logs to the use of maintenance KPIs for decision making in manufacturing has been found. This paper bridges the gap between MWO data elements and maintenance KPIs.

The rest of the paper is structured as follows: Section 1 defines the data elements from MWOs used for maintenance KPIs defined in Section 2. Subsection 2.1 discusses common problem hot spot KPIs related to diagnostics and trend analysis, subsection 2.2 provides machine KPIs, and subsection 2.3 defines maintenance technician expertise KPIs. Each KPI subsection describes a KPI using elements defined in Section 1 and gives multiple examples of specific KPIs and how they relate to maintenance decisions. These KPIs are calculated from commonly measured elements in industry. Future work will expand these KPIs and link them to possible maintenance decisions as is discussed in Section 3.

Maintenance Element Definitions 1

Common MWO elements (measurements) defined in this section have been generated using actual MWO data from multiple manufacturers, and are almost certainly not all-inclusive. The source MWO data ranged from hand-written text (and entered into spreadsheets) to fully automated CMMS system read-

outs. At least one unstructured, natural language text field was present in every work order studied. The elements in this paper were not measured by every manufacturer studied. Instead, the list was generated using commonalities among the datasets. This list provides a reference for calculating maintenance KPIs from MWOs.

1.1 Date and Time Elements

The date and time elements address the timing involved in a maintenance work order. These elements may include only dates or both dates and times. More accurate results require more precise measurements (for example calculating date and time to the second). The most commonly calculated time elements are Work Order Start Time and Work Order Completion Time. Section 1.2 describes the importance of measuring each of these time steps for more precise decision making. The following elements, represented by the variable on the right, are used in the remainder of the paper to formulate KPIs.

Machine Down Time-stamp	Md
The date and time when the machine goes down.	
Work Order Start Time-stamp	Wd
The date and time when the work order is started.	
Maintenance Technician Arrives Time-stamp	T _d
The date and time when the maintenance technician a at the machine	rrives
Problem Found Time-stamp	Sd
The date and time when the problem is found at the ma	chine.
Part(s) Ordered Time-stamp	$P_d^{(i)}$
The date and time when part <i>i</i> is ordered.	u
Part(s) Received Time-stamp	$P_u^{(i)}$
The date and time when part <i>i</i> is received. This can be inventory or from an outside supplier.	from
Problem Solved Time-stamp	Su
The date and time when the problem is solved at the ma	chine.
Machine Up Time-stamp	Mu
The date and time when the machine returns to operation	on.
Work Order Completion Time-stamp	W_u
The date and time when the work order is completed.	

1.2 Calculated Time Elements

Calculated Time Elements illustrate the amount of time spent during different time periods of the MWO described in Table 1. These elements are not always directly measured, but can be calculated using the Date and Time Elements in Section 1.1. The variable k represents the k^{th} issue for a MWO. The Calculated Time Elements provide a better understanding of the time spent on each portion of the maintenance work order, compared to the Date and Time Elements. For example, a maintenance manager can analyze the amount of time spent repairing an issue as compared to the amount of time only spent diagnosing the issue. A maintenance technician could then be trained on how

Copyright © 2018 by ASME

2

Brundage, Michael: Hoffman, Michael: Moccozet, Sascha: Morris, Katherine: Sexton, Thurston,

[&]quot;Developing Mainteance Average Static States, Morris, Nathenne, Sexton, Thurston. "Developing Maintenance Key Performance Indicators from Maintenance Work Order Data." Paper presented at 46th SME North American Manufacturing Research Conference (NAMRC) and the ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18, 2018 - June 22, 2018.

to better diagnose certain types of issues, if diagnosing the issue takes disproportionately long for the MWO.

1.3 Human Elements

Human elements identify the humans and their traits as seen in Table 2. They include the operator, maintenance technician, and the skill level required for the MWO. A heuristic often determines what skill level is needed for a particular issue. Techniques using historical data to determine the required skill level to solve a problem will be explored in future work.

1.4 Machine Elements

Machine elements consist of machine related attributes as seen in Table 3. This data includes information about the machine, components of the machine, and the part in progress.

1.5 Raw Text Elements

The Raw Text elements consist of free text descriptions of the issue from a MWO as seen in Table 4. Often times, no standard format for this information exists. Examples provided in the table contain misspellings and different formats to illustrate the inconsistency and variation of the information captured.

One important step to defining maintenance KPIs is transforming the unstructured raw text data into a structured format for analytics. The next subsection describes the structured elements and the procedure for generating them.

1.6 Tag Elements

The Tag Elements represent the actions taken and items of interest within each part of the MWO. They are described in Table 5. These elements provide structure for the unstructured data found in the raw text elements. Previous work in [9, 10] researched automated natural language processing (NLP) methods to provide structured data from unstructured natural language input in MWOs using a procedure called "tagging." A tag element is a single unique token that represents a particular action or item.

Figure 1 shows the difference between Raw Text Elements and Tag Elements. The example shows a combination of multiple problems and solutions in the Raw Text Elements. The Raw Text Elements are first separated into Problem Items and Solution Items and Problem Actions and Solution Actions categories. To facilitate the calculation of several later-discussed KPIs, those elements are then combined into Problem Item & Action and Solution Item & Action.

Without this tagged data, data analysis and KPI measurement is difficult. If a maintenance manager analyzed the commonality of the "Description of Problem" fields, the results would only include specific descriptions. For example if "Brush unit forward" was part of the MWO 3 times and "Brush unit stuck forward - motor not spinning" was part of the MWO 2 times, both descriptions contain the problem "Brush unit forward." Only analyzing the "Description of Problem" field leads to two separate issues, instead of the common issue "Brush unit forward" due to the inconsistency in the data input. This use of natural language including pervasive jargon and abbreviations is the norm in the MWOs studied and one of the primary impediments to reusing this data.

2 Maintenance Key Performance Indicators

Maintenance KPIs are calculated from these MWO data elements to aid in maintenance operations management decisions. This paper describes a selection of KPIs that can be calculated using these MWO data elements. A guideline on creating and selecting maintenance KPIs while accounting for different stakeholders using ASTM E3096-17 [11,12] will be explored in future work.

2.1 Common Problem Hot Spot Indicators

Common problem hot spot KPIs for diagnostics and problem tracking throughout the facility are described in this section. Tag Elements are required data for common problem hot spot KPIs. To perform trend analysis, comparing the facility performance to either a baseline historical level or to other facilities, is necessary. Including Machine Elements or Calculated Time Elements allows for more in depth analysis of common problem hot spots. These elements are summarized below:

COMMON PROBLEM HOT SPOTS: ELEMENTS Required:

Tag Elements

Optional:

Machine Elements; Calculated Time Elements

Common problem hot spot KPIs with Machine Elements identifies trouble spots, provides comparisons of problems at different machines, and analyzes machine performance for specific problems. Calculating common problem hot spot KPIs with Calculated Time Elements tracks time spent on specific problems. Examples of common problem hot spot KPIs are discussed in the following subsections.

2.1.1 Common Problem Items The "Common Problem Items" KPI investigates the number of issues at a facility for specific items. For example:

MWO issued 40 times for a Bearing.

MWO issued 20 times for Gears.

This KPI is used to compare problem items for outliers in the number of MWOs issued for a specific item or to compare problems among different facilities or years. For example:

2012 MWO issued 30 times for bearings.

Copyright © 2018 by ASME

Brundage, Michael; Hoffman, Michael; Moccozet, Sascha; Morris, Katherine; Sexton, Thurston. "Developing Maintenance Key Performance Indicators from Maintenance Work Order Data." Paper presented at 46th SME North American Manufacturing Research Conference (NAMRC) and the ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18, 2018 - June 22, 2018.

Name	Description	Formula	Notes on Data
Time Between Failure	The time between machine failures (from machine up $M_u(k\!-\!1)$ to machine down $M_d(k).$	$\mathbf{M}_{\mathrm{d}}(k)-\mathbf{M}_{\mathrm{u}}(k-1)$	This is the time between the last failure (k-1) and the current failure (k)).
Time to Repair	The time from when the machine is down $M_d(k)$ until the machine is up $M_u(k)$.	$\mathbf{M}_{\mathbf{u}}(k) - \mathbf{M}_{\mathbf{d}}(k)$	This can be misleading as the entire duration of this time period might not be spent repairing a machine.
Work Order Completion Time	The time the maintenance work order is issued $W_d(k)$ until when it is closed $W_u(k)$.	$W_u(k) - W_d(k)$	This is often used to approximate Time to Repair for a work order.
Time to Dispatch	The time from when the machine is down $M_d(k)$ until the maintenance technician arrives $T_d(k)$.	$\mathbf{T}_{\mathbf{d}}(k) - \mathbf{M}_{\mathbf{d}}(k)$	
Time to Return to Opera- tion	The time from when the maintenance technician arrives $T_d(k)$ until the machine is up $M_u(k)$.	$\mathbf{M}_{\mathbf{u}}(k) - \mathbf{T}_{\mathbf{d}}(k)$	
Time to Issue Work Or- der	The time from when the machine is down $M_d(k)$ until the work order is issued $W_d(k)$.	$W_d(k) - M_d(k)$	
Time to Travel	The time from when the work order is issued $W_d(k)$ until the maintenance technician arrives $T_d(k)$.	$T_d(k) - W_d(k)$	
Time to Solve Problem	The time from when the maintenance technician arrives $T_d(k)$ until the time when the problem is solved $S_u(k)$.	$S_u(k) - T_d(k)$	
Time to Diagnose	The time from when the maintenance technician arrives $T_d(k)$ until the time when a problem cause is found $S_d(k)$.	$S_d(k) - T_d(k)$	This can be difficult to calculate since it is hard to calcu- late when diagnosing stops and fixing a problem begins.
Time to Order	The time from when the problem cause is found $S_d(k)$ until the time when a part is ordered $P_d^{(i)}(k)$.	$\mathbf{P}_{\mathrm{d}}^{(\mathrm{i})}(k) - \mathbf{S}_{\mathrm{d}}(k)$	This is only necessary if a part i needs to be replaced to solve the problem.
Lead Time for Part	If a part <i>i</i> is ordered, the time from when the part is ordered $P_d^{(i)}(k)$ until the technician begins work on repairing the problem $P_u^{(i)}(k)$.	$\mathbf{P}_{\mathrm{u}}^{(\mathrm{i})}(k) - \mathbf{P}_{\mathrm{d}}^{(\mathrm{i})}(k)$	This can be applied if a part is in inventory.
Time to Fix	The time the maintenance technician begins work on repairing the problem $P_u^{(i)}(k)$ until the problem is solved $S_u(k)$.	$\overline{\mathbf{S}_{\mathbf{u}}(k)} - \mathbf{P}_{\mathbf{u}}^{(i)}(k)$	This can be difficult to calculate since it is hard to calcu- late when diagnosing stops and fixing a problem begins.
Time to Turn On	The time from when the problem is solved $S_u(k)$ until the machine is up $M_u(k)$.	$M_u(k) - S_u(k)$	

TABLE 1: CALCULATED TIME ELEMENTS

TABLE 2: HUMAN ELEMENTS

Name	Description	Examples	Notes on Data
Maintenance Technician	The person responsible for fixing the machine.	"John Smith" "J.S."	The format might not be consistent as illustrated by the examples. Mul- tiple technicians or a third party might solve the problem. This type of information should be captured here.
Operator	The person operating the machine when a MWO was issued.	"Warren, Bill" "B.W."	The format might not be consistent as illustrated by the examples. It is possible to have one operator for multiple machines. This can be null if the machine is automated.
Skill or Craft	The required skill level or craft needed to respond to the MWO.	"Mechanical" "Electrical"	The format for this field will depend on the company. Smaller companies might not have this field.

2013 MWO issued 600 times for bearings.

This information enables managers to investigate the increase in issues in a given year for a given part.

2.1.2 Common Problem Items per Time Between Failure The "Common Problem Items per Time Between Failure" KPI investigates the time between issues at a facility for

Copyright © 2018 by ASME

4

Brundage, Michael; Hoffman, Michael; Moccozet, Sascha; Morris, Katherine; Sexton, Thurston. "Developing Maintenance Key Performance Indicators from Maintenance Work Order Data." Paper presented at 46th SME North American Manufacturing Research Conference (NAMRC) and the ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18, 2018 - June 22, 2018.

TABLE 3: MACHINE ELEMENTS

Name	Description	Examples	Notes on Data
Machine Name	The machine where the problem occurred.	"Milling Machine 1" "HURCO 1" "Machine H1" "HURCO VMX24i"	The format depends on the company.
Machine Manu- facturer	The manufacturer who made the machine where the problem occurred.	"HURCO" "Hydromat"	The manufacturer might not be known for older ma- chines.
Machine Type	The type of process the machine was performing when the MWO was issued.	"Milling" "Drilling"	The machine might be capable of multiple processes.
Machine Loca- tion	The area where the machine was located.	"Machining Line 1" "Line A" "Mechanical Cell"	This cannot be calculated if the production line is not broken into distinct areas.
Part in Process	The part being processed by the machine when the MWO was issued.	"Part A" "CA10110"	The format depends on the company.
Necessary Part	If a part is necessary for repair, this is the part that was ordered or taken from inventory.	"Gear 1012" "Bearing A2"	The format will be dependent on the company and ma- chine.

TABLE 4: RAW TEXT ELEMENTS

Name	Description	Examples	Notes on Data
Description of Problem	A free text description of a main-	"Hydraulic leak/leaking valve"	This element can include both "cause"
	tenance problem.	"Noise at spindle because of lose bearing"	and 'effect" information.
Description of Observed Symp-	A free text description of the ob-	"Hydraulic leak at accumulator"	A "cause" versus an "effect" is often difficult to decipher.
toms (Effects)	served effects.	"Noise at spindle"	
Description of Cause	A free text description of the cause of a problem.	"Leakin valve caused hydraulic leak" "loose bearing"	It can often be difficult to capture what is truly the root cause of a problem.
Description of Solution	A free text description of the solu-	"Repiar valve and remove gear"	A MWO might involve multiple solu-
	tion taken to solve the problem.	"replace bearing at machine h1"	tions.

TABLE 5: TAG ELEMENTS

Name	Description	Examples	Notes on Data
Item	The item(s) described in the description of problem, effects, cause, or solution.	"Hydraulic System" "Bearing"	The item described can be different for problem, effect, cause, and solution.
Action	The action(s) described in the description of problem, effects, cause, or solution.	"Broken" "Noise" "Repair"	The action described can be different for problem, ef- fect, cause, and solution.
Item & Action	The item and action pair(s) described in the description of problem, effects, cause, or solution.	"Broken Gear" "Replace Hose"	These item & action pairs can be different for problem, effect, cause, and solution.

specific items. For example:

MWO issued every 10 days on average for bearings

MWO issued every 20 days on average for gears.

Similarly to "Common Problem Items", comparing issues against other facilities or previous year's data is possible with this KPI. Estimating failure rates of specific parts is possible, but this is not accurate as a MWO does not always indicate a failure

of a part.

2.1.3 Common Problem Items by Machine Type The "Common Problem Items by Machine Type" KPI investigates the number of issues at a certain machine type for specific items. For example:

MWO issued 40 times for Milling Machine Bearings

Copyright © 2018 by ASME

5

Brundage, Michael; Hoffman, Michael; Moccozet, Sascha; Morris, Katherine; Sexton, Thurston. "Developing Maintenance Key Performance Indicators from Maintenance Work Order Data." Paper presented at 46th SME North American Manufacturing Research Conference (NAMRC) and the ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18, 2018 - June 22, 2018.

Raw Text	Description of Problem		Description of Solution		
Elements "Brush unit stuck forward – motor not spinning" "Unjammed s			indle, replaced worn pulley, belt too tight"		
Tag Elements					
Problem Act	tions	Problem Items	Solution Items	Solutions Actions	
{stuck, will_not_spin, ja	am, worn, tight} {brus	h_unit, motor, spindle, pulley, belt}	{spindle, pulley}	{unjam, replace}	
<u>Pi</u>	Problem Action & Items			n & Items	
{brush_unit stuck} & {motor will_not_spin} &			{unjam spin	dle}&	
{spindle jar	n} & {pulley worn} & {b	elt tight}	replace pt}	ılley}	

FIGURE 1: Illustrating the procedure to develop structured data (Tag Elements: Table 5) from unstructured, natural language text (Raw Text Elements: Table 4). This example shows how the data is transformed from **Description of Problem** and **Description of Solution** to Items and Actions and finally to Item & Action pairs. The problem and solution actions and items can come from both the Description of Problem and Description of Solution elements. Adapted from [10].

MWO issued 30 times for Milling Machine Gears compared to

MWO issued 60 times for Drilling Machine Bearing MWO issued 80 times for Drilling Machine Gears

This KPI compares performance of specific machine types. It can also serve as a baseline to compare different machines of the same machine type. For example:

MWO issued 200 times for Milling Machine #1 Gears

MWO issued 20 times for Milling Machine #2 Gears

when the average for all Milling Machines is known to be 30 Gear MWOs — Milling Machine #1 is likely an outlier.

2.1.4 Common Problem Item & Action per Time to Repair The "Common Problem Item & Action per Time to Repair" KPI investigates the amount of time to solve an issue for a specific Item & Action. For example:

Broken Gears: 20 hours total repair.

Worn Bearings: 10 hours total repair.

Comparing types of issues for specific Item & Action elements allows for identification of outliers of types of failures that need to be addressed.

2.2 Machine Performance Indicators

Tracking performance of a machine and simulation of machine failures are possible with this set of Machine Performance KPIs. The required elements are Machine Elements and Calculated Time Elements. To analyze machine performance as related to specific issues requires Tag Elements.

MACHINE PERFORMANCE ELEMENTS **Required:**

Machine Elements; Calculated Time Elements

Optional: Tag Elements

Some possible examples of machine KPIs are detailed in the following subsections.

2.2.1 Machine per Time Between Failure The "Machine per Time Between Failure" KPI investigates the amount of time between failures for a given machine. For example:

Machine #1: MWO Issued every 10 days on average.

Machine #2: MWO Issued every 20 days on average.

Comparing how often a maintenance work order is issued for a specific machine is provided by this KPI. Tag Elements are needed to determine the specific issues for each work order. This KPI can also be used for simulation of a facility by inserting random downtime events for specific machines.

2.2.2 Machine by Problem Action per Time Between Failure The "Machine by Problem Action per Time Between Failure" KPI investigates the amount of time between failures for a given machine for a given issue. For example:

Machine #1	Broken every 10 days on average
	Cleaned every 30 days on average
Machine #2:	Broken every 100 days on average
	Cleaned every 50 days on average

Simulation of a facility by inserting random downtime distributions for machines with multiple failure types is possible with this KPI. More specific information is provided with this KPI as compared to the "Machine per Time Between Failure" KPI for how often and why the MWO was issued for a machine.

Copyright © 2018 by ASME

6

"Developing Mainteance Average Static States, Morris, Nathenne, Sexton, Thurston. "Developing Maintenance Key Performance Indicators from Maintenance Work Order Data." Paper presented at 46th SME North American Manufacturing Research Conference (NAMRC) and the ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18, 2018 - June 22, 2018.

Brundage, Michael: Hoffman, Michael: Moccozet, Sascha: Morris, Katherine: Sexton, Thurston,

2.2.3 Machine Type per Time to Repair The "Machine Type per Time to Repair" KPI investigates the time to repair for a given machine type. For example:

Milling Machines: 20 hours spent solving issues

Drilling Machines: 50 hours spent solving issues.

Analysis of the length of repairs for different machine types is provided with this KPI. This KPI provides comparison between specific machines within the same machine type. For example:

Milling Machine #1: 50 hours spent solving issues

Milling Machine #2: 40 hours spent solving issues

where the average time spent solving Milling Machines issues is 20 hours. A maintenance manager can investigate why time spent on issues for Milling Machine 1 and 2 is much higher than the average Milling Machine.

2.2.4 Machine Type per Problem Item per Time to **Repair** The "Machine Type per Problem Item per Time to Repair" KPI investigates the time to repair specific items for a given machine type. For example:

Milling Machines: 20 hours spent solving issues for gears 10 hours spent solving issues for bearings Drilling Machines: 80 hours spent solving issues for gears

100 hours spent solving issues for bearings

This KPI provides comparison between types of items that have issues on different types of machines. A maintenance manager can analyze the usage of a machine to investigate why a part might fail more often in one type of machine compared to another type of machine.

2.3 Maintenance Technician Expertise Indicators

Indicators related to maintenance technician expertise are used for tracking the number of times a maintenance technician addresses an issue. Training procedures to help maintenance technicians acquire more expertise in an area can be developed with the information provided by these KPIs. Lastly, dispatching procedures can be developed for sending the most experienced maintenance technician to solve a type of problem at a specific machine. For this set of KPIs, the required elements are Human Elements as well as either Tag Elements and/or Machine Elements. If Tag Elements are utilized, analysis of a maintenance technician's expertise with certain problems or solutions is provided. If Machine Elements are used in the KPI calculation, maintenance technician experience for specific machines or machine types/manufacturers is calculated. Adding in Calculated Time Elements to these KPIs allows analysis on the time spent on maintenance issues by a specific technician. These elements are summarized below:

MAINTENANCE TECHNICIAN EXPERTISE ELEMENTS

Required:

Human Elements; Tag Elements or Machine Elements **Optional:**

Calculated Time Elements

Some examples of maintenance technician expertise KPIs are detailed in the following subsections.

2.3.1 Maintenance Technician Expertise per Problem Item The "Maintenance Technician Expertise per Problem Item" KPI investigates the number of times a maintenance technician works on a specific item. For example:

Maint. Tech. A:	40 times working on gears
	20 times working on bearings
Maint. Tech. B:	30 times working on gears
	100 times working on bearings

Comparing different maintenance technician's expertise on specific items within a machine, determining where a maintenance technician needs training on specific items, or dispatching a maintenance technician based on expertise are all possible with this KPI.

2.3.2 Maintenance Technician Expertise per Problem Action The "Maintenance Technician Expertise per Problem Action" KPI investigates the number of times a maintenance technician spends on a specific action. For example:

Maint. Tech. A:	40 times working on leaks
	20 times working on low pressure
Maint. Tech. B:	30 times working on leaks
	100 times working on low pressure

This KPI compares maintenance technician's expertise for specific actions and illuminates possible training for technicians on how to respond to various problems.

2.3.3 Maintenance Technician Expertise per Problem Item & Action per Time to Repair The "Maintenance Technician Expertise per Problem Item & Action per Time to Repair" KPI investigates the amount of time a maintenance technician spends repairing an issue. For example:

Maint. Tech. A:	40 hours working on hydraulic leaks
	20 hours working on bearing failure
Maint. Tech. B:	30 hours working on hydraulic leaks
	100 hours working on bearing failure

Dispatching a maintenance technician for repair based on the current observed issues at a machine is possible with this KPI.

2.3.4 Maintenance Technician Expertise per Problem Item & Action per Time to Diagnose The

Copyright © 2018 by ASME

7

Brundage, Michael; Hoffman, Michael; Moccozet, Sascha; Morris, Katherine; Sexton, Thurston. "Developing Maintenance Key Performance Indicators from Maintenance Work Order Data." Paper presented at 46th SME North American Manufacturing Research Conference (NAMRC) and the ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18, 2018 - June 22, 2018.

"Maintenance Technician Expertise per Problem Item & Action per Time to Diagnose" KPI investigates the amount of time a maintenance technician spends diagnosing an issue. For example:

Maint. Tech. A:	40 hours diagnosing hydraulic leaks 20 hours diagnosing bearing failure
Maint. Tech. B:	30 hours diagnosing hydraulic leaks

Dispatching a maintenance technician to diagnose an issue based on the current observed issues at a machine is possible with this KPI. This would involve sending one technician to diagnose the problem and then dispatching the technician with the most expertise to fix the given issue.

2.3.5 Maintenance Technician Expertise per Solution Item & Action per Time to Fix The "Maintenance Technician Expertise per Problem Item & Action per Time to Fix" KPI investigates the amount of time a maintenance technician spends solving an issue. For example:

Maint. Tech. A:	40 hours repairing leaks
	20 hours replacing accumulators
Maint. Tech. B:	30 hours repairing leaks
	100 hours replacing accumulators

Dispatching a maintenance technician to solve the problem more efficiently, once the problem is diagnosed, is possible with this KPI.

2.3.6 Maintenance Technician Expertise per Machine Type per Time to Repair The "Maintenance Technician Expertise per Machine" KPI investigates the number of times a maintenance technician spends on a particular machine type. For example:

Maint. Tech. A:	40 hours working on milling machines
	20 hours working on drilling machines
Maint. Tech. B:	70 hours working on milling machines
	100 hours working on drilling machines

This KPI illuminates comparison of a maintenance technician's expertise for different machine types and can identify which technicians need training for a type of machine.

3 Conclusions & Future Work

The different possible elements of a MWO and associated KPIs for common problem hot spots, machine performance, and maintenance technician expertise are discussed in this paper. These examples do not provide every possible KPI, but they illustrate how the MWO elements can be assembled into indicators used for decision making. Each KPI type is presented at a high level to illustrate which elements are necessary for calculation and what decisions can be made with that type of KPI. More detailed examples are provided to illustrate how a specific KPI can be calculated with maintenance work order elements depending on the needs of the factory. Equations and analysis techniques will be explored in future work to demonstrate how to calculate the KPIs discussed in this paper.

Formal guidelines on how to properly use MWO data for maintenance decisions will enable widespread adoption of this work. Many of the KPIs required structured data for accurate results. The guidelines will include instructions on structuring data, data storage, cleaning data (using the method developed in [10]), calculating KPIs, and making the correct decisions based on KPI results.

Several areas for exploration follow from this work:

- 1. Incorporation of sensor data into the information framework: Merging MWO data with sensor data to aid in maintenance decision making is an important research topic as sensor technologies become cheaper and easier to implement.
- 2. Implications for system level decision making: KPIs that are useful for machine level decisions are presented in this paper. More data needs to be collected to properly make decisions at the system level. This system level information allows maintenance decisions to be linked to operations to perform maintenance procedures without affecting productivity of the facility.
- 3. Roles of different decision makers and divergent interfaces to the data: Roles, such as the operator, maintenance technician, and maintenance manager are not discussed in this paper. Future work will investigate which KPIs are important to each decision maker and how to properly visualize this data.

REFERENCES

- [1] IEC 62264-3:2016, 2016. Enterprise-control system integration - Part 3: Activity models of manufacturing operations management. International Organization for Standardization.
- [2] ISO 22400-1:2014, 2014. Automation systems and integration - Key performance indicators (KPIs) for manufacturing operations management - Part 1: Overview, concepts and terminology. International Organization for Standardization
- [3] Muchiri, P., Pintelon, L., Gelders, L., and Martin, H., 2011. "Development of maintenance function performance measurement framework and indicators". International Journal of Production Economics, 131(1), pp. 295-302.
- [4] Wireman, T., 2005. Developing Performance Indicators for Managing Maintenance, 2nd ed. Industrial Press, New York.
- [5] Horenbeek, A. V., and Pintelon, L., 2014. "Development of

Copyright © 2018 by ASME

8

"Developing Mainteance Average Static States, Morris, Nathenne, Sexton, Thurston. "Developing Maintenance Key Performance Indicators from Maintenance Work Order Data." Paper presented at 46th SME North American Manufacturing Research Conference (NAMRC) and the ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18, 2018 - June 22, 2018.

Brundage, Michael: Hoffman, Michael: Moccozet, Sascha: Morris, Katherine: Sexton, Thurston,

a maintenance performance measurement frameworkusing the analytic network process (anp) for maintenance performance indicator selection". *Omega*, 42(1), pp. 33 – 46.

- [6] Garg, A., and Deshmukh, S. G., 2006. "Maintenance management: literature review and directions". *Journal of Quality in Maintenance Engineering*, 12(3), pp. 205–238.
- [7] Labib, A. W., 2004. "A decision analysis model for maintenance policy selection using a cmms". *Journal of Quality in Maintenance Engineering*, **10**(3), pp. 191–202.
- [8] Rastegari, A., and Mobin, M., 2016. "Maintenance decision making, supported by computerized maintenance management system". IEEE, pp. 1–8.
- [9] Sharp, M., Sexton, T., and Brundage, M. P., 2017. *Toward Semi-autonomous Information*. Springer International Publishing, Cham, pp. 425–432.
- [10] Sexton, T., Brundage, M. P., Morris, K., and Hoffman, M., 2017. "Hybrid datafication of maintenance logs from aiassisted human tags". IEEE Big Data 2017, pp. 1–8.
- [11] ASTM 3096-17, 2017. New guide for definition, selection, and composition of key performance indicators for environmental aspects of manufacturing processes. ASTM International.
- [12] Kibira, D., Brundage, M. P., Feng, S., and Morris, K., 2017. "Procedure for selecting key performance indicators for sustainable manufacturing". *Journal of Manufacturing Science and Engineering*.

Copyright © 2018 by ASME

9

Proceedings 2018 ASME International Manufacturing Science and Engineering Conference **MSEC2018** June 18-22, Texas A&M University, USA

MSEC2018-6603

MONITORING, DIAGNOSTICS, AND PROGNOSTICS FOR ROBOT TOOL CENTER ACCURACY DEGRADATION

Guixiu Qiao National Institute of Standards and Technology Gaithersburg, Maryland, USA

Brian A. Weiss National Institute of Standards and Technology Gaithersburg, Maryland, USA

KEYWORDS

Condition Monitoring, Diagnostics, Prognostics Maintenance, Manufacturing Processes, Robot Systems, Robot Performance Degradation

ABSTRACT

Over time, robots degrade because of age and wear, leading to decreased reliability and increased the potential for faults and failures. The effect of faults and failures impacts robot availability. Economic factors motivate facilities and factories to improve maintenance techniques and operations to monitor robot degradation and detect faults, especially to eliminate unexpected shutdowns. Since robot systems are complex, with sub-systems and components, it is challenging to determine these constituent elements' specific influences on the overall system performance. The development of monitoring, diagnostic, and prognostic technologies, which is collectively known as Prognostics and Health Management (PHM), can aid manufacturers in maintaining the performance of robot systems by providing intelligence to enhance maintenance and control strategies. This paper presents the strategy of integrating top level and component level PHM to detect robot performance degradation (including robot tool center accuracy degradation), supported by the development of four-layer sensing and analysis structure. The top level PHM can quickly detect robot tool center accuracy degradation through the advanced sensing and test methods developed at the National Institute of Standards and Technology (NIST). The component level PHM supports the deep data analysis for root cause diagnostics and prognostics. A reference data set is collected and analyzed using the integration of top level PHM and component level PHM to understand the influence of temperature, speed, and payload on robot's accuracy degradation.

INTRODUCTION

Robot systems play important roles in many manufacturing environments including automotive, electronics, consumer packaged goods, and aerospace manufacturing [1, 2]. The

applications of robots in manufacturing systems bring benefits through both improving flexibility and reducing costs [3-5]. Robot work cells have become more complex, especially when considering robot-robot and human-robot operations [6-9]. More complexity leads to more sources of faults and failures, which can compromise the efficiency, quality, and productivity of a manufacturing system. Moreover, new innovative technologies are making robot work cells more accurate and intelligent, enabling them to be applied to some new applications [4, 10, 11]. New applications include material removal, high precision assembly, two-side drilling and fastening, in-process inspection, and three-dimensional (3D) composite material layout. New technologies often introduce new types of challenges that may not be fully understood. The afore-mentioned applications require high accuracy in both robot position and path. The degradation of robot tool center accuracy can lead to a decrease in manufacturing quality and production efficiency. It is important to understand robot accuracy degradation so that maintenance and control strategies can be optimized.

There are many challenges in developing monitoring, diagnostics, and prognostics for robot tool center accuracy degradation. First, robot tool center accuracy degradation may be difficult to detect, in a timely manner, because the robot may still be operating without any obvious signs of degradation, e.g., the robot being frozen or performing an undesirable activity. Second, with more diverse systems, sub-systems, and components integrated to increase robot work cell capabilities, further challenges are introduced in determining an element's specific influence(s) on the overall system performance [12, 13]. Third, continuous changes to an existing system give rise to new relationships that may lead to greater complexity. This complexity may include, dynamic robotic configurations (e.g., reconfiguration of the instrument layout and production processes), working parameters (e.g., program changes, temperature, payload, speed, part/grasp changes which causes force and torque changing), controller changes (e.g., control strategy, proportional-integral-derivative (PID) tuning), and worn parts [14]. To address these barriers and challenges,

Qiao, Guixiu; Weiss, Brian. "Monitoring, Diagnostics, and Prognostics for Robot Tool Center Accuracy Degradation." Paper presented at ASME International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22, 2018.
measurement science is needed, including performance metrics, use case scenarios, test methods, reference datasets, and software tools, to promote unbiased assessment of robot system accuracy degradation and to verify and validate health assessment strategies. Health monitoring, diagnostics, prognostics, and maintenance, which is collectively known as Prognostics and Health Management (PHM), have gained more and more attention within the robot system domain. The objective of PHM for robotics is to maintain the performance of robot systems by providing intelligence to enhance maintenance and control strategies. Robot systems within manufacturing environments can benefit from PHM where PHM technologies can reduce unscheduled downtime and costs.

Degradation monitoring, diagnosis, and prognosis at a system's highest level are defined as top level PHM. The same efforts at the component level are defined as component level PHM [15]. The top level PHM is difficult due to the great complexity and from the interactions among multiple subsystems and components that comprise the system. A considerable body of knowledge has been accumulated on component level PHM. Researches are conducted in the development of reasoning algorithms and in establishing failure precursors for components [15-18]. Research approaches for solving PHM problems (for both top level PHM and component level PHM) are typically either physics-based or data-driven; while a hybrid combination also exists, it is but usually dominated by one of them [8, 19]. Physics-based approaches typically involve building technically comprehensive mathematical models to describe the physics of a system and its failure modes. For most industrial applications, physics-based approaches might not be the most practical solutions, especially for component level PHM, since the fault type in question is often unique from one component to another. It might also be hard to identify the fault without interrupting operations (e.g., needing special instruments to measure a component's performance or health that may interrupt the production operations) [20]. Data-driven approaches attempt to derive models directly from routinely-collected condition monitoring data instead of building models based on comprehensive system physics and human expertise. They are built using historical information and produce prediction outputs directly in terms of condition monitoring (CM) data [21]. Data-driven approaches may be the more available solution in many practical cases since it is easier to gather data than to build accurate system physics models [20]. However, data-driven approaches raise the questions of what data to collect, when to collect, how to collect (what sensors to use), and how to quantify the impact of degradation on the output of the robot work cell (process or final products).

One research effort at NIST is the Prognostics, Health Management, and Control (PHMC) project, which is developing the measurement science to promote advanced monitoring, diagnostic, and prognostic strategies within the manufacturing domain [22]. Part of the research effort for robot system PHM,

from the PHMC project, is presented in this paper. The Research Background and Approach section presents the development of the PHMC for robot systems structure and its four-layer data sensing. The Integration of Top Level and Component level PHM section presents the development of using top level PHM to directly detect tool center accuracy degradation, and using component level PHM for deep data analysis (including root cause analysis) and PHM solution development. The Reference Data Collection and Analysis section presents the data sets collected on a Universal Robot¹ (UR5) to measure the degradation of the robot tool center position accuracy. The Conclusion wraps up the paper and highlights future work.

RESEARCH BACKGROUND AND APPROACH

PHMC for robot systems research is being conducted at NIST. The efforts include the development of performance metrics, test methods, reference datasets, and supporting tools to help the manufacturing community enhancing their PHM capabilities [12]. The research effort includes: 1) identifying data and information needed to make an informed decision with respect to robot systems setting and updating control points: 2) determining appropriate structure, organization, and analysis of data to gain actionable intelligence for robot systems; 3) enabling feedback of intelligence through the robot system to update control for optimal production; and 4) developing use cases to implement PHMC for robot systems in industrial applications. The goal is to deliver vendor-neutral approaches and plug-andplay solutions to improve decision-making support and automation [13].

Structure for the Robot Performance Degradation PHM

As a subset of the research, a structure and the key building elements are developed for robot performance degradation PHM (including robot tool center accuracy degradation) as shown in Fig. 1. The first key building element is an advanced sensing module shown in the upper left of Fig. 1. PHM efforts cannot be performed without proper sensing to understand the status of the system and components. The advanced sensing module provides important inputs for the research to promote monitoring, diagnosing, and predicting the system's health. This module has four layers of sensing that will be detailed in the next subsection. The second key building element is the data processing module as shown in the second upper left of Fig. 1. This module centers on reference algorithms development to fuse data. Data are captured from multiple sensors that are employed in the advanced sensing module. The third key building element is the development of algorithms for robot system health assessment. which is collectively shown in the cost function, degradation, prognostic, and visualization tools modules in Fig. 1. The purpose is to develop algorithms for robot system health assessment. The cost function works on solving robot error model to identify the possible root causes of faults and failures.

¹ https://www.universal-robots.com/

Qiao, Guixiu; Weiss, Brian. "Monitoring, Diagnostics, and Prognostics for Robot Tool Center Accuracy Degradation." Paper presented at ASME International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22, 2018.



Figure 1. Structure and the key building elements for robot performance degradation PHM

The fourth key building element presents a closed-loop implementation of the PHM solution within the control, which is shown as the action module and the PHM remedy module in Fig. 1.

Expanded from the first key building element, the advanced sensing module consists of four layers of data sensing and analysis as shown in the callout picture at the bottom of Fig. 1. These layers are: the controller layer sensing and analysis, the environmental layer sensing and analysis, the add-on layer sensing and analysis, and the upper layer sensing and analysis. From the upper layer to the controller layer, information is becoming more granular by detailing sensing information in specific elements of the system.

Details for the Advance Sensing Module

Upper layer sensing and analysis aim to actively assess the health of the overall system by taking into account the system architecture, system functions, and process-related parameters [23]. For upper layer accuracy degradation analysis, integrated

sensors are needed to efficiently assess the overall system's health degradation. Using multiple 1-D (one dimensional) or 2-D (two dimensional) sensors should be avoided since the setup is complex and will introduce error stacking. The research approach at this layer emphasizes the development of advanced sensing and test methods (including models and analysis algorithms) that can quickly and efficiently assess the tool center accuracy degradation.

Add-on layer sensing and analysis are developed to collect pre-designed features, e.g., using force and torque sensors to understand the influence of payload and the unbalanced tool mounting, from the targeted sub-systems. The add-on sensing promotes the involvement of additional sensors for additional information that the controller and upper layers may not provide. The research at this layer emphasizes the key subsystem/module (e.g., motor within a robot arm) identification and suitable sensing methodology selection. The design of the add-on system needs to be easily integrated into the system's controller(s) without complex interfaces or wirings.

Qiao, Guixiu; Weiss, Brian. "Monitoring, Diagnostics, and Prognostics for Robot Tool Center Accuracy Degradation." Paper presented at ASME International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018

- June 22, 2018

The environmental layer sensing and analysis are developed to collect information about environmental conditions and settings when a robot is performing a task. Information includes design data (e.g., the program that a robot is running), process data, system integration control data, and external programmable logic controller (PLC) data. The environmental layer sensing and analysis can help to clarify the operational settings of the robot (e.g., speed of the robot, payload changes, etc.) when an anomaly is detected (by the upper layer sensing and analysis), or the parameters of an on-going robot operation when a dataset is collected from a controller. There are still challenges of how to integrate and align the environmental layer data with the controller layer data and the add-on layer data for deeper data analysis.

Controller layer sensing and analysis extracts data, for example, commanded and actual joint positions, commanded and actual speed, joint current, etc., from the robot controllers and/or embedded sensors. The controller layer sensing is not the direct measurement of the tool center accuracy degradation, but can highlight issues in the system through data analysis. NIST's research at this layer is to develop methods and algorithms for deep data analysis, including root cause analysis.

After the four-layer sensing and analysis are structured, a strategy is developed to use the structure to support the monitoring, diagnostics, and prognostics for robot tool center accuracy degradation. The next section will present the idea of integrating the top level PHM and the component level PHM by utilizing the four-layer sensing and analysis.

INTEGRATION OF TOP LEVEL AND COMPONENT LEVEL PHM

The integration of the top level PHM and the component level PHM is adopted for this robot tool center accuracy monitoring, diagnostic, and prognostic research. Instead of using stacks of data from components to quantify the robot tool center accuracy degradation, which may miss some influencing components since the measurements are indirect, the tool center accuracy degradation is directly measured from the upper layer sensing and analysis. The measured deviation errors can be directly used as a comparison against the robot's task specification and tolerance to aid in the prediction of faults and failures. Top level PHM can quickly detect problems and give the robot tool center a quick health assessment if the condition of environmental conditions changes, or reconfigurations occur for the work cell, or manufacturers need to make sure the robot has not experienced a degradation when an important part is put in the work cell. At the same time, the component level PHM is needed because once a problem is detected from the top level PHM, the root cause needs to be found, and remedies can be applied to the problematic components by understanding the components' health status, e.g., a calibration needs to be performed (components are in good condition, but the mechanical relationship between them needs to be recalibrated) or a gearbox needs to be changed (components have some failure).

To support the integration of top level PHM and the component level PHM, the four-layer sensing and analysis have the following functionalities:

- Upper layer sensing and analysis: perform quick health assessment
- Add-on layer sensing and analysis: provide extra information that is missing from other layers of sensing
- Environmental layer sensing and analysis: identify key environmental information needed for PHM purpose
- Controller layer sensing and analysis: collect/monitor robot controller data.

Top Level PHM - the Quick Health Assessment Methodology Supported by Upper Layer Sensing and Analysis

The upper layer sensing and analysis support the top level PHM through the development of the quick health assessment methodology [24]. The quick health assessment methodology assesses the robot tool center position and orientation accuracy degradation. Developments of the quick health assessment include: 1) advanced sensing to measure the tool center position and orientation; 2) test methods and model to assess the health status of the full robot working volume using limited measurements; and 3) algorithms to solve the test method model, which handles the geometric and non-geometric robot errors, and the uncertainties of the measurement system [24].

The advanced sensing developed at NIST is a 7-D (seven dimensional - time, X, Y, Z, roll, pitch, and yaw) measurement system, which includes a vision-based measurement instrument and a special target (under consideration for a patent). The developed advanced sensing system can quickly acquire the position and orientation information of a robot tool center accuracy [24]. Existing position and orientation measurement technologies include laser tracker systems and optical tracking systems [10, 25]. These systems are expensive. The laser tracking systems need to keep the line-of-sight of its laser beam. Otherwise, the loss-and-reconnect beam processes significantly slow down the measurement speed. The beam may be easy to break if a target is mounted on the robot's end effector when the robot arm rotates. As for the condition of the optical tracking systems, optical tracking uses infrared (IR) technology. The accuracy and efficiency are influenced by ambient lights. Since objects that need to be tracked are equipped with retro-reflective markers, the images of the optical tracker's IR image sensor can only contain the markers. They cannot see the measurement objects or an environment. When ambient lights exist, the reflected lights from ambient objects or targets will be treated as real targets. There is no redundancy when applications are used under a complex industrial environment. The 7-D measurement system developed at NIST is a vision-based system. The 7-D measurement instrument uses two high-speed color cameras. The reasons to use vision-based design are because: (1) a visionbased system can measure position and orientation information simultaneously; (2) novel camera technologies enable the achievement of camera sub-pixel accuracy. The sub-pixel accuracy converts to the measurement system's high degree of accuracy after optical triangulations; and (3) camera technology

Qiao, Guixiu; Weiss, Brian. "Monitoring, Diagnostics, and Prognostics for Robot Tool Center Accuracy Degradation." Paper presented at ASME International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22, 2018.

is getting mature. A vision-based system is relatively costeffective to integrate [26]. The 7-D measurement system doesn't use infrared cameras, but selected high-speed color cameras. Redundant information from color images and advanced color image processing technologies are utilized to get more accurate target detection results. A high-performance computer will be used to perform the image processing. A special target is designed to work with the measurement instrument to measure the robot position and orientation information. Software tools are developed to perform the measurements. The 7-D information is captured with a time synchronization feature. Data synchronization is important for fusion of this data with the data from other layers to support root cause analysis [27-29].

Test methods and algorithms are developed to analyze the tool center accuracy degradation in a volumetric way (i.e., evaluate tool center errors from different directions in 3D space) because the error magnitudes and directions are different depending on the specific joint movements to achieve the desired tool center. A robot arm fixed loop motion is designed. The fixed loop motion needs to be evenly distributed in both joint space and Cartesian space within the robot working volume [24]. The even distribution in joint spaces prevents any errors from being missed or from being too heavily weighted. The even distribution in Cartesian spaces enables the evaluation of the arm accuracy and rigidity throughout the robot working volume, including near positions or fully extended positions. While the tool center is moving to these pre-programmed positions, the 7-D measurement system captures the position and orientation information of the robot tool center. Ideally, periodic data would be collected to track accuracy degradation with minimal disruptions to production. Analyzed time, position, and orientation data will be used to measure the robot tool center accuracy degradation when compared to original specifications and prior measurements.

Detail Data Analysis – Integration of Component level PHM

Once accuracy degradations are detected at the top level PHM, data from other layers are added to the analysis. The environmental layer, add-on layer, and controller layer sensing and analysis support the integration of the component level PHM to top level PHM, for deep data analysis (including root cause analysis) and proposed solutions. The environmental and add-on layer sensing and analysis provide the operational settings and system setups when an issue occurs (the environmental and addon layer sensing are not the focus of this paper, so they are minimally discussed). The controller layer sensing and analysis provide detailed component information about abnormal issues that may influence the robot's tool center accuracy. When the robot is performing the fixed loop motion for top level PHM, the controller layer data are also collected, including joint positions, joint velocities, joint current, joint temperature, etc. The controller layer sensing and analysis will focus on the section of data where the degradation conditions are detected (by the top level PHM). With known quantified deviation detected by the top level PHM, multiple problematic conditions are analyzed to find the abnormal components to identify the factors that

influence the system performance (for diagnostic purposes). These analyses can be used to build the knowledge for quantifying precursors used for the PHM prognostics purpose. During the occurrence of a fault or a failure, the combination of abnormal component features (for example, a particular pose, speed, and payload) is captured. When historical data exist, searches can be aimed using the condition under which the specific fault or failure occurred. It is more efficient for the targeted condition search, comparing with the general search of abnormal changes in the historical data, which is usually difficult without clear precursors.

Historical data can be collected before and after maintenance is performed on a robot. The top level PHM can measure and record the robot's performance changes before and after maintenance. Component level PHM data can capture the components' performance changes, thereby understanding the influences from maintenance activities (e.g., a gearbox changes). This approach can help to develop the link of the top level PHM and component level PHM to automatically capture failure precursors that can be used for prognostic analysis.

Data synchronization is important for the integration of top level PHM and component level PHM. Data are collected from different layers using different frequencies. It takes substantial effort to align unsynchronized data. A global timer is used to time-stamp all the data. In dynamic motion-related analyses, data synchronization has strict latency requirements. When robots are moving at fast speeds, the data may shift a lot even with a small amount of delay time. Controller layer data is collected at the highest allowable speed (e.g., 125 Hz for the robot used in this research). Both static and dynamic errors are captured, along with the transition errors (following error, gear cyclic errors, etc.). These data can provide insights of what from the robot's controller is influencing the robot's degradation.

A test was conducted at NIST to provide an example and create reference data sets for the robot tool center accuracy degradation monitoring, diagnostics, and prognostics using the integration of top level PHM and component level PHM.

REFERENCE DATA COLLECTION AND ANALYSIS

A data set was collected from the UR5 robot to measure the degradation of robot tool center position accuracy. For this test, a precision checkerboard (20 µm accuracy) was mounted on the robot arm's tool flange as shown in Fig. 2. The 7-D measurement system (shown in Fig 2. with two cameras mounted on a tripod) developed at NIST, was placed in front of the robot to detect and measure the center of the checkerboard. The tool center accuracy is described using six degrees of freedom (6DOF), containing position accuracy (3DOF - x, y, z) and orientation accuracy (anther 3DOF - yaw, pitch, and roll). For this test, only the tool center's accuracy (x, y, and z) was measured using the checkerboard target by the 7-D measurement system. For the orientation accuracy (yaw, pitch, and roll), extra information need to be taken. A special target was developed (under consideration for a patent) to measure both position and orientation information (this will be presented in a future publication). For this test, the tool center accuracy degradation

Qiao, Guixiu; Weiss, Brian. "Monitoring, Diagnostics, and Prognostics for Robot Tool Center Accuracy Degradation." Paper presented at ASME International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22, 2018.



Figure 2. Setup of the test in robot work cell

can indicate a degradation in the robot's accuracy health. Most of the robot operations in manufacturing need to guarantee the position accuracy (x, y, z). When position accuracy degrades, there is a possibility that some problems may exist in the orientation accuracy too.

A motion path for the robot was preprogrammed using a simulator and offline programming software. A script program was generated from the offline programming software and later copied to the robot, which can run from the robot teach pendant. A software tool was developed to communicate with the robot controller. The 7-D measurement system software and the communication software tool used the same global time stamp allowing the tool center measurement and the controller data to be automatically aligned. When the script program started running, the robot moved along the preprogrammed path. At the same time, the controller data was captured at 125 Hz and saved to a file at the conclusion of the robot's motion. A software tool was developed to use the collected, actual joint data from the controller to playback the robot motion in the simulator software. The change of joint angles was monitored within this software tool. When a joint angle change was bigger than one degree (e.g., user-defined criteria), a new "teach point" in the simulator was created. All of the teaching points created the playback path of the robot motion (as shown in Fig.3). Other factors (e.g., time), can also be used as the criteria to create "teach points." For example, a "teach point" was created every 0.2 seconds. This kind of playback drew 3D paths of the robot's real physical motion (as shown in Fig. 3). The playback can be saved and replayed like a simulation program, but it reflects the real robot motions in its actual condition. Different from a video recording, the playback can be viewed from different angles and can call out detailed position and orientation information. The playback



Figure 3. Playback of robot motion

can be used as a user-friendly tool to help users understand the robot pre-programmed motion. Being user-friendly was an important requirement of the PHM tools developed at NIST to transfer the technology to the manufacturing community.

In this test, the 7-D system measured the tool center's x, y, and z positions using the checkerboard target. Measurements were taken when the robot arrives at a waypoint and remains stationary (dynamic measurements will be performed using the special target in future work). Deviations were calculated from the measured positions to the nominal positions. At the same time, controller data was collected. The controller data contained time information (the time elapsed since the controller was started - a global timestamp replaces the controller time when data collection starts), target joint positions, actual joint positions, target joint velocities, actual joint velocities, etc. Details are shown in Table 1.

	Table	1	Controller	data	sets
--	-------	---	------------	------	------

Meaning	Туре	Number	Size in	Notes
ROBOT_TIME	double	1	8	Time elapsed since the controller was started
ROBOT_TARGET_JOINT_POSITIONS	double	6	48	Target joint positions
ROBOT_ACTUAL_JOINT_POSITIONS	double	6	48	Actual joint positions
ROBOT_TARGET_JOINT_VELOCITIES	double	6	48	Target joint velocities
ROBOT_ACTUAL_JOINT_VELOCITIES	double	6	48	Actual joint velocities
ROBOT_TARGET_JOINT_CURRENT	double	6	48	Target joint currents
ROBOT_ACTUAL_JOINT_CURRENT	double	6	48	Actual joint currents
ROBOT_TARGET_JOINT_ACCELERATIONS	double	6	48	Target joint accelerations
ROBOT_TARGET_JOINT_TORQUES	double	6	48	Target joint torques
ROBOT_JOINT_CONTROL_CURRENT	double	6	48	Joint control currents
ROBOT_CARTESIAN_COORD_TOOL	double	6	48	Actual Cartesian coordinates of the tool: (x,y,z,rx,ry,rz), where rx, ry and rz is a rotation vector representation of the tool orientation
ROBOT_TCP_FORCE	double	6	48	Generalised forces in the TCP
ROBOT_JOINT_TEMP	double	6	48	Temperature of each joint in degrees celsius

For the test, the same program ran with different conditions to understand the influences of position degradation from temperature, speed, and payload. A one-second motion halt was added to the script program at the waypoint positions. Since the controller retrieved data at 125 Hz, the one-second motion halt contained the continuous data stream of 125 samples. It is important to observe how soon the robot arm could stop

Qiao, Guixiu; Weiss, Brian. "Monitoring, Diagnostics, and Prognostics for Robot Tool Center Accuracy Degradation." Paper presented at ASME International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018

completely at a waypoint. The 125 data-sample during this time showed the dynamic performance of the joint motor, especially under the influences from different speeds, temperatures, payloads, and the tool mounting configurations. Fig. 4 shows the tool center position deviation (combined changes of x, y, and z as distance deviation) calculated from the target joint position and the actual joint position during the motion halt status (all joint speeds are zero). Each cluster is one second worth of data (125 points). The x-axis of Fig. 4 is the number of points measured. The vertical axis shows the distance deviation in mm. The 7-D measurement system measured the static positions at the middle of motion halt as shown in Fig. 4 (the red dots labeled with 7-D measurement data). The deviations that the 7-D system measured are larger than the calculation error from joint kinematics (actual joint positions minus target joint positions). One reason is that the joint kinematics calculations didn't include the imperfection of joints, links, and deformations. Another reason is that the arm is still not fully settled when the 7-D system began taking measurements. For this test, the robot works at the condition of cold start, full speed (1m/s), and payload of 2 kg (maximum designed load of this platform is 5 kg). Fig. 4 shows that there are significant fluctuations of the position deviation (from 80 µm to 180 µm) and overshoots when the robot stops at waypoints. To understand why the big deviations of tool center positions occurred, speed vs. joint deviation data were studied. Fig. 5 shows the joint's deviation at the cold start, half of the full speed (the full speed is 1 m/s), and 2 kg payload condition.

The target joint positions and the actual joint positions retrieved from the controller were used to calculate the change in joint positions (as shown in the primary y-axis on the left side of Fig. 5). This reflects the condition when a robot is programmed to move to a position; the arm actually moves to a position that is slightly deviated from the commanded position. The smaller the deviation, the more accurate the robot joint moved. A small joint angle error can result in a large tool center position, and orientation deviation since the effect is enlarged with the kinematic chain of arm length. Velocity may also influence the deviation. The velocity is drawn on the secondary y-axis on the right side of Fig. 5 to show the relationship with



Figure 4. Tool center position deviation



Figure 5. Joint's velocity vs. joint position deviation

joint position deviations. The x-axis presents the number of the point measured. Among the six joints of the test robot, the fifth joint (as shown in Fig. 3) shows large vibration patterns (as shown in Fig. 5 in the ROBOT_ACTUAL_JOINT_VELOCITIES data). The correlation of the velocity and joint position deviation shows that when the robot is moving, the deviation is about four times larger as compared to when the robot is stationary (the fluctuation range is ~ 0.04 degrees while moving compared to ~ 0.01 degrees while static). Therefore, the tool center accuracy will be influenced by the fluctuation. If a robot is moving to perform material removal or a composite additive layout, any tool center position degradations will impact the part quality. Accuracy degradation during dynamic operations needs to be carefully monitored for these applications that require relatively high-precision motion. To understand the payload influences, the same test and analysis were performed after removing the payload. The phenomenon still existed. It was determined that payload was not the cause of joint vibration in this test.

Fig. 6 shows the analysis when the system was allowed a two-hour warm-up where joint 5's (J5) temperature increases by 10 degrees Celsius. Under this temperature condition, Fig. 6 shows when the robot is stationary, the deviation curve remains relatively "flat"; when the robot is moving, J5's position deviation range increases to 0.06 degrees. The higher operating



Figure 6. Temperature influence on delta joint position

Qiao, Guixiu; Weiss, Brian. "Monitoring, Diagnostics, and Prognostics for Robot Tool Center Accuracy Degradation." Paper presented at ASME International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018



Figure 7. Velocity influence on joint position deviation

temperature made the joint position degradation worse, especially when the robot was in motion. Fig. 7 presents the data when the joint's velocity is at full speed. Under the full speed condition, J5's position deviation range increased to 0.07 degrees; when the robot is stationary, the deviation curve shows "bumps" as highlighted in the red circle in Fig. 7. These "bumps" can cause tool center position jitter. This jitter would need to be eliminated if the robot were to support some high accuracy applications.

The above datasets show that a robot's tool center position errors can be enlarged by various settings. When wear and tear exist, the change and degradation become more complex. On the factory floor, debugging and making parameter adjustment for robot programming/teaching is time-consuming, especially when new tasks are being programmed or changes are made to existing programs. Some task failures of low reliability (unexpected failures occur, although for most of the time the robot works fine) are due to the robot motion being influenced by the combinations of changes (speed, payload, unbalance tool mounting, etc.). In addition, robots of different models and from different manufacturers may have different levels of reliability.

CONCLUSION

Accuracy degradation impacts a robot's performance. In this paper, the NIST's development of measurement science to support the PHM for robot system performance degradation is presented. PHM outputs are expected to reduce unexpected downtime, improve productivity, efficiency, quality, and optimize maintenance strategy. The robot performance degradation PHM structure and its four-layer data sensing were constructed. Innovative approaches for advanced sensing (the 7-D measurement system and special target), test methods, and the fusion of multiple layers' sensing analysis were used. The integration of the top level PHM and the component level PHM was promoted. Data sets are collected to provide an example for robot degradation PHM. The next steps will continue the efforts of developing and the implementing the special target (used by the 7-D measurement system) to measure both position and orientation degradation in static or dynamic robot's conditions. Additional use cases will be developed for further applications. Additionally, datasets will explore what can be accomplished in terms of generating actionable diagnostic and prognostic intelligence.

NIST DISCLAIMER

Certain commercial entities, equipment, or materials may be identified in this document in order to illustrate a point or concept. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

REFERENCES

- [1] J. S. Manyika, R. Dobbs, G. Strube, "Manufacturing the future: The next era of global growth and innovation," McKinsev Global Institute, 2012.
- [2] J. Rossmann, "eRobotics meets the Internet of Things Modern Tools for Today's Challenges in Robotics and Automation," in Proceedings 2015 International Conference on Developments in Esystems Engineering Dese 2015, ed, 2015, pp. 318-323.
- [3] R. DeVlieg, "Expanding the use of robotics in airframe assembly via accurate robot technology," SAE Int. J. Aerospace, vol. 3, pp. 198-203, 2010.
- [4] T. M. Anandan, "Aerospace Manufacturing on Board with Robots," in https://www.robotics.org/contentdetail.cfm/Industrial-Robotics-Industry-Insights/Aerospace-Manufacturing-on-Board-with-Robots/content id/5960, ed: RIA, 2016.
- A. Murphy, "Industrial: Robotics Outlook 2025," in [5] http://loupventures.com/industrial-robotics-outlook-2025/, ed: Loup ventures, June 2017.
- [6] L. Lin, Y. Song, Y. Yang, H. Feng, Y. Cheng, and H. Pan, "Computer vision system R&D for EAST Articulated Maintenance Arm robot," Fusion Engineering and Design, vol. 100, pp. 254-259, 2015.
- [7] Z. Pan, J. Polden, N. Larkin, S. Van Duin, and J. Norrish, "Recent progress on programming methods for industrial robots," Robotics and Computer-Integrated Manufacturing, vol. 28, pp. 87-94, 2012.
- K. L. Tsui, N. Chen, Q. Zhou, Y. Hai, and W. Wang, [8] "Prognostics and Health Management: A Review on Data-Driven Approaches," Mathematical Problems in Engineering, vol. 2015, pp. 1-17, 2015.
- [9] J. A. Marvel, J. Falco, and I. Marstio, "Characterizing Task-Based Human-Robot Collaboration Safety in Manufacturing," IEEE Transactions on Systems Man Cybernetics-Systems, vol. 45, pp. 260-275, Feb 2015.
- [10] B. Greenway, "Robot accuracy," Industrial Robot, vol. 27, pp. 257-265, 2000.
- [11] C. Feng, Y. Xiao, A. Willette, W. McGee, and V. R. Kamat, "Vision guided autonomous robotic assembly and as-built scanning on unstructured construction sites," Automation in Construction, vol. 59, pp. 128-138, 2015.

- June 22, 2018

- G. Qiao and B. A. Weiss, "Advancing Measurement [12] Science to Assess Monitoring, Diagnostics, and Prognostics for Manufacturing Robotics " International Journal of Prognostics and Health Management, vol. 7, 2016.
- [13] B. A. Weiss and G. Oiao, "Hierarchical Decomposition of a Manufacturing Work Cell to Promote Monitoring, Diagnostics, and Prognostics," in ASME 2017 12th International Manufacturing Science and Engineering Conference, Los Angeles, CA, USA, 2017.
- [14] B. A. Weiss, M. M. Helu, G. W. Vogl, and G. Qiao, "Use Case Development to Advance Monitoring, Diagnostics, and Prognostics in Manufacturing Operations " Intelligent Manufacturing Systems, vol. Austin, TX, December 5-7, 2016
- V. Sathish, S. D. Sudarsan, and S. Ramaswamy, "Event [15] Based Robot Prognostics using Principal Component Analysis," 2014 IEEE International Symposium on Software Reliability Engineering Workshops, pp. 14-17, 2014
- [16] F. Massi, N. Bouscharain, S. Milana, G. Le Jeune, Y. Maheo, and Y. Berthier, "Degradation of high loaded oscillating bearings: Numerical analysis and comparison with experimental observations," Wear, vol. 317, pp. 141-152, Sep 2014.
- J. H. Shin and J. J. Lee, "Fault detection and robust fault [17] recovery control for robot manipulators with actuator failures," Proceedings 1999 IEEE International Conference on Robotics and Automation 1999, Detroit, Michigan, pp. 861-866.
- [18] G. J. Liu and A. A. Goldenberg, "Uncertainty decomposition-based robust control of robot manipulators," IEEE Transactions on Control Systems Technology, vol. 4, pp. 384-393, Jul 1996.
- [19] Y. Peng, M. Dong, and M. J. Zuo, "Current status of machine prognostics in condition-based maintenance: a review," International Journal of Advanced Manufacturing Technology, vol. 50, pp. 297-313, Sep 2010.
- [20] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," Mechanical Systems and Signal Processing, vol. 23, pp. 724-739, 2009/04/01/ 2009.
- K. Javed, "A Robust & Reliable Data-driven [21] Prognostics Approach Based on Extreme Learning Machine and Fuzzy Clustering," Thesis, 2014.
- [22] B. A. Weiss, G. W. Vogl, M. Helu, G. Oiao, J. Pellegrino, M. Justiniano, et al., "Measurement Science for Prognostics and Health Management for Smart Manufacturing Systems: Key Findings from a Roadmapping Workshop," presented at the Annual Conference of the Prognostics and Health Management Society 2015, Coronado, CA, 2015.
- B. Abichou, A. Voisin, and B. Iung, "Bottom-up [23] capacities inference for health indicator fusion within

multi-level industrial systems," IEEE Conference on Prognostics and Health Management (PHM), pp. 1-7, 18-21 June 2012.

- [24] G. Qiao, C. Schlenoff, and B. Weiss, "Quick Positional Health Assessment for Industrial Robot Prognostics and Health Management (PHM)," in International Conference on Robotics and Automation (ICRA) Singapore, 2017, pp. 1815-1820.
- [25] R. Mautz, "Overview of current indoor positioning systems," Geodesy and Cartography, vol. 35, pp. 18-22, 2009.
- [26] M. Švaco, B. Šekoranja, F. Šuligoj, and B. Jerbić, "Calibration of an Industrial Robot Using a Stereo Vision System," Procedia Engineering, vol. 69, pp. 459-463, 2014.
- D. Mahto and A. Kumar, "Application of root cause [27] analysis in improvement of product quality and productivity," Journal of Industrial Engineering and Management, vol. 1, pp. 16-53, 2008.
- [28] S. De, A. Das, and A. Sureka, "Product failure root cause analysis during warranty analysis for integrated product design and quality improvement for early results in downturn economy," International Journal of Product Development, vol. 12, pp. 235-253, 2010.
- [29] M. H. C. Li, A. Al-Refaie, and C.-Y. Yang, "DMAIC approach to improve the capability of SMT solder printing process," IEEE Transactions on Electronics Packaging Manufacturing, vol. 31, pp. 126-133, 2008.

Qiao, Guixiu; Weiss, Brian. "Monitoring, Diagnostics, and Prognostics for Robot Tool Center Accuracy Degradation." Paper presented at ASME International Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22, 2018.

Proceedings of the 2018 Manufacturing Science and Engineering Conference **MSEC2018** June 18-22, 2018, College Station, TX, USA

MSEC2018-6617

ROBOTIC WORK CELL TEST BED TO SUPPORT MEASUREMENT SCIENCE FOR MONITORING, DIAGNOSTICS, AND PROGNOSTICS

Alexander Klinger

National Institute of Standards and Technology Gaithersburg, MD, USA

Brian A. Weiss National Institute of Standards and Technology Gaithersburg, MD, USA

ABSTRACT

Industrial robotics users, integrators, and manufacturers are implementing advanced monitoring, diagnostics, and prognostics (collectively known as Prognostics and Health Management (PHM)) techniques and technologies. PHM can take many different forms when implemented, and measures of effectiveness are highly dependent on the techniques implemented. A test bed has been built, and a use case designed, to represent common manufacturing tasks performed in robot work cells where PHM can provide greater equipment and process health intelligence. The physical and functional relationships within the work cell are mapped using a hierarchical deconstruction method to gain a better understanding of the propagation of effects of both equipment and process degradation. The test bed has been built so PHM techniques and technologies can be integrated and tested in a realistic scenario. Data is recorded for post processing and analysis for the verification and validation (V&V) of the implemented PHM techniques. The test bed will serve as a platform to develop, test, verify, and validate PHM techniques at the National Institute of Standards and Technology (NIST), and provide industry participants a standard platform for testing their PHM technologies. Having a common testing platform will provide industry a foundation for sets of tests to evaluate PHM. This paper presents the test bed and use case, the relationships therein, and the data management and collection approaches used to enable future research.

INTRODUCTION

Once a robot system is implemented in a manufacturing environment, it must maintain a level of health to meet its necessary performance targets [1-3]. As with most automation technologies, robot system effectiveness is measured using metrics tailored to the process that is being performed [4, 5]. Different metrics are used by the manufacturing community to increase awareness of a robot system's health state and the operation(s) it performs. Most metrics present current and/or historical information. If any predictive performance metrics are

used, they often do not consider future health degradations of system components [6-9].

The manufacturing community is motivated to monitor the degradation and predict the future health states of their robot systems to prevent faults and failures which ultimately impact performance [10, 11]. Advanced monitoring, diagnostics, and prognostics (collectively known as Prognostics and Health Management (PHM)) is an active area of research to support robot manufacturers, integrators, and users [7, 12-15]. There is a need to develop techniques to verify and validate (V&V) PHM technologies for robotics within manufacturing operations.

Across the manufacturing robot systems community, there is no uniform way to implement PHM. Similarly, there is no standard means of performing V&V of PHM methods as applied to robot systems. Standards do exist for some PHM methods, yet they are often specific to a piece of equipment (i.e., machine tool monitoring and diagnostics) or an application within the robotics community [12, 16-18]. The lack of broad PHM standards for industrial robot systems is in part due to the inherent complexity of robot systems and their wide use across many applications with supporting automation and peripheral equipment. There are many complex elements that interact with one another to form an industrial robot. To integrate a robot into a system adds more complexity, physical interactions, and functional relationships that influence system health. PHM can also be applied at multiple levels within a system. Examples of the various levels where PHM can be applied within a manufacturing process include components (e.g., actuators, robots), work cells, assembly lines, and factories [2, 6, 19].

The National Institute of Standards and Technology (NIST) is conducting research to develop the necessary means to V&V PHM technologies. A part of this research is focused on PHM for an industrial robot work cell. A robot work cell is defined as a system, including robot(s), controller(s), safety equipment, and other peripherals [20]. Parts typically flow into a work cell where they are manipulated and/or transformed before leaving the work cell. Common types of industrial robot work cells include material handling (e.g., machine tending, sortation), dispensing (e.g., adhesive dispensing), and assembly (e.g., welding).

Klinger, Alexander; Weiss, Brian. "Robotic Work Cell Test Bed to Support Measurement Science for Monitoring, Diagnostics, and Prognostics ." Paper presented at ASME Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22, 2018.

There is no widely-accepted metric that assesses the health of a robot work cell. However, overall equipment effectiveness (OEE) is a commonly-used metric that can provide some insight on the value of maintaining work cell health. OEE is derived from three critical metrics: quality, performance (or productivity), and availability [21, 22]. All manufacturers seek to maximize their OEE. Because of this, OEE is used as a starting point in selecting what data is collected in the test bed to ensure that PHM methods being verified detect degradations which impact the metrics and measures that contribute to OEE. System health can either increase OEE (under healthy conditions) or decrease OEE (under degraded conditions). Quality of part output is related to robot position accuracy in work cells where the relationship between a part and tool is critical to the operation being performed. This is visible in operations such as dispensing, welding, and deburring. In this case, the robot's end effector is the tool and works directly on the part or the part is moved by a robot around a static tool. Performance is tied to the operational speed of components within a work cell including robot joints. conveyors, other moving parts, and the signals that control eventbased movement [23, 24]. A reduction in speed has a clear impact on material handling tasks where the timing of movements is critical to maintain throughput. Availability is directly related to the health of a robot work cell. A robot that is not available is either not healthy or the robot is adversely impacted by a relationship it may have with an unhealthy element. Ideally, implementing PHM will maximize availability of a system by minimizing unplanned downtime and reducing planned downtime [25, 26].

The health of a robot work cell is dictated by the health of the components that comprise the work cell. In most work cells, there are complex relationships and interactions between components. These relationships and interactions can influence process time and quality. Part production can be influenced by almost any component of the system and generally requires every component to perform as intended to yield parts within specification. Mechanical components must be dimensionally accurate, parts must be within design tolerance, sensors must operate properly, sensor signals must be interpreted properly, and the algorithms (programming) that processes information must operate as designed. With the possibility of any component degrading, the relationships between the components must be understood.

This paper presents a test bed designed and constructed at NIST to serve as a platform to develop, test, and V&V PHM techniques and technologies. The following (second) section presents the test bed configuration and use case. The third section briefly discusses a hierarchical decomposition methodology and its preliminary application to the use case. The fourth section introduces the data management and collection approach to capture performance, process, and quality data. The fifth section presents sample data captured from the test bed. The final section concludes the paper and highlights next steps in this NIST research effort.



FIGURE 1. LAYOUT OF THE PHM TEST BED

TEST BED CONFIGURATION AND USE CASE

A test bed has been built to develop, test, verify, and validate PHM methods [27]. The test bed is configured to perform a process comprised of tasks that are similar to those found in common industrial robot applications. while making data (sensor data, robot data, etc.) available for PHM method V&V. The test bed is capable of real-time monitoring and control, and recording data for post processing and analysis. The test bed includes two industrial robots, end effectors for each robot, parts, and fixtures integrated to form a work cell (Fig. 1). The use case is a work cell that performs the following actions: receives parts, moves parts to a position to be drawn on, draws on the parts, then moves the parts to an output to be removed from the work cell. There is a single input location and a single output location. There are two available work fixtures where parts can be drawn on.

The use case process can be decomposed into three main tasks that are performed by the two industrial robots in the following order (per part): 1) the material handling robot performs a pick and place operation, moving a raw part from the input to a work fixture; 2) the drawing robot draws on the part, transforming the raw part into a completed part; 3) the material handling robot performs a pick and place operation, moving a completed part from a work fixture to the output (Fig. 2). A gravity-driven ramp inputs parts within reach of the material handling robot; the output consists of a box that parts can be dropped into where they are considered removed from the work cell. Once a part is placed on a work fixture by the material handling robot, the part is held in place by locating pins integral to the fixture. These pins mate with cylindrical locating holes on the bottom of the parts (Fig. 3). The locating pins have tapered ends to allow the part to self-center when there is a slight misalignment between part and fixture during part placement.

"Robotic Work Cell Test Bed to Support Measurement Science for Monitoring, Diagnostics, and Prognostics ." Paper presented at ASME Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22,

Klinger, Alexander; Weiss, Brian



FIGURE 2. THREE TASKS PERFORMED BY ROBOTS IN THE TEST BED: 1) MOVE PART TO THE WORK FIXTURE, 2) DRAW ON THE PART, AND 3) MOVE PART TO OUTPUT.



FIGURE 3. WORK FIXTURE WITH LOCATING PINS AND PART (BOTTOM SHOWN) WITH LOCATING HOLES.

The material handling robot is equipped with an electricallyactuated parallel finger gripper-type end effector. Tasks (including waypoints) and task frames (i.e., coordinate systems) are preprogrammed in the robot controller before task execution. The task frames are programmed for each possible pick and place location (i.e., input(s), output(s), and work fixtures). Tasks are assigned to the robot by assigning a pick frame and a place



FIGURE 4. COMPLIANT PEN HOLDING END EFFECTOR: A) NOT DRAWING (RESTING POSITION), B) DRAWING (COMPLYING TO PART)

frame. The drawing robot is equipped with a spring-loaded pen holder end effector. The spring-loaded end effector can be seen in the natural resting state (A) and in the drawing state where the spring is compressed (B) in Fig. 4. The drawing robot's task, including all waypoints, are hard-coded relative to a frame dictated by the work fixture (task frame) specified in the task assignment. This results in the same programmed motion profile taking place relative to the work fixture specified by the task assignment. The frame of each work fixture which holds the parts is programmed in the robot controller before task execution. The drawing robot's motion profiles are similar to motion profiles of glue dispensing, deburring, and other processes requiring robot trajectory accuracy.

Coordination, including task assignment, part position monitoring, and process state tracking, is managed by a programmable logic controller (PLC). The PLC tracks the status of each part by monitoring robot task events. Each robot task is composed of three distinct activities: 1) the robot moving into a 'ready' position to perform an action (moving to the start position for drawing or into position to grasp a part); 2) the robot performing the action (drawing, or picking, moving, and releasing a part); 3) the robot moving to a position clear of the part, completing the task, and enabling the robot to be ready to receive a new task. Robot tasks are assigned by transmitting the part number for the task along with pick and place location for the material handling robot or work fixture and task number (i.e., drawing profile, if multiple profiles or part types exist) for the drawing robot. For the current test bed configuration and use case, there is a single part type and single drawing task available for operations (Fig. 5).

Klinger, Alexander; Weiss, Brian. "Robotic Work Cell Test Bed to Support Measurement Science for Monitoring, Diagnostics, and Prognostics ." Paper presented at ASME Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22,



Future configuration plans include the addition of part presence detection at each work fixture, the input, output, and on the material handling robot's end effector to add greater awareness for part tracking and to provide additional data for PHM methods. Though not currently being used, the test bed is designed to accommodate a tray or batch style input and output where a tray with multiple raw parts in multiple locations is fed into the work cell. Additionally, the gravity fed input can be replaced with a motorized conveyor for a more complex input configuration.

HIERARCHICAL DECOMPOSITION

Robot work cells can be complex, posing challenges when designing, deploying, verifying, and validating PHM techniques. Ongoing robot work cell research includes the development of a hierarchical decomposition method to promote advanced monitoring, diagnostic, and prognostics within a robot work cell. The decomposition method is actively being developed and can offer insight where PHM should be implemented within the test bed and use case. The insight gained from the decomposition method will lead to increased equipment and process health awareness which should further inform the PHM evaluation process with the understanding of relationships between elements of the work cell.

The application of this hierarchical decomposition method to the use case is motivated by the desire to bring transparency to how a work cell's physical elements, functional tasks, and corresponding health metrics relate to one another to enhance PHM efficiency [28]. As noted earlier, the initial use case and test bed configuration are purposefully relatively simple as a starting point to prove the overall approach. As this effort continues to make progress, both the complexity of the use case and test bed will expand making it much more difficult to determine the influences and impacts that the health of specific physical elements have on functional tasks along with health metrics.

The overall objective of the hierarchical decomposition method is to develop a cost-effective, methodical approach to guiding the manufacturing community through the PHM design and deployment process when all possible fault and failure modes are not known ahead of time. This method is based upon the Multi-Relationship Evaluation Design framework which generates test-plan blueprints from several categories of input to quantitatively and qualitatively evaluate advanced and emerging technologies at the system, component, and capability levels [29, 30]. The hierarchical decomposition method's foundational elements are helping to manage use case and test bed complexity. At a relatively high level, the steps of the method are:

- 1. Physical Decomposition decompose the physical work cell into its constituent components, sub-components, etc. based upon the maintenance-driven boundaries
- 2. Functional Decomposition – decompose the overall process into its constituent tasks, sub-tasks, etc. based upon control strategy, modularity of the tasks and sub-tasks for reconfiguration or improvement, and what is important to monitor
- 3. Process and Task Metric Identification – identify all metrics that are captured (or should be captured when a new manufacturing system is being designed) at the overall process level and constituent task and sub-task levels. This task also includes identifying the relationships between these metrics (e.g., which lower level metrics are rolled up into higher level metrics).
- Risk Identification identify the sources of risk in the robot 4 work cell in terms of what has been known to go wrong, what could potentially go wrong, the likelihood of something going wrong, and impact of something going wrong.
- Risk Reduction explore ways to reduce the risk of work 5. cell faults and failures. Reductions can range from eliminating the potential for the fault or failure altogether (e.g., a robot is known to fail so removing the robot from the design would eliminate this failure) to incorporating preventative (scheduled) maintenance plans and being ready to perform reactive maintenance if something does fail
- 6. Data to Collect and Collection Approach identify the necessary data to be collected and the best strategy (e.g., sensor selection, sensor location) in which to do this.
- 7. Physical Element Metric Identification identify metrics and measures that can inform on physical system, component, and sub-component health and relate these metrics back to the previously-identified process metrics.
- Relationship Mapping and Quantification form the connections up-and-down and across the three (physical, functional, and metric) hierarchical dimensions to document both the influences and the impacts faults and failures can cause throughout the multiple dimensions of a work cell.

4

Klinger, Alexander; Weiss, Brian. "Robotic Work Cell Test Bed to Support Measurement Science for Monitoring, Diagnostics, and Prognostics ." Paper presented at ASME Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22,

2018



FIGURE 7. FUNCTIONAL DECOMPOSITION FOR THE USE CASE

It should be noted that the first time a user encounters Step 6, they will determine the necessary data to collect based upon the metrics identified in Step 3. Proceeding to Step 7, metrics that are solely focused on physical elements are identified (which may include metrics identified in Step 3). After Step 7, the user returns to Step 5 to determine if risk can be further reduced and Step 6 to identify additional data for collection based upon any metrics identified in Step 5 and the previous iteration of Step 7. The user proceeds to Step 8 after no additional metrics can be identified in Step 7. Step 6 is not the last step because it is possible that the desired data, and corresponding collection strategy, may be unattainable. In that case, the user is going to be limited in the metrics that can be captured and must narrow the focus to specific physical elements.

The first two steps, physical and functional decompositions, are completed for the robot work cell use case and presented at a high-level.

Figure 6 shows the physical element decomposition and begins with the PLC being physically connected to the controllers of both the material handling and drawing robots. The solid lines between the PLC and these two robot controllers indicate that this physical connection is continuous and fixed so long as the work cell is in operation. Both robot controllers are directly connected to their respective robot arms. In turn, both robot arms have a physical connection to their respective endeffectors, the gripper for the material handling robot and the drawing end effector for the drawing robot. Solid lines are used to indicate the fixed connection between these elements. Both the gripper and the drawing end effector directly interact with the parts that enter and exit the work cell. These interactions are dynamic and temporary, based upon the specific task (i.e., material handling, drawing) at hand, so these connections are represented with dashed lines. Several of the physical systems presented in Fig. 6 can be decomposed into constituent components and sub-components based upon their physical construction and what is commonly monitored in a manufacturing environment. For example, both robot arms can be decomposed into each of their six joints, J1 through J6. At minimum, each joint is comprised of numerous sub-components including a motor, a gear reduction, and an encoder. If data is available at a deeper sub-component level, diagnostic and maintenance activities can also be done at this deeper level. Additional details of this physical decomposition will be presented in future efforts.

Figure 7 presents the functional decomposition for the use case at the task and sub-task levels. The overall process is broken down into three specific tasks identical to what was presented in the Test Bed Configuration and Use Case Section of this paper. Each task (1, 2, 3) is then broken down further into three specific sub-tasks (1.1 - 1.3, 2.1 - 2.3, 3.1 - 3.3). Similar to how metrics can be captured to isolate the health of each task level, metrics can also be generated to determine the health of each sub-task. Several fault or failure conditions may warrant sub-task-level inspection to determine where, specifically in the process, errors are occurring. For example, suppose that the cycle time for the overall process is longer than the expected baseline. Upon examination at the task level, a determination could be made as to which specific task(s) are taking longer than expected. Additional examination at the sub-task level can reveal more details of the problem and likely lead to the root cause (e.g., closer inspection of an anomaly in a pick and place operation reveals that sub-task 1.2 – MH robot gripping part is taking longer than expected because the gripper is moving slower than baseline expectations. This could reveal a problem with the

"Robotic Work Cell Test Bed to Support Measurement Science for Monitoring, Diagnostics, and Prognostics ." Paper presented at ASME Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22,

gripper and not the positioning of the gripper). Depending upon the information conveyed by the metrics and the inherent complexity of a given process, it is possible that a task, sub-task, and deeper-level investigation may need to occur. Some of the sub-tasks can be broken down further (e.g., 2.2 'Drawing robot draws on part' can be broken down into granular drawing motions). This will be presented in future work.

The remaining steps of the hierarchical decomposition method will be applied to the use case in future efforts. This preliminary application of the method demonstrates some of the expected value of deconstructing the physical features of the work cell and the functional process it enables, providing a foundation on which PHM can be designed and implemented. Applying the next step, 'Process and Task Metric Identification,' requires an understanding of the data that can be captured to inform on the overall process. This leads to the development of a data management and collection approach.

DATA MANAGEMENT AND COLLECTION APPROACH

Devising an appropriate data management and collection strategy is critical to capture PHM data, generate necessary intelligence, and verify its effectiveness. Due to the reconfigurable nature and broad target applications of automation and robotics equipment, there is a large amount of data generated by equipment commonly found in robot work cells for use in PHM methods. The data being generated across automation equipment, and in the test bed, has many data types, units, and sample frequencies. The test bed uses a PLC as both a data aggregator and real-time processor for the data. The PLC is managing two categories of data: 1) external sensor (e.g., discrete presence sensor) and component-generated (e.g., robot) data that is not processed outside the components' internal processing implemented by the original equipment manufacturer. This includes robot-generated data such as joint positions and velocities, temperatures, and controller status; and 2) process-generated data based upon programming on the PLC or other user programmable devices. This data includes part status information, task assignments, and any other calculated values. All PLC data, whether from internal or external sources, is capable of being used in real-time processing within PHM methods being tested and can be recorded for post processing. Though the PLC in the test bed is configured to have access to all data published by the robot controllers in real-time or nearreal-time, the controllers do not publish all the data they generate to external devices. Data generated by the robots but unavailable to the PLC is logged (if possible) on the robot controllers and accessible for post processing and analysis.

Specific PLC-generated data that is recorded in the test bed is discussed in the following sub-sections. This internallygenerated data includes the building blocks for the OEE metric and can be split into three categories: performance, process, and quality data. Raw data generated by the robots and their controllers will be discussed when the data applications in PHM methods are presented in future work.

Performance Data

Performance measurement can provide data for PHM methods and enable verification of system performance during PHM method testing. Performance is a metric based upon the timing of events in the context of OEE. Performance can be measured at many levels of the system (e.g., the work cell, each robot task, each part). If events, which influence performance of a system, occur on time and/or at a defined rate, the system is performing to specification. The test bed is designed to record timestamps of many state changes and events throughout the execution of the use case.

Timestamps are recorded at predefined state changes and events on the PLC and robot controllers, respectively. Each robot controller and the PLC have their own internal clocks and, due to hardware limitations, the timestamps recorded on each device are relative to their own internal clocks. All timestamp data collected on the PLC is available for real-time calculations and is recorded. The timestamps collected on the robots are only available as recorded data for post-processing and analysis.

The timestamps collected on the PLC correspond to 14 part state changes (to be presented in detail in future work) throughout the processing of a part (see the 'Part State Changes' row in Fig. 8). Timestamps are recorded when PLC-monitored triggers are activated by internal processing (PLC trigger origin) or after the PLC receives an input from a robot controller (robot trigger origin) (see the 'Trigger Origin' row in Fig. 8). Records generated from PLC-originated triggers include parts entering the work cell, assignment of robot tasks, and parts leaving the work cell. PLC-originating triggers are activated by either internal algorithms or sensors which are monitored directly in the PLC Inputs/Outputs (I/O). Records generated from a robotoriginated trigger include when a robot begins operating on a part, when the task operation is complete, and when the robot has physically cleared the fixture area and is ready for a new task assignment. Robot-originating triggers are activated by PLC I/O.

In addition to the part state changes that are recorded on the PLC, for each robot task, four events which correspond to unique part state changes are recorded on the robot controllers. The material handling robot records four events for each pick and place operation. There are two pick and place operations for each part, one instance to place the part in the work fixture and another instance to remove the part from the work fixture (see 'Material Handling Robot Events' row in Fig. 8). The drawing robot records four events for the drawing task on each part. The four events correspond to the beginning and end of the three parts of a robot task discussed in the Test Bed Configuration and Use Case section (see the 'Drawing Robot Events' row in Fig. 8).

Each part to be processed is assigned a unique part number as it enters the work cell. The part number is transmitted to the robot controllers when tasks are assigned and each record on the robot includes the part number which is an element in the timestamp records. All timestamp records use the part number as a common index. Records from the individual robot controllers and the PLC can be compiled in post-processing to allow for any individual process segment performance to be calculated.



Process Data

Analysis at the work cell level requires an understanding of the variability of the process taking place to differentiate designed variability vs. degradation variability (caused by process/equipment health changes). In a variable process like the test bed use case, process data is required to provide context to performance data. For example, if a part is being moved by the material handling robot from the input to work fixture 1, it will take a different amount of time as moving a part from the input to work fixture 2 because work fixture 1 and 2 are in different physical locations and the robot moves with constant speed constraints. This timing differential will be seen in timestamps recorded based upon the movement of parts. It may also result in a different total part processing time (from part entering work cell to part leaving work cell) but that is not guaranteed depending on the process, slack time allowance, and existence of bottlenecks.

Process data collected in the test bed are the variable pieces of process information. This includes the input location (single option in the initial configuration presented in this paper), the output location (single option in the initial configuration presented in this paper), the work fixture location, the part number counted from startup, and the part type (task number for drawing robot). A summary of part-related process data being recorded is presented in Table 1. As the test bed is used and more

variability is identified and introduced, it is anticipated that more process data will be captured.

TABLE 1. RECORDED PROCESS DATA

Recorded Data	Description				
Part	Unique identifier for each part processed.				
Number	•				
Input Number	Identifies the numbered location where a				
	part entered the work cell. This number is				
	unique to a specific physical location.				
Output Number	Identifies the numbered location where a				
	part is placed to be removed from the work				
	cell. This number is unique to a specific				
	physical location.				
Work Fixture	Identifies the numbered location of the				
Number	work fixture where the part was placed for				
	the drawing task. This number is unique to				
	a specific physical location.				
Part Completed	Boolean identifier if part was completed				
	(PLC interpreted the successful				
	completion of the drawing task)				
Part Type	Part-type identifier reserved for the				
	possibility of part type variable processes.				

Klinger, Alexander; Weiss, Brian. "Robotic Work Cell Test Bed to Support Measurement Science for Monitoring, Diagnostics, and Prognostics ." Paper presented at ASME Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22,

Quality Data

Similar to performance measurement, quality measurement can provide data for PHM methods as well as enable the verification of part quality degradation, or lack thereof, throughout the testing of PHM methods. As work cell components degrade, there may be a direct impact on part quality. This is due to the mechanical allowances built into the system [27]. Quality measures are use case-specific and present differing challenges in their measurement. Many automated processes use some form of automated inspection which generates quality data that is usable in PHM method implementation and assessment.

In this use case, because there is a passive compliance device holding a pen, a degradation of the drawing robot which manifests itself as a positioning error in-line with the compliance will not impact part quality until the deviation from nominal is greater than the allowance provided by the end effector. However, if the robots' degradation manifests itself as a positioning error in the plane parallel with the parts drawing surface, the quality will be directly and immediately impacted.

Two measures of quality have been identified for the current use case:

- 1. Position of the drawing on the 2D surface
- 2. Quality of line drawn (e.g., solid continuous, discontinuous)

Two manual measurement methods have been developed. These measurements will provide data that can be used in an offline process to verify the success or failure of PHM methods designed to detect degradations that impact quality. In the test bed, inspections will be done during post processing making the data available for method verification.

To measure the position of the 2D drawing, a transparent template is mated with the completed part, covering the drawing. (see Fig. 9) The transparent template presents a copy of the drawing in the nominal position. The deviation of the completed drawing on the part to the template's nominal position can be measured at any segment of the drawing. For the task currently implemented in the use case, the drawing includes four locations where lines intersect, creating crosshairs for measurement. Alternatively, position can be gauged using a series of transparent templates which have gauging areas on them where if the drawn crosshair is within the gauging lines, the line is at least as accurate as that gauge specification. The quality of the line is manually measured by a person inspecting the completed part.

The quality measures developed will measure the final product and capture both the expected deviations from nominal that are present due to the tolerance stack of physical component position error as well as any additional deviations due to degradations. The nominal variance due to the tolerance stack will be made up of multiple possible component variances including robot repeatability, part-to-fixture mating variation, and the part's quality (is the drawing surface where it should be). For the use case in this test bed, the drawing program is hardcoded with waypoints in a fixed relationship to each other.



FIGURE 9. QUALITY MEASUREMENT TRANSPARENT GAUGE SHOWN ON AND OFF A COMPLETED PART.

Both fixture and part locations are hardcoded and do not change. Any measure of quality correlates to the repeatability of the robot or the location of the part/drawing surface (this will be further evident upon complete application of the hierarchical decomposition method). This use case was designed to limit the number of potential points of degradation that can impact completed part quality by forcing the robot to perform a repetitive task with hardcoded waypoints.

In a use case where there is dynamic programming, additional sources of error can impact part quality. For example, a vision system could be used to localize a part in a work area. One scenario could dictate the vision system transmit the part's position to the drawing robot which the robot would then use to calculate its kinematics and determine specific waypoints. It would be expected that there would be slight variance in the vision system's interpretation of the part position which would then create greater variation in the completed part. Another example is if the drawing robot did not have a spring-loaded compliance device and instead used force feedback to control pen pressure on the part, any variation in the force feedback could create variation in pen pressure and the final product quality. The domino effect of added variability poses real challenges in monitoring work cell health. It is anticipated that as the research progresses, future test bed configurations and use cases will include vision-based part localization while addressing the challenges of additional variance.

SAMPLE DATA

A sample dataset was collected on the test bed by processing a single part through all 14 state changes presented in the performance data section. Robot and process data were collected as described. In depth analysis of this data and the meaning of it is scheduled for a future publication. Fig. 10 is comprised of

Klinger, Alexander; Weiss, Brian,

"Robotic Work Cell Test Bed to Support Measurement Science for Monitoring, Diagnostics, and Prognostics ." Paper presented at ASME Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22,

2018.



PROCESS SHOWING (TOP TO BOTTOM) PART STATE CHANGES, DRAWING ROBOT JOINT POSITIONS, AND MATERIAL HANDLING ROBOT JOINT POSITIONS.

timeseries plots of a subset of the sample data collected, showing the time of each part state change (vertical lines) along with the positions of each of the six joints on both robots throughout the part process. By capturing part state changes, specific operational regimes can be isolated making it possible to analyze each segment independently and in the context to its operation. In this example each robot is stationary for a position of its time. During these stationary times, any position fluctuations in joint position may be indicative of a degradation of that joint.

This sample was collected without a second part being processed simultaneously to allow a clear view of the expected movements of each robot joint for a single part process. When multiple work fixtures and parts are active within the work cell, timeseries data segments will not be as clear to an observer due to robots performing multiple tasks on multiple parts in a variable order and frequency. The process data collected differentiates the individual segments of robot data and how they correlate to the parts and process. This allows a link to be made between all physical and functional components within the work cell during analysis.

CONCLUSION

A test bed has been built at NIST to collect work cell level data that can be used to verify PHM methods. The use case has been designed as a relatively simple process representative of articulated arm robots performing industrial operations. The data recording structure is such that raw and processed data can be accessed both in real-time and after-the-fact to implement and verify PHM methods. Events throughout the use case process have been defined and allow the recording of timestamps for use in performance measurement. Quality measures have been developed and defined. Process information has been identified and is recorded to contextualize data and the combination of these types of data can be used to inform, monitor, diagnose, and predict failure of elements found in robot work cells. The test bed has been built to serve as a platform to test and evaluate PHM methods developed both at NIST and in industry. This platform has the potential to enable the development of device-agnostic monitoring, diagnostic, and prognostic techniques to enhance manufacturing asset availability. Likewise, the platform promotes unbiased verification and validation of emerging PHM techniques to further inform industry on the capabilities of available technologies. Future work includes characterizing the variability of performance and quality metrics in the base (undegraded) use case, adding capabilities to induce degradation through physical manipulation of components and/or simulation of known degradations, and further applying the hierarchical decomposition method to produce a greater understanding of the physical, functional, and metric relationships within the work cell.

NIST DISCLAIMER

The views and opinions expressed herein do not necessarily state or reflect those of NIST. Certain commercial entities, equipment, or materials may be identified in this document to illustrate a point or concept. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

ACKNOWLEDGMENTS

We would like to acknowledge Joshua Chin for his contributions to preliminary test method development.

REFERENCES

[1] Helu, M., and Weiss, B. A., "The current state of sensing, health management, and control for small-to-medium-szed manufacturers," Proc. ASME 2016 Manufacturing Science and Engineering Conference, MSEC2016.

[2] Jin, X., Siegel, D., Weiss, B. A., Gamel, E., Wang, W., Lee, J., and Ni, J., 2016, "The present status and future growth of maintenance in US manufacturing: results from a pilot survey," Manufacturing Rev., 3, p. 10.

[3] Yamada, A., and Takata, S., 2002, "Reliability Improvement of Industrial Robots by Optimizing Operation Plans Based on Deterioration Evaluation," CIRP Annals - Manufacturing Technology, 51(1), pp. 319-322.

[4] Paes, K., Dewulf, W., Elst, K. V., Kellens, K., and Slaets, P., 2014, "Energy Efficient Trajectories for an Industrial ABB Robot," Procedia CIRP, 15, pp. 105-110.

[5] Muchiri, P., Pintelon, L., Gelders, L., and Martin, H., 2011, "Development of maintenance function performance

Klinger, Alexander; Weiss, Brian.

2018.

measurement framework and indicators," International Journal of Production Economics, 131(1), pp. 295-302.

[6] Jin, X., Weiss, B. A., Siegel, D., and Lee, J., 2016, "Present Status and Future Growth of Advanced Maintenance Technology and Strategy in US Manufacturing," International Journal of Prognostics and Health Management, 7 (Special Issue on Smart Manufacturing PHM)(Sp5), p. 18.

[7] Weiss, B. A., Vogl, G. W., Helu, M., Qiao, G., Pellegrino, J., Justiniano, M., and Raghunathan, A., 2015, "Measurement Science for Prognostics and Health Management for Smart Manufacturing Systems: Key Findings from a Roadmapping Workshop," Annual Conference of the Prognostics and Health Management Society 2015, P. Society, ed., PHM Society, Coronado, CA, p. 11.

[8] Melnyk, S. A., Stewart, D. M., and Swink, M., 2004, "Metrics and performance measurement in operations management: dealing with the metrics maze," Journal of Operations Management, 22(3), pp. 209-218.

[9] Ghalayini, A.M., and Noble, J.S., 1996, "The changing basis of performance measurement," International Journal of Operations & Production Management, 16(8), pp. 63-80.

[10] Chang, H. T., Huang, P. F., Wang, M., Lu, Z. Y., and Ieee, 2014, "Locked-joint Failure Identification for Free-floating Space Robots," 2014 Ieee International Conference on Information and Automation (Icia), pp. 170-175.

[11] Kahan, T., Bukchin, Y., Menassa, R., and Ben-Gal, I., 2009, "Backup strategy for robots' failures in an automotive assembly system," International Journal of Production Economics, 120(2), pp. 315-326.

[12] Vogl, G. W., Weiss, B. A., and Donmez, M. A., 2014, "Standards Related to Prognostics and Health Management (PHM) for Manufacturing," No. NISTIR 8012, National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, USA.

[13] Vogl, G. W., Weiss, B. A., and Helu, M., 2016, "A review of diagnostic and prognostic capabilities and best practices for manufacturing," Journal of Intelligent Manufacturing.

[14] Weiss, B. A., Helu, M., Vogl, G. W., and Qiao, G., 2016, "Use Case Development to Advanced Monitoring, Diagnostics, and Prognostics in Manufacturing Operations," IMS2016 -Intelligent Manufacturing SystemsAustin, Texas, p. 6.

[15] Qiao, G., and Weiss, B. A., 2017, "Accuracy Degradation Analysis for Industrial Robot Systems," ASME International Manufacturing Science and Engineering Conference, ASME, Los Angeles, California, p. 9.

[16] Vogl, G. W., Weiss, B. A., and Donmez, M. A., "Standards for prognostics and health management (PHM) techniques within manufacturing operations," Proc. Annual Conference of the Prognostics and Health Management Society 2014.

[17] Zhou, Y., Bo, J., and Wei, T., 2013, "A review of current prognostics and health management system related standards," Chemical Engineering Transactions, 33, pp. 277-282.

[18] Jeswiet, J., and Helferty, R., 1996, "Measuring robot repeatability an application of ISO and ANSI standards," Advanced Robotics, 10(5), pp. 503-520.

[19] Lee, J., Ghaffari, M., and Elmeligy, S., 2011, "Selfmaintenance and engineering immune systems: Towards smarter machines and manufacturing systems," Annual Reviews in Control, 35(1), pp. 111-122.

[20] RobotWorx, 2017, "What is a robot cell?," https://www.robots.com/fag/show/what-is-a-robot-cell.

[21] International Organization for Standardization, 2013, "ISO/WD 22400-2 - Manufacturing operations management -Key performance indicators - Part 2: Definitions and

descriptions of KPIs." [22] Liker, J. K., 2014, "Is OEE a Useful Key Performance

Indicator?," Industry Week. [23] Malinowski, M., Beling, P., Haimes, Y., LaViers, A.,

Marvel, J., and Weiss, B., 2015, "System Interdependency Modeling in the Design of Prognostic and Health Management Systems in Smart Manufacturing," Annual Conference of the Prognostics and Health Management Society 2015, PHM Society, Coronado, CA, p. 12.

[24] Mahr, C., 2000, "Beyond the standard cycle: increasing total work cell productivity by optimizing robot placement and other key performance influencers," Industrial Robot: An International Journal, 27(5), pp. 334-337.

[25] Ahmad, R., and Kamaruddin, S., 2012, "An overview of time-based and condition-based maintenance in industrial application," Computers & Industrial Engineering, 63(1), pp. 135-149.

[26] Moyne, J., Iskandar, J., Hawkins, P., Furest, A., Pollard, B., Walker, T., and Stark, D., "Deploying an equipment health monitoring dashboard and assessing predictive maintenance," Proc. Advanced Semiconductor Manufacturing Conference (ASMC), 2013 24th Annual SEMI, IEEE, pp. 105-110.

[27] Weiss, B. A., and Klinger, A. S., "Identification of Industrial Robot Arm Work Cell Use Cases and a Test Bed to Promote Monitoring, Diagnostic, and Prognostic Technologies," Proc. 2017 Annual Conference of the Prognostics and Health Management (PHM) Society, PHM Society, p. 9.

[28] Weiss, B. A., and Qiao, G., 2017, "Hierarchical Decomposition of a Manufacturing Work Cell to Promote Monitoring, Diagnostics, and Prognostics," ASME 2017 International Manufacturing Science and Engineering Conference (MSEC2017), ASME, Los Angeles, California, p. 11.

[29] Weiss, B. A., and Schmidt, L. C., "The multi-relationship evaluation design framework: creating evaluation blueprints to assess advanced and intelligent technologies," Proc. Proceedings of the 10th Performance Metrics for Intelligent Systems Workshop, ACM, pp. 136-145.

[30] Weiss, B. A., Schmidt, L. C., Scott, H. A., and Schlenoff, C. I., "The Multi-Relationship Evaluation Design framework: Designing testing plans to comprehensively assess advanced and intelligent technologies," Proc. ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, pp. 603-616.

Klinger, Alexander; Weiss, Brian. "Robotic Work Cell Test Bed to Support Measurement Science for Monitoring, Diagnostics, and Prognostics ." Paper presented at ASME Manufacturing Science and Engineering Conference, College Station, TX, United States. June 18, 2018 - June 22,

46th SME North American Manufacturing Research Conference, NAMRC 46, Texas, USA

Experimental Investigation of Residual Stress and its Impact on Machining in Hybrid Additive/Subtractive Manufacturing

Jarred C. Heigel*, Thien Q. Phan, Jason C. Fox, Thomas H. Gnaupel-Herold

National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA

* Corresponding author. Tel.: +1-301-975-3620 E-mail address: jarred.heigel@nist.gov

Abstract

Hybrid manufacturing leverages the advantages of both additive manufacturing (AM) with machining to create parts that have complex geometries, tight tolerances, and good surface finish. However, the residual stresses induced by both processes present a challenge. These high stresses can induce significant distortion that can damage the part during the additive process or lead to unanticipated distortion during the machining process. The current work investigates the impact of the residual stresses generated during the AM process on the part distortion during the subsequent machining operation. Laser powder bed fusion is used to create stainless steel cylinders and their geometries are measured before and after machining their outer diameter, allowing the resulting distortion to be calculated. The measured distortion is counter to what is anticipated based on the trends exhibited in the residual stress in the cylinders measured before the machining operation using neutron diffraction. This suggests that the machining process imposes significant stresses that, in this case, counter act the remaining stresses from the AM process and/or trigger a material phase change.

© 2018 The Authors. Published by Elsevier B.V. Peer-review under responsibility of the scientific committee of the 4th International Conference on System-Integrated Intelligence.

Keywords: Hybrid manufacturing, additive manufacturing, residual stress, distortion

Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

1. Introduction

Hybrid additive/subtractive manufacturing is a process that combines both additive manufacturing (AM) and subtractive manufacturing, such as machining, to create parts with high complexity, tight tolerances, and good surface finish [1]. This processing can be done in-envelope, where both AM and machining operations are performed by the same machine [2], or out-of-envelope, where two separate machines are used for each process [3], requiring the part to be moved between each machine. Whichever hybrid approach is used, the AM processing enables complex geometries to be created while the machining creates better geometric accuracy and surface finish than is possible solely using AM [4]. A key difference between the two approaches however, is the ability to manage the residual stresses from the AM process. While the out-of-envelope approach can easily incorporate heat treatments to alleviate the residual stress before machining, the in-envelope approach prohibits intermediate heat treatments. This can be a challenge since the residual stress generated in AM is significant.

AM generates very high residual stresses due to the thermal history imposed by the process [5]. These stresses can cause the part to crack and separate from the build plate or to distort significantly, causing the part to fail or not meet the design requirements. For instance, Denlinger et al. demonstrated through experiment and simulation that the stresses in a 4 m long part fabricated using wire-fed electron beam directed energy deposition (DED) could experience distortion greater than 50 mm [6]. Although simulation can be used to predict the residual stress and distortion so that they can be mitigated or compensated for [7], it is unclear how subsequent machining of the highlystressed AM material impacts part distortion.

A few studies in the literature have focused on the impact of residual stresses generated during AM on the subsequent machining operations. Lane et al. created stainless steel disks with laser powder bed fusion (LPBF) that were subsequently machined using a turning operation that achieved 2-D orthogonal cutting [8]. A substantial amount of chip curl out of the cutting plane was observed, which was not observed when cutting wrought disks of the same material. This outof-plane curl was attributed to the residual stress distribution in the disks from the LPBF process, and indicates that residual stresses from the AM process can impact chip formation during machining.

Concerning the impact on geometric error and distortion, Salonitis et al. studied the impact of residual stress and distortion in cylinders created using laser cladding on the subsequent machining operation [9]. Using simplified finite element models, they found that the residual stresses could be significant enough that the part could deform during the following machining operation to the point that material does not even engage with the tool. This leads to the part not meeting the design tolerances. However, this work was purely a modeling effort with no experimental validation.

The objective of the current work is to investigate the impact of residual stress on the geometric accuracy of parts produced using hybrid additive/subtractive manufacturing. Scan strategy, heat-treatment, and machining strategy are varied to quantify their impact on distortion. Stainless steel cylinders are created using powder bed fusion (PBF), then machined and the distortion from the machining operation is measured. Cylinders are produced using two different scan strategies that result in different thermal histories. Half of the cylinders are heat-treated to relieve the residual stress, while the other half are not, capturing one of the key differences between in-envelope and out-ofenvelope hybrid strategies. The residual hoop and axial stress of select cylinders are measured using neutron diffraction and the as-produced surface is characterized using a laser confocal microscope. The inner diameter (ID) of each cylinder is machined to provide a reference surface that is measured with a coordinate measuring machine (CMM) before and after three different machining strategies are used to remove material from the outer diameter (OD). The results are then presented and discussed.

2. Experiment Setup

Figure 1 presents the experimental steps used in this study. First, cylinders are created using a LPBF process, then the inner and outer dimensions of those cylinders are measured using a CMM. Two identical sets of cylinders are created on two different substrates, one of which is used to study the asproduced residual stresses while the other is heattreated to relieve the residual stresses. The dimensions of the heat-treated cylinders are re-measured using the CMM. Four cylinders are then removed from each substrate for post-process measurements: neutron diffraction to measure lattice spacing and from which residual stress is calculated, and laser confocal microscopy to characterize the roughness and profile

of the outer surface. The ID of each cylinder remaining on the substrates is machined to create a smooth reference surface, which is re-measured using the CMM. After this, material is removed from the outer perimeter using an end-mill and three different machining strategies. The cylinders are measured a final time using the CMM and the distortion from the OD machining is calculated. While this experimental method is akin to the out-of-envelope hybrid approach. the two different substrates (heat-treated and asproduced) are intended to study the anticipated difference between in-envelope and out-of-envelope hybrid strategies. The following sections describe each of these steps in greater detail before the results are presented and discussed.

	Experiment Step	Description
1)	Two separate builds are performed to create 12 cylinders on 2 different build plates (Referred to as Builds A and B).	Section 2.1
2)	The dimensions of each cylinder on both builds are measured using a CMM.	Section 2.2
3)	Heat-treat Build A to reduce residual stress.	Section 2.3
4)	Use a CMM to remeasure the dimensions of the cylinders of Build A.	Section 2.2
5)	Remove cylinders and sections of the build plate for additional measurements.	
6)	Characterize the outer surface of select cylinders.	Section 2.4
7)	Measure the residual stress in select cylinders using neutron diffraction.	Section 2.5
8)	Machine the ID of the remaining 8 cylinders on each plate.	Section 2.6
9)	Use a CMM to remeasure cylinders.	Section 2.2
10)	Machine the OD of the cylinders.	Section 2.7
11)	Use a CMM to remeasure cylinders.	Section 2.2

Figure 1 - Processing and measurement steps.

2.1. Cylinders Created using Powder Bed Fusion

Two substrates are machined from wrought American Iron and Steel Institute (AISI) 17-4 stainless steel. Each substrate is a 100 mm square and 12.7 mm thick with countersunk holes in each corner for fixturing. A single substrate is mounted on the build surface inside an EOSint M270 LPBF machine. EOS GP1 powder [10], which is equivalent in chemistry to AISI 17-4, is used to create 12 cylinders that are 12 mm tall and have nominal inner and outer diameters of 13 mm and 16 mm, respectively, as

shown in Figure 2. The powder used in this study has been recycled multiple times, but is sieved using a 65 µm mesh to remove any large particles from previous builds. Once the cylinders are built on the first substrate, it is removed from the machine so the second substrate can be mounted and the processes repeated.



Figure 2 - A) Image acquired during the build and B) the final part. The different colors are a result of different thermal histories caused by different scan strategies.

Two different scan strategies are used to produce the 12 cylinders on each substrate, as illustrated in Figure 3. Half of the cylinders on each plate are created using the standard scan strategy, in which the laser scans back and forth in a raster pattern in predesignated stripes that rotate by 67° each layer. The cylinders processed using the default scan strategy are in the lower right portion of the plate in Figure 2B and appear darker in color. The raster scans are performed with a laser power of 195 W traveling at a speed of 1 000 mm/s with 0.1 mm between adjacent raster scan tracks (hatch width). The width of the raster is 4 mm. The inner and outer perimeter of each cylinder is

scanned using a different set of contour processing conditions before and after the raster scan.



Figure 3 - The two scan strategies used in this study. A) Default raster strategy and B) concentric circles.

The other half of the cylinders are created by scanning the laser in 15 concentric circles. The laser power, speed, and hatch width are the same as the raster scan. In addition, the inner and outer perimeters of each cylinder are scanned according to the same contouring method. However, it should be noted that an error occurred while creating this scan pattern that caused the ID to be slightly smaller ($\approx 200 \ \mu m$) than the default strategy. This scan strategy is utilized to change the cooling rate and thermal history of the cylinders. Considering the scan speed and inner and outer diameters, each scan takes 40 ms to 50 ms to complete before the next laser scan begins to melt the adjacent powder and re-heat the newly solidified track. In contrast, a single default raster scan can be no longer than 9.3 mm in these parts and would take at most 9.3 ms to complete before the material is reheated by the next adjacent scan track. Therefore, the default cylinders will experience higher temperatures for longer periods of time and will have slower cooling rates, potentially affecting the residual stress. The longer time spent at higher temperatures for the default scan strategy is evident by the darker color in those cylinders, located in the lower right portion of the substrate in Figure 2B. The darker color is a result of oxidation with the atmosphere in the LPBF processing chamber, which has an oxygen content between 0.5 % and 1.5 % during the process. Table 1 presents a summary of the PBF strategy used to create each cylinder and the subsequent processes and measurements performed on each cylinder. These processes and measurements are described in greater detail in the following sections.

Build designation	Cylinder number	PBF scan strategy R aster or <u>C</u> oncentric	Heat treated	Separated from build	Surface measured	Stress measured	OD machined (number of passes)	Distortion measured	
А	1	R	х	Х	Х				
А	3	С	Х	Х	Х				
А	10	R	Х	Х		Х			
А	12	С	Х	Х		Х			
А	2	R	Х				Х	Х	
А	4	R	Х				Х	Х	
А	5	R	Х				Х	Х	
А	6	С	Х				Х	Х	
А	7	R	Х				Х	Х	
А	8	С	Х				Х	Х	
А	9	С	Х				Х	Х	
А	11	С	Х				Х	Х	
В	1	R		Х	Х				
В	3	С		Х	Х				
В	10	R		Х		Х			
В	12	С		Х		Х			
В	2	R					Х	Х	
В	4	R					Х	Х	
В	5	R					Х	Х	
В	6	С					Х	Х	
В	7	R					Х	Х	
В	8	С					Х	Х	
В	9	С					Х	Х	
в	11	С					x	х	

Table 1 - Summary of the PBF strategy of each cylinder and the subsequent processing steps and measurements performed on each.

2.2. CMM Measurements

The inner and outer diameters of each cylinder are measured with a CMM (International Organization for Standardization (ISO) 10360-2 maximum permissible error (MPE) of 5 µm) using a 2 mm diameter probe tip. The diameters are measured at various heights in

1 mm steps, beginning at a height of 2.5 mm from the top surface of the base plate, and concluding at a height of 11.5 mm, as illustrated in Figure 4. The diameter is calculated using eight measurements spread evenly around the circumference of the cylinder using the CMM inspection software. The variance is calculated and reported.



Figure 4 - Illustration of the CMM measurement locations. The ID (blue dots) and the OD (red dots) are measured at ten vertical locations using eight measurement points, equally spaced around the circumference.

2.3. Heat-Treatment

Build A is heat treated to relieve the residual stresses that are created during the PBF process. The part is cleaned using acetone and alcohol in an ultrasonic bath before heat treatment. It is then placed in a furnace and heated in a vacuum at a rate of 5 °C/min to reach a temperature of 400 °C, where it is held for 20 min. This dwell is a necessary step to allow any existing carbonates to de-gas. Heating resumes at the same rate until a temperature of 650 °C is achieved and held for 1 h. The part is then allowed to cool in the furnace with He gas. In addition to relieving residual stress, the heat treat may also change the fraction of retained austenite, which is likely to occur in this material during PBF processing [11]. However, microstructure characterization is beyond the scope of the current study.

2.4. Characterization of the Outer Cylinder Surface

934

Two cylinders (Cylinders 1 and 3) and the plate material they are built on are removed from each of the substrates using electrical discharge machining (EDM) and their outer surfaces are characterized using a laser confocal microscope with a 10x objective. The microscope scans along the height of the cylinder adjacent to the build plate edge, allowing the side surface of the build plate to be included in the scan as a reference. The scanning is performed over an area 1.3 mm wide and 12.7 mm tall. Figure 5 presents images of Cylinders 1 and 3 from Build B with yellow rectangles indicating the approximate area scanned by the microscope.



Figure 5 - The surface is characterized over an area on the outer diameter of select samples using a laser confocal microscope

2.5. Stress Measurement using Neutron Diffraction

Neutron diffraction is used for the non-destructive evaluation of the tri-axial stresses at the National Institute of Standards and Technology Center for Neutron Research (NCNR) using the stress diffractometer [12, 13]. Measurements are made on Cylinders 10 and 12 at 18 locations in the cylinder wall, as illustrated in Figure 6. The orientation and location dependent lattice spacings are measured from the neutron diffraction angle. The cross section of the diffracted beam and the neutron beam defines the gage volume, which in this work is approximately 0.5 mm wide, 2 mm tall, and 1 mm thick. Each measurement location averages the lattice spacing measured within this gage volume. The measurements are repeated on the opposite side of the cylinder to check for consistency. By assuming the radial stress through the cylinder is 0 MPa, and the material properties of 17-4 stainless steel, the average hoop and axial stresses at each measurement location are calculated. The stresses are qualitatively verified by ensuring the axial and hoop directions are in equilibrium.



Figure 6 - Neutron diffraction measurement locations.

2.6. Inner Diameter Machining Operation

A small amount of material is removed from the inner diameter during this operation to create a smooth surface that can be reliably measured with the CMM, enabling the distortion resulting from the OD operation to be measured. The inner diameter of each cylinder remaining on the plate in Step 8 is machined based on the prior CMM measurements using a 10 mm diameter 3 flute end mill (SECO 910100R100-MEGA). The interior of the cylinder is machined from the top down using a cutting speed of 70 m/min, and a feed rate of 0.1 mm per tooth, in 64 µm steps, to a diameter of 13 mm, concluding at a height 0.5 mm above the substrate surface. Coolant is used. Although the cylinder ID is nominally 13 mm, the actual inner diameter is smaller than that, as will be shown later.

2.7. Outer Diameter Machining Operations

Figure 7 illustrates the OD machining operation. The same cutting tool from the ID machining operation is used to machine the OD of all cylinders. The location of each cut is based on location of each cylinder relative to the substrate as determined by the prior CMM measurement. The OD of each of the remaining eight cylinders is cut using climb milling at an axial depth of cut of 11.5 mm, a cutting speed of 70 m/min, and a feed rate of 0.1 mm per tooth. The tool engages each cylinder 20° offset from the +Y direction, makes one complete revolution then retracts. Coolant is used during this operation. Cylinders 2, 4, 6, and 8 are machined to an OD of 15 mm using a single pass. Cylinders 5 and 11 are machined to that diameter using two passes, with the

first taking the OD to 15.5 mm, and the second to a diameter of 15 mm. Cylinders 7 and 9 are machined to a diameter of 15.5 mm in a single pass. The variety of OD machining strategies allows the impact of the radial engagement to be investigated.



Figure 7 - The OD machining strategy for the remaining eight cylinders on each plate.

3. Results

The residual stress calculations from the neutron diffraction measurements of Cylinders 10 and 12 from Build B (as-produced) are presented in Figure 8. As shown in Figure 8A, hoop stress is primarily in tension near the outer circumference of the cylinder and primarily in compression near the inner circumference. The bottom of each cylinder is in tension because the cylinder is still attached to the substrate, which is preventing the cylinder from contracting. Qualitatively, these trends agree with the literature [9, 14]. The stress distribution is a result of the thermal processing: as the newly solidified material cools and contracts, it compresses the material below it, which in turn, resists the contraction of the new material. This explains the higher amounts of compression in the middle and tension in the top of each cylinder.



Figure 8 - Residual stress measurements of Cylinders 10 and 12 from the as-produced plate (Plate B). A) Hoop stress, B) axial stress. Error bars represent an estimated measurement uncertainty of 50 MPa

Residual stress was not calculated from the neutron diffraction measurements of Cylinders 10 and 12 from Build A (heat-treated) because the heat treatment induced a phase change in the material that created a textured (multi-phase) material, which complicates the residual stress calculation. LPBF processed 17-4 stainless steel has been shown to be primarily metastable austenite [11] that easily transforms to martensite with subsequent heat treatments or cold work. The diffraction measurements confirm the presence of a second phase in the heat-treated cylinders. Unfortunately, the reduction in residual stress from the heat-treat cannot be confirmed.

Figure 9 shows a profile taken from the center of the optical scan for one of the cylinders in this study (Cylinder 1, Build A). From this profile, the arithmetic mean roughness, Ra, and root-mean-square roughness, Rq, are calculated based on the ISO 4287 standard using a digital gaussian filter with cutoff wavelength of 0.8 mm. Figure 9A shows that the surface is very rough (Ra of 16.5 µm and Rq of 19.7 µm) and consistent with the literature [15]. However, the form

Fox, Jason; Gnaupel-Herold, Thomas; Heigel, Jarred; Phan, Thien. "Experimental Investigation of Residual Stress and its Impact on Machining in Hybrid Additive/Subtractive Manufacturing." Paper presented at 46th SME North American Manufacturing Research Conference, NAMRC 46, College Station, TX, United States. June 18, 2018 - June 22, 2018.

error can be extracted from this measurement as shown in Figure 9B and shows that the cylinder does curve inward slightly. CMM measurements of the OD and ID have been acquired, however the rough surface makes the variance greater than the measured curvature.

Figure 10 shows a cylinder after each machining step. The surface resulting from the ID machining operation (Step 8) is smooth and shiny, as shown in Figure 10A, while the outer surface remains rough. This smooth surface is easy to measure with the CMM and is used as a reference surface to quantify the amount of distortion that results from the OD machining. The surface generated by the OD machining is shown in Figure 10B.



Figure 9 - Surface profile measurement of Cylinder 1 from Plate A. A) Profile with outliers removed, B) extracted form error.



Figure 10 - Example cylinders after A) ID machining, Step 8 and B) OD machining, Step 10.

Figure 11 shows the CMM measurements of the inner diameter of each cylinder before and after OD machining. Measurements acquired before the operation are represented by dashed lines, and those acquired after are represented by solid lines. In all cases the diameter decreases after the OD machining, with the greatest decrease near the top of the cylinders. The average concentricity of the ID and OD measurements is 15 µm.



Figure 11 - ID measured before (dashed lines) and after (solid lines) the OD machining operation (Steps 9 and 11). A) Plate A, heattreated B) Plate B, as-produced. Error bars represent the diameter measurement variance.

4. Discussion

There are several notable observations from these results. The first is that the cylinders distorted in the opposite direction than expected after the OD is machined. Figure 12 shows that based on the measured residual stress distribution in the part, the OD machining removed a majority of the material in tension. Removing this material would allow the remaining material that was primarily in compression to reach a new equilibrium, which would cause the cylinder to expand. However, the diameter measurements presented in Figure 11 clearly show that the cylinders contracted. There are three hypothesis that explain the decreasing diameter:

Fox, Jason; Gnaupel-Herold, Thomas; Heigel, Jarred; Phan, Thien. "Experimental Investigation of Residual Stress and its Impact on Machining in Hybrid Additive/Subtractive Manufacturing." Paper presented at 46th SME North American Manufacturing Research Conference, NAMRC 46, College Station, TX, United States. June 18, 2018 - June 22, 2018.

SP-442

- The residual stress measurements may be 1 incorrect.
- 2. Machining may induce a phase-change in the material that affects residual stress and its equilibrium state.
- 3. The machining operation may induce compressive stresses that counteract the remaining stresses from the LPBF process, causing the cylinders to distort inward to reach a new equilibrium.

The first hypothesis seems unlikely considering the measured stress trends agree well with literature [9, 13]. The second hypothesis is far more likely, since it has been reported in the literature that the austenite in this AM material is meta-stable and can easily change to martensite through additional heat treatment or cold-work [11], both of which occur during metal cutting. The third possibility is just as likely as the second, since it is well known that metal cutting operations can induce high residual stresses in the workpiece [16].



Figure 12 - Illustration of the assumed stress distribution and the anticipated impact of removing material from the OD on the ID measurements. A) Hoop stress, B) axial stress.

The second observation from the results in this study is that the chosen processing strategies used in this study have little to no significant impact on the amount of residual stress or distortion. Figure 13 shows that for every cylinder on each plate, the inner diameter decreased by 20 µm to 30 µm (from Z = 2 mm to Z = 10 mm), regardless of the processing conditions.



Figure 13 - Change in the inner diameter caused by machining the outer diameter. A) Plate A, heat-treated, B) Plate B, as-produced. Error bars represent the variance of the diameter calculations using the eight measurement points.

The difference between the two different scan strategies clearly has no effect on the residual stress or distortion. The residual stress measurements presented in Figure 8 show no difference between the asproduced and heat-treated cylinders. Furthermore, the gray and black lines in Figure 13 lie almost perfectly on top of each other. While scan strategy has been shown in the literature to impact distortion and residual stress [5], those instances are for larger scan areas and builds. However, the results for the cylinders in this current study are consistent with the trends reported by Dunbar et al. in their study using similarly sized cylinders and a similar LPBF process [14]. Their changes in scan strategy, albeit less drastic than those in the current study, had no significant impact on the measured distortion or simulated stresses.

The machining strategies used in the current study also appear to have no significant impact on distortion. Figure 13 shows that each cylinder within a certain build experienced the same amount of distortion due

to the OD machining operation, regardless of the radial engagement (approximately 0.5 mm in the even numbered cylinders, compared to approximately 0.25 mm in the odd numbered cylinders) or the number of cutting passes used to achieve the 15 mm OD (one pass for the even numbered cylinders vs. two passes for Cylinders 5 and 7). However, the tool geometry and cutting conditions were the same for all cases, so the observations regarding the impact on machining operations are limited to this study. It is expected that different tooling or cutting conditions would have a significant impact on the stresses induced in the cylinders and the resulting distortion.

The impact of the stress-relief heat treatment is less clear. Figure 13 shows that the heat-treated cylinders on Build A experienced slightly more distortion on average compared to Build B from Z = 2.5 mm to Z = 8.5 mm, 26 µm compared to 21 µm, though this difference is within the variance of the measurement $(8 \mu m)$. However, the trends near the top of the plate are significantly different. The heat-treated cylinders in Build A (Figure 13A) exhibit consistent distortion for most of the height, except for the last measurement location at 11.5 mm, which ranges between 66 µm and 102 µm. In contrast, the distortion in the non-heattreated cylinders on Build B (Figure 13B) begin to gradually increase at a height of Z = 9.5 mm, reaching maximum values ranging from 34 µm to 49 µm at the height of Z = 11.5 mm.

These distortion results suggest that the heattreatment did alter the residual stresses in the cylinders (which unfortunately could not be confirmed using neutron diffraction due to the creation of a second phase). They also suggest that the machining process used in this study induced compressive stresses along the outer perimeter of the cylinders, supporting the second hypothesis presented earlier. If great enough, the induced machining stresses would cause the cylinders to contract despite the residual stresses generated by the LPBF process, which should have caused the cylinders to expand with the removal of the material along the outer perimeter. The distortion through the bulk of the as-produced cylinders in Build B (Figure 13B) is less than Build A because the compressive residual stresses from the LPBF process resisted the distortion induced by the OD machining operation compared to the heat-treated Build B. However, more work is required to better understand the stresses created in both AM and machining processes to fully understand these results and the

Fox, Jason; Gnaupel-Herold, Thomas; Heigel, Jarred; Phan, Thien. "Experimental Investigation of Residual Stress and its Impact on Machining in Hybrid Additive/Subtractive Manufacturing." Paper presented at 46th SME North American Manufacturing Research Conference, NAMRC 46, College Station, TX, United States. June 18,

interactions complexity of their in hybrid additive/subtractive manufacturing.

5. Conclusions

This work investigated the part distortion that arises from the interaction of the multiple processes encountered during hybrid additive/subtractive manufacturing. Specifically, AM processes inherently create parts with high residual stresses that, in hybrid manufacturing, have portions of the material removed using machining operations. As a result, the stresses in the remaining material, and those potentially introduced by the machining process, must reach a new equilibrium, causing the part to distort. Cylinders that are 12 mm tall with a 16 mm outer diameter were created with powder bed fusion. The residual stresses in select as-produced cylinders were measured using neutron diffraction and the surfaces were characterized using a laser confocal microscope. After a reference surface was created on the inner diameter, the cylinders were then machined using a 3-flute end mill and the resulting distortion was measured using a CMM

The residual stress in the as-produced cylinders measured using neutron diffraction qualitatively matches what is reported in the literature, where the outer material is primarily in tension and the inner material is primarily in compression. Based on the stress distribution in the cylinder, it was assumed that they would expand when a portion of the outer material was removed through machining, since the remaining material that was in compression would reach a new equilibrium. However, measurements revealed that the cylinders contracted during the outer diameter machining operation, suggesting that the chosen tooling and cutting conditions used in this study induced compressive stresses along the outer surface that counter-acted those remaining from the LPBF process. It is also possible that the machining operation induced a phase transformation of the metastable austenite, changing the stress balance. However, additional work is required to confirm these hypotheses and to better understand the complex interaction between additive and subtractive manufacturing processes.

Acknowledgements

The authors would like to thank the following individuals for their assistance. Kil-Won Moon performed the heat treatments, Jared Tarr performed the EDM work, Jay Nanninga performed the CMM measurements, and Tyler Gervasio and Rick Lake were instrumental in the machining operations.

References

- [1] J. M. Flynn, A. Shokrani, S. T. Newman, and V. Dhokia, Hybrid additive and subtractive machine tools Research and industrial developments. International Journal of Machine Tools and Manufacture, 101, (2016) 79-101.
- [2] J. Jones, P. McNutt, R. Tosi, C. Perry, and D. Wimpenny, Remanufacture of turbine blades by laser cladding, machining and in-process scanning in a single machine, Proceedings of the 23rd Annual International Solid Freeform Fabrication Symposium, (2012) 821-827
- [3] G. Manogharan, R. Wysk, O. Harrysson, and R. Aman, AIMS-A Metal Additive-hybrid Manufacturing System: System Architecture and Attributes, Procedia Manufacturing, 1, (2015) 273-286.
- [4] F.-W. Goudsmit, G. Oosterhuis, E. Sluiter, and T. Peijnenburg, Typical requirements for precision mechanical parts and opportunities for additive manufacturing, Proceedings of the ASPE Topical Meeting: Dimensional Accuracy and Surface Finish in Additive Manufacturing, Berkeley, CA, (2014) 51-55.
- [5] P. Mercelis and J.-P. Kruth, Residual stresses in selective laser sintering and selective laser melting, Rapid Prototyping Journal, 12(5), (2006) 254-265.
- [6] E. R. Denlinger, J. Irwin, and P. Michaleris, Thermomechanical Modeling of Additive Manufacturing Large Parts, Journal of Manufacturing Science and Enginering, 136(6), (2014) 61007-61007.
- [7] E. R. Denlinger and P. Michaleris, Mitigation of distortion in large additive manufacturing parts, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 231(6), (2017) 983-993.
- [8] B. Lane, S. Moylan, and E. P. Whitenton, Post-process machining of additive manufactured stainless steel, Proceedings of the 2015 ASPE Spring Topical Meeting: Achieving Precision Tolerances Additive in Manufacturing, Raleigh, NC, (2015) 27-29.
- [9] K. Salonitis, L. D'Alvise, B. Schoinochoritis, and D. Chantzis, Additive manufacturing and post-processing simulation: laser cladding followed by high speed machining, The International Journal of Advanced Manufacturing Technology, 85(9-12), (2015) 2401-2411
- [10] R. Stirling, EOS StainlessSteel GP1 for EOSINT M 270, EOS GmbH - Electro Optical Systems, Material Data Sheet, (2009).
- [11]S. Cheruvathur, E. A. Lass, and C. E. Campbell, Additive Manufacturing of 17-4 PH Stainless Steel: Post-processing Heat Treatment to Achieve Uniform Reproducible Microstructure, JOM, 68(3), (2016) 930-942.

- [12] P. Brand, H. Prask, and T. Gnaeupel-Herold, Residual stress measurements at the NIST reactor, Physica B: Condensed Matter, 241 (1997) 1244–1245.
- [13]T. Gnäupel-Herold, J. Slotwinski, and S. Moylan, Neutron measurements of stresses in a test artifact produced by laser-based additive manufacturing, presented at the AIP Conference Proceedings, 1581, (2014) 1205–1212.
- [14]A. J. Dunbar, E. R. Denlinger, M. F. Gouge, and P. Michaleris, Experimental validation of finite element modeling for laser powder bed fusion deformation, Additive Manufacturing, 12 (2016) 108–120.
- [15]A. Triantaphyllou, C.L. Giusca, G.D. Macaulay, F. Roerig, M. Hoebel, R.K. Leach, B. Tomita, K.A. Milne, Surface texture measurement for additive manufacturing, Surface Topography: Metrology and Properties, 3(2), (2015) 24002.
- [16]J. D. Thiele, S. N. Melkote, R. A. Peascoe, and T. R. Watkins, Effect of cutting-edge geometry and workpiece hardness on surface residual stresses in finish hard turning of AISI 52100 steel, Journal of Manufacturing Science and Engineering, 122(4), (200) 642–649.

Low-GWP Refrigerants: Options and Issues

<u>Piotr A. Domanski</u>^{*(a)}, Riccardo Brignoli^(a), J. Steven Brown^(b), Andrei F. Kazakov^(c), Mark O. McLinden^(c)

^(a)National Institute of Standards and Technology, 100 Bureau Dr., Gaithersburg, MD 20899, USA
^(b)The Catholic University of America, 620 Michigan Ave, NE, Washington, DC 20064, USA
^(c)National Institute of Standards and Technology, 325 Broadway, Boulder CO 80305, USA
*Corresponding author, Tel: 1-301-975-5877; Fax: 1-301-975-8973
E-mail: piotr.domanski@nist.gov

ABSTRACT

This research explored the possibilities for refrigerants having low global warming potential (GWP) with a particular focus on high-pressure applications. The search for suitable refrigerants relied on screening of a comprehensive database using filters of environmental acceptance, chemical stability in the system, low toxicity, coefficient of performance (COP), volumetric capacity (Q_{vol}), and flammability. Among different screening steps, the performance potential (COP and Q_{vol}) of the candidate fluids was assessed using a simulation model that included both thermodynamic and transport properties, and heat exchangers with optimized refrigerant circuitry. The study showed that the low-GWP refrigerant options are very limited, particularly for fluids with volumetric capacities similar to those of R-410A and R-404A. The identified fluids with good COP and low toxicity are at least mildly flammable. Accepting the thermodynamic argument that viable refrigerants are restricted to small molecules, it is our contention that the screened database contains all suitable candidates and our screening process yielded a list of the 'best' low-GWP fluids. Refrigerant blends allow for tradeoffs between COP, volumetric capacity, flammability and GWP. The probability of finding 'ideal', better-performing low-GWP fluids is minimal.

KEYWORDS: Air conditioning, coefficient of performance, refrigeration, thermophysical properties, working fluids

1. INTRODUCTION

An amendment to the Montreal Protocol adopted in October 2016 mandates a significant reduction of the weighted value of global warming potential (GWP) of fluids used in air-conditioning (AC) and refrigeration equipment. Consequently, hydrofluorocarbon (HFC) refrigerants having a high GWP will be replaced with low-GWP fluids. Earlier research demonstrated the availability of suitable hydrofluorocolefins (HFOs) for low-pressure applications (water chillers) and medium-pressure applications (e.g., automotive AC), with medium-pressure, low-toxicity HFOs that are mildly flammable. However, prior studies did not identify low-GWP, low-toxicity, nonflammable candidates that could replace the high-pressure refrigerants R-410A or R-404A. The goal of this research was to search for suitable low-GWP refrigerants for high-pressure applications.

2. DATABASE SCREENING

To search for the most suitable low-GWP refrigerants, we used the PubChem database listing over 60 million unique chemical molecules. Our study relied on establishing filters representing different refrigerant selection criteria and applying them to these molecules contained. We accepted the thermodynamic argument that viable refrigerants are restricted to small molecules, and we considered PubChem to be exhaustive for such molecules [1].

We limited our search to molecules with 18 or fewer atoms and comprising only the elements C, H, F, Cl, Br, O, N, or S, following the observation that only a small portion of the periodic table would form compounds volatile enough to serve as refrigerants. Despite their ability to deplete stratospheric ozone, we included Cl and Br since molecules including Cl or Br might have a negligible ODP and might be acceptable if they have a very short atmospheric lifetime. These restrictions reduced the pool of molecules for further consideration to 184 000. The next two filters were GWP₁₀₀ and T_{crit} . The PubChem database does not contain these values for the vast majority of the compounds, so most had to be estimated using novel methods developed within this study based solely on the molecular structure. The prediction of GWP₁₀₀ combined estimates of the radiative efficiency and atmospheric reactivity with hydroxyl radicals. The applied filters of GWP₁₀₀ < 1000 and 320 K < T_{crit} < 420 K reduced the pool of candidates to 138 fluids. The next filters were for chemical stability and toxicity; both were applied manually. Compounds with generally unstable functional groups were dropped from further consideration [1].



International Conference on Cryogenics and Refrigeration 2018, April. 12-14, 2018, Shanghai

ICCR2018-0000

3. PERFORMANCE EVALUATION

The filter of 320 K $< T_{crit} < 420$ K ensured that the accepted 138 low-GWP fluids had the COP and Q_{vol} in a general range of the current refrigerants used in small AC and refrigeration units. For accurate estimations of COP and $Q_{\rm vol}$, we performed two rounds of thermodynamic cycle simulations at different levels of detail. The second simulation round used an advanced cycle model, which accounts for the effects related to transport properties and for the drop of saturation temperature with respect to pressure, dT_{sat}/dp . The model optimized the refrigerant circuitry in heat exchangers to maximize COP, i.e., it arrived at the optimized refrigerant mass flux that enhanced the refrigerant heat transfer coefficient at an acceptable penalty of pressure drop. This feature is important for a realistic representation of optimized heat exchangers relying on refrigerant forced-convection heat transfer.

Following cycle simulations, we generated a list of 28 single-component fluids deemed to be the best for consideration in small AC systems [2]. In this list, 21 fluids are a subset of the 138 fluids, i.e., those with 320 K $< T_{\rm crit} < 420$ K after unstable or higher-toxicity fluids were removed along with those having low $Q_{\rm vol}$ or COP. With the intent to produce a complete list of refrigerant options, we added four fluids based on evidence of commercial interest, and an additional three fluids with T_{crit} < 320 K, including carbon dioxide (R-774). HFOs constitute the largest group in the final list. The other fluids are halogenated alkanes, halogenated oxygenates, hydrocarbons, halogenated nitrogen and sulfur compounds, and inorganic compounds. R-744 is the only lowtoxicity and nonflammable fluid in the list, but it would operate in a transcritical cycle because of its low $T_{\rm crit}$.

The identified fluids with good COP and low toxicity are at least mildly flammable. Most of the fluids are relatively large molecules with large molar heat capacities. As such, they experience greater throttling losses in the basic vapor-compression cycle than currently-used HFCs and markedly benefit from the economizer cycle (Figures 1a and 1b). The list includes six novel fluids. None of them are particularly compelling from a performance standpoint, particularly given that the hazards they may present are unknown.



Figure 1: COP and Q_{vol} of selected fluids referenced to R-410A values; (a) basic cycle and (b) economizer cycle; (excludes fluids with low T_{crit}; includes R-22, R-134a, and R-125 for comparison purposes) [2].

4. CONCLUSIONS

Low-GWP refrigerant options are very limited, particularly for fluids with volumetric capacities similar to those of R-410A and R-404A. Fluids with good COP and low toxicity are at least mildly flammable. The probability of finding better-performing, low-GWP fluids is minimal. The limited refrigerant options and environmental concerns make it imperative to use refrigerants judiciously.

ACKNOWLEDGEMENT

This work was supported by the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy under contract no. DE-EE002057 with A. Bouza and B. Habibzadeh serving as Project Managers.

REFERENCES

- [1] McLinden, M. O., Brown, J. S., Kazakov, A. F., Brignoli, R., Domanski, P. A, 2017. Limited options for lowglobal-warming-potential refrigerants. Nature Communications, 8:14476, doi:10.1038/ncomms14476.
- [2] Domanski, P.A., Brignoli, R., Brown, J.S., Kazakov, A.F., McLinden, M.O., 2017. Low-GWP Refrigerants for Medium and High-Pressure Applications, Int. J. Refrig., 84:198-209, doi:10.1016/j.ijrefrig.2017.08.019.



International Conference on Cryogenics and Refrigeration 2018, April. 12-14, 2018, Shanghai

Brignoli, Riccardo; Brown, J; Domanski, Piotr; Kazakov, Andrei; McLinden, Mark. "Low-GWP Refrigerants: Options and Issues." Paper presented at Int. Conference on Cryogenics and Refrigeration 2018, April. 12-14, 2018, Shanghai., Shanghai, China. April 12, 2018 -

Proceedings of the 2018 Manufacturing Science and Engineering Conference MSEC 2018 June 18-22, 2018, College Station, TX, USA

MSEC2018-6550

A STANDARDS AND TECHNOLOGY ROADMAP FOR SCALABLE DISTRIBUTED MANUFACTURING SYSTEMS

Thomas Hedberg, Jr.*, Moneer Helu, Timothy Sprock

Systems Integration Division, Engineering Laboratory National Institute of Standards and Technology 100 Bureau Drive, Stop 8260 Gaithersburg, MD 20899 USA

ABSTRACT

The increasing decentralization of manufacturing has contributed to the growing interest in scalable distributed manufacturing systems (DMSs). The emerging body of work from smart manufacturing, Industrie 4.0, Industrial Internet of Things (IIoT), and cyber-physical systems can enable the continued development of scalable DMS, particularly through the digital thread. However, significant challenges exist in understanding how to apply the digital thread most appropriately for scalable DMS. This paper describes these major challenges and provides a standards and technology roadmap developed from the digital thread viewpoint and consensus-built industrial standards to realize scalable DMS. The goal of this roadmap is to guide research that enables manufacturers to take advantage of opportunities provided by scalable DMS, including improved agility, flexibility, traceability, dynamic decision making, and utilization of manufacturing resources.

1 Introduction

Several trends have contributed towards the growing decentralization of manufacturing systems. For example, with the growth of global production networks original equipment manufacturers (OEMs) have increasingly become system integrators rather than just manufacturers. There is also a growing desire to design and manufacture products closer to the end user, which requires customer involvement in the product development process and enables more individualized products [1]. Furthermore, both small and large manufacturers have identified opportunities to leverage external capacity to provide flexibility. For example, small-to-medium manufacturers often have excess capacity that can be monetized by providing services to OEMs that need resources to meet planned or unplanned production changes. These trends have contributed to a growing interest in distributed and federated systems, which we will call distributed manufacturing systems (DMSs).

DMSs are manufacturing systems composed of heterogeneous components that have a means of semantic interoperability that enables the coordination and control of activities. Figure 1 provides a schematic representation of a DMS. There are a variety of examples of DMSs, including tiered supply chains common in the aerospace and automotive sectors, traditional job shops that may use equipment of different capability and vintage from a variety of vendors, and emerging organizations focused on providing manufacturing services on-demand. While the scale of each of these examples may be substantially different, they are architecturally similar at a meaningful level of abstraction, which enables the scalability of DMSs. Scalable DMSs can provide useful capabilities to manufacturers, including agility, flexibility, traceability, dynamic decision making, and ultimately improved utilization of manufacturing resources.

The growing body of work from smart manufacturing, Industrie 4.0 (I4.0), Industrial Internet of Things (IIoT), and cyberphysical systems (CPSs) can support the continued development of scalable DMSs. These concepts all promise to link the various phases, viewpoints, and systems of the product lifecycle, which allows manufacturers to deliver higher-quality products to mar-

1

Hedberg, Thomas; Helu, Moneer; Sprock, Timothy. "A Standards and Technology Roadmap for Scalable Distributed Manufacturing Systems." Paper presented at 2018 ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18,

^{*}Corresponding author. Email: thomas.hedberg@nist.gov



FIGURE 1. Cyber-physical representation of a scalable distributed manufacturing system (DMS)

ket in ways that are faster, cheaper, and more sustainable. Linking each portion of the lifecycle requires a means to integrate information and the structure of that information, which is the digital thread [2]. The digital thread is an integrated information flow that connects all of the phases of the product lifecycle using an accepted authoritative data source (e.g., technical data package [3], three-dimensional (3D) computer-aided design (CAD) model) [2, 4, 5]. The digital thread provides the infrastructure needed to link heterogeneous systems to support decision making, knowledge generation, and control.

Both distributed and federated systems can be integrated using the digital thread, but manufacturers require a standards and technology roadmap to understand how to apply the digital thread most appropriately for scalable DMSs. One example of research in this area is the digital surrogate, which is an application of the digital thread to the manufacturing shop-floor environment where integrated information flows are leveraged to digitize production systems and apply modeling and simulation to enable dynamic control (i.e., the Digitization and Realization boxes in Figure 1). The digital surrogate has also been called the digital twin [6, 7, 8, 9]. The term "twin" implies a one-to-one match between a digital model and a physical asset, but one must consider that models are simply representation of things. Thus, models have an inherent uncertainty that must be attributed to them, which is why digital surrogate is a more precise term. The lack of understanding about this inherent uncertainty further motivates the need for a standards and technology roadmap. The goal of this paper is to describe such a roadmap to guide the research needed to realize scalable DMSs.

2 Challenges Hindering Scalable Manufacturing Systems

ISA-95 [10] is an enterprise-control system integration framework, which is intended to help organizations integrate business processes, manufacturing operations, control, and lowlevel processes. Figure 2 presents the scope and hierarchy of the ISA-95 framework. The framework is broken into five levels - level four is the highest level, which is focused on business planning and logistics. Level four is where systems such as enterprise-resource planning (ERP) are deployed. Level three deals with manufacturing operations management and is where a manufacturing-execution system (MES) is utilized. Level two and level one deal with control, where level two focuses on monitoring and level one focusing on sensing. The lowest level is level zero, which deals with the individual physical processes of manufacturing (e.g., milling, turning).

ISA-95 provides a sound foundation for describing the different types of systems and functions that exist in a manufacturing system. Part three [12] of ISA-95 standardizes several activity models of manufacturing operations. These activity mod-

Hedberg, Thomas; Helu, Moneer; Sprock, Timothy. "A Standards and Technology Roadmap for Scalable Distributed Manufacturing Systems." Paper presented at 2018 ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18,



FIGURE 2. Scope and hierarchy of the ISA 95 enterprise-control system integration framework (based on [11])

els provide concise guidance to organizations for the purposes of integrating the various systems encountered within each level. However, in practice, it is difficult to integrate vertically across the levels of ISA-95 [13, 14, 15, 16]. Further, the framework implies single monolithic MES and ERP systems. If this is true, are monolithic integrations capable of being scaled or distributed? This is a question that must be answered before the full benefits of a framework like ISA-95 could be achieved to enable a scalable DMS.

ISA-95 was developed by the process industry. The framework was not purpose built for discrete manufacturing. Enabling scalable DMSs requires both vertical integrations within domains and horizontal integrations between domains. Horizontal integration requires connecting heterogeneous information and systems, which implies the need for semantic mediation. Thus, are the information models (or behaviors, methods, interfaces, services) described with sufficient precision to enable translation of data (e.g., syntactic integration) and/or semantic interoperability? The answer is, not completely.

Each phase of the product lifecycle has different viewpoints and concerns, which lead to different levels of abstraction in modeling and simulation [17, 18, 19]. The various viewpoints lead to information models and systems being developed for a specific purpose, which results in different information models across phases of the lifecycle to look at the same data in different ways. A "fit for purpose" approach to modeling is recommended because it enables "expert systems" that support the user (i.e., human), in a specific function and role, to make decisions in a contextual way.

Conversely, purpose-built models are not scalable. Data requires context when related to decisions [20]. Data alone is not sufficient for decision making because the decision maker must understand the scope and type of the problem the decision is intended to solve. As the scope of the problem changes, the models must also change. Thus, connecting of heterogeneous information and systems to enable scalable DMSs introduces a paradox to the steadfast approach of purpose-driven modeling. A tradeoff of how purpose-built to make a model versus how scalable to make a model must be considered with the shift towards scalable DMSs

Scalable DMSs require an effective and efficient forward and backward communication backbone. This requires integrating domains in multiple directions while providing scalable contextual models. Overcoming these challenges to a scalable DMS is not easy, but we believe a standards-based approach, using the digital thread, provides the best opportunity for maximizing the successful deployment of scalable DMSs.

3 Defining Use Case for Research

Addressing the challenges and barriers to scalable manufacturing systems described in Section 2 requires defining an appropriate use case reflective of industrial practice, but manageable in the context of research. The use case we propose is a flexible machining cell composed of a numerical control (NC) machine tool, coordinate-measurement system (CMS), robot, in-process metrology system (e.g., cutting-tool metrology), and material-handling system (e.g., automated-guide vehicle); see Figure 3. Such a cell would act as an on-demand, pull manufacturing system that autonomously produces finished parts from stock material and resources using the following sample workflow:

- 1. Humans set the initial work plans
- 2. Autonomous systems execute the work
- 3. Control systems study the execution of the work while all systems communicate with each other
- 4. Autonomous systems self-adjust the work plans and act based on inputs from the control system and other systems

The objective of the proposed research is to develop a reference implementation for a flexible machining cell that will scale to larger production systems. The research would leverage and extend the lifecycle information framework and technology proposed by Hedberg et al. [21]. The framework consists of three layers: (1) linked product-lifecycle data, (2) data certification and

Hedberg, Thomas; Helu, Moneer; Sprock, Timothy. "A Standards and Technology Roadmap for Scalable Distributed Manufacturing Systems." Paper presented at 2018 ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18,



FIGURE 3. Representation of flexible machining cell for on-demand, pull production

traceability, and (3) data-driven applications. The framework and technology supports both vertical (i.e., intra-phase) and horizontal (i.e., inter-phase) integration in the product lifecycle.

The proposed research would enable all decisions (e.g., process parameters, schedule, routing) to be made based on the measured and predicted status or capability of the physical components of the manufacturing system as well as additional part and process information (e.g., product and manufacturing information (PMI), order and delivery schedules, process constraints and specifications) from production management and control databases (e.g., product-data management (PDM), ERP, MESs, and quality-management systems (QMSs)). Data and information would be shared across the manufacturing system based on consensus-built industry standards, including Standard for the Exchange of Product Model Data Application Protocol 242 (STEP AP242) [22], MTConnect [23], and Quality Information Framework (QIF) [24]. Such a manufacturing system would represent a CPS paradigm for manufacturing where control of the physical elements would be determined using modeling, simulation, and optimization that occurs in a cyber or digital environment (see Figure 1).

To achieve the goals of the proposed research, the cell must be composed of industry-standard equipment and must process parts designed to reflect authentic production components. Ideally, the development of such a cell would be supported through collaborations between academia, government, and industry partners to ensure that all stakeholder needs are recognized and satisfied. A longer-term goal of this work is to deploy several of these cells with industry partners to verify and validate the implementation and its scalability.

4 Proposed Standards and Technology Roadmap

Distributed decision making is an important component to successfully enabling a scalable DMS. Flexibility and agility is added to the manufacturing system by decoupling the decision directives from the specification requirements - allowing the manufacturing system to focus on managing itself for how best to schedule and make product. This idea moves manufacturing systems towards logistics systems where the manufacturing systems acts as an agent in a larger supply chain [25, 26, 27]. However, several areas of research must be expanded and integrated to support the shift towards scalable DMSs. We identify four immediate areas of need: (1) cyber-infrastructure integration, (2) physical-infrastructure integration, (3) modeling and simulation, and (4) analytics and data science for manufacturing. Successful extension of these four areas of research, coupled with emerging IIoT concepts, provides opportunities to integrate cyber and physical systems in meaningful ways to support the realization of a scalable DMS as depicted in Figure 1.

4.1 Cyber-infrastructure integration

The Digital Thread for Smart Manufacturing project¹ at National Institute of Standards and Technology (NIST) defined the conceptual framework [21], described in Section 3, for lifecycle information management and integration of emerging and existing technologies that support data curation, discovery, and reuse in manufacturing. The project provided a standardized infrastructure to richly represent lifecycle data and place it into the appropriate context to generate useful knowledge. The framework leverages and promulgates existing data standards such as STEP AP242, MTConnect, and QIF.

Needs for cyber-infrastructure integration should include a focus on further extending the digital thread - particularly integrating the elements of the cyber-world depicted in the top portion of Figure 1 while supporting information flow on all the arrows. This area of focus would build upon previous work in the areas of dynamic knowledge-base management, decision support, requirements management, and architectures [16, 21, 28, 29, 30]. Research components for extending the previous work on reference architectures should include the use of agents and on-demand micro-services [31, 32, 33].

Furthering linked-data and semantic web concepts for manufacturing, including the use of handles [34], is required for realizing the goal of autonomous knowledge generation and decision diagnostics. Semantic web for manufacturing would be enabled by standardizing the interfaces between functions, roles, and phases of the product lifecycle. Standard interfaces could be achieved by continuing to identify normalized elements to support a minimum information model [35, 36]. However, gaps [37] in integrating product design, manufacturing, and quality data must be resolved to successfully link the phases of the product lifecycle using standard interfaces.

In addition, authentication, authorization, and traceability cannot be ignored. Hedberg et al. [38] proposed using X.509 Public Key Infrastructure (X.509-PKI) [39] for embedding trace-

4

Hedberg, Thomas; Helu, Moneer; Sprock, Timothy. "A Standards and Technology Roadmap for Scalable Distributed Manufacturing Systems." Paper presented at 2018 ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18,

¹https://www.nist.gov/programs-projects/digital-thread-smart-manufacturing
ability metadata into artifacts. The embedded metadata would assist functions of the lifecycle in determining what data is, how the data can be used, and who did what to the data. This could all be done in support of trustworthiness. Additional work is required to determine a metadata schema that enables trusted exchange of artifacts throughout the product lifecycle. Emerging technologies, such as blockchain and distributed ledgers, could also support trustworthiness requirements.

4.2 Physical-infrastructure integration

The Prognostics, Health Management, and Control (PHMC) project² at NIST delivered methods, protocols, and tools for robust sensing, diagnostics, prognostics, and control that enable manufacturers to respond to planned, new, and un-planned performance changes towards the goal of enhancing the efficiency of smart manufacturing systems. The project promoted advanced sensing, prognostics and health management, and control from ISA-95 manufacturing levels zero through three. The resulting impact is improved decision-making support and greater automation with a focus on vendor-neutral approaches and plug-andplay solutions.

In addition, NIST provides the public with real manufacturing data of a contract manufacturer through the Smart Manufacturing Systems (SMS) Test Bed. The SMS Test Bed is comprised of a computer-aided technologies (CAx) lab containing several computer-aided technology tools, a manufacturing lab mimicking the configuration of a contract-manufacturing shop, and data publication web services. The goal of the SMS Test Bed is to extend existing production-focused concepts by designing and developing an architecture [30] for a test bed that enables smart manufacturing research and development across the product lifecycle [40].

Needs for physical-infrastructure integration should include focusing on leveraging the reference architecture from the SMS Test Bed [30] while coupling the work from the PHMC project with part-process methods and an integration of fabrication systems, robotic systems, and metrology systems. This focus would address the the physical-world depicted in the bottom portion of Figure 1 and would define standard interfaces for the inputs from realization and outputs to digitization.

Sensing, monitoring, and control methods would be further studied to provide a full integration of ISA-95 from level zero to level four. This would require support from the cyberinfrastructure integration work to support connecting and contextualizing the representation of ERP and MES systems. Further, methods to control physical machines (e.g., fabrication, robots, metrology) requires an innovative shift towards modelbased control.

Manufacturing has reached the fundamental limits of what

its tools and processes can manage. For example, G-code was developed at a point when hardware and computing power was a limiting factor and the programming approach was not intended for accessing the machine controller directly [41]. However, computing power today far exceeds the capabilities of G-code for programing machine tools. Hardware and computing are no longer the limiting factors - G-code has become the functional limits of what a machine tool can achieve. Model-based control systems are feasible now because of the advancement in computing power. Fabrication systems would benefit significantly by enabling a controller to interact directly with a part and process model [42].

The research direction we propose for physicalinfrastructure integration would enable a model-based transformation of manufacturing control. Models, coupled with sensing and monitoring, would be used to plan processes, execute (do) those processes, study the execution, and take action to ensure effective and efficient performance of the manufacturing system. Applying the Deming-Shewhart Plan-Do-Study-Act cycle [43] to our proposed approach of scalable DMSs would ensure the system is achieving its goals and that all decisions are determined based on measured physical components of the system and part and process information.

4.3 Modeling and simulation

Two related projects at NIST are focused on increasing access to and availability of analytical capabilities to support smart manufacturing [16]. The Systems Analysis Integration for Smart Manufacturing Operations (SAISMO) project³ has developed methods and protocols to facilitate analysis of smart manufacturing operations by enabling efficient integration of smart manufacturing systems models and engineering analysis models. The Modeling Methodology for Manufacturing System Analysis project⁴ has developed new ways of applying analytical and empirical methods using domain-specific modeling methodologies tailored for smart manufacturing.

Modeling and simulation research addresses the needs of the analysis environment of Figure 1. Grand challenges in modeling and simulation for manufacturing systems have been addressed in [44, 45]. An overarching challenge is time and expertise required to develop useful models and simulations of manufacturing systems and populate (or update) them with information gathered from the system.

System models developed using the Systems Modeling Language (SysML) [46] provide a 'single source of truth' for organizing and integrating heterogeneous models and information gathered from multiple engineering disciplines involved in sys-

5

Hedberg, Thomas; Helu, Moneer; Sprock, Timothy. "A Standards and Technology Roadmap for Scalable Distributed Manufacturing Systems." Paper presented at 2018 ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18,

²https://www.nist.gov/programs-projects/prognostics-health-managementand-control-phmc

³https://www.nist.gov/programs-projects/systems-analysis-integrationsmart-manufacturing-operations

⁴https://www.nist.gov/programs-projects/modeling-methodology-smartmanufacturing

^{2018 -} June 22, 2018.

tem design and operation. Formal system models convey unambiguous, shared meaning of manufacturing concepts that extends beyond common information models. Model-based systems engineering (MBSE) methods integrate concerns from multiple engineering disciplines to support the design, operation, and maintenance of product and production systems [17, 47], including the many viewpoints and abstractions used to construct analysis models [18, 48].

One area of intense focus is the formalization of part and process models and linking those models to shop floor data collection. Part models capture details about what should be produced (e.g., CAD models) and how well the resulting object was actually produced (e.g., QIF information). Process models describe both what needs to be done to produce a part (capabilities) and capture options for sequencing the execution of those capabilities (process steps). These models capture execution information, including duration and machines that executed each step. They may provide a common representation of manufacturing capabilities, both required by parts and provided by machines. Incorporating semantic meaning with raw data enables consistent interpretation across applications using that data. This information informs many operational decisions, such as scheduling, enabling these decisions to be made dynamically using real-time feedback from the shop floor.

Enabling dynamic decision making, such as scheduling, in smart manufacturing environments requires access to robust decision-support, often provided by simulation and optimization tools. Integrating simulation and optimization tools with sources of system information can be enabled by system models that describe standard interfaces to analysis tools [49].

The research direction we propose for addressing the analysis environment requirements for smart manufacturing systems focuses on modeling and simulation challenges. Standard reference models containing reusable domain-specific part, process, and resource definitions provide a semantic foundation for integrating simulation and optimization models, methods, and tools with shop floor decision-support.

4.4 Analytics and data science for manufacturing

The Data Analytics for Smart Manufacturing project⁵ at NIST developed standards, software tools, methodologies, and guidelines to enable small-and-medium enterprises to apply data analytics services to improve decision-making and performance in smart manufacturing systems. The project studied four key areas: information standards, measurement methods, integration framework, and a data analytics testbed. The project proposed two new models to extend Predictive Model Markup Language (PMML) - Gaussian process regression and Bayesian networks. In addition, the project supporting the development of the American Society of Mechanical Engineers (ASME) subcommittee for verification and validation of computational modeling for advanced manufacturing⁶.

Needs for analytics and data science in manufacturing should include a focus on extending the information standards, measurements methods, and integration framework further into predictive and prescriptive analytics. The goal would be to further mature manufacturing models for predictive analytics, domain-specific languages for performing predictive analytics, standard interfaces for data analytics tools, and further study of uncertainty quantification. Then, models could be developed to move beyond predictive analytics toward prescribing what needs to be done based on an ability to predict the future and understand the interactions of the various operation and materials on the systems. Enhancing analytics and data science in manufacturing would support decisions in and interactions with each component represented in Figure 1.

The study of analytics and data science in manufacturing must include both data analytics [50, 51] and visual analytics [52]. This would leverage our proposed modeling and semantics work to enable effective decision making by both machines and humans. Further, leveraging the traceability, security, verification, validation, and data provenance methods from our proposed cyber-infrastructure integration work, and combining it with enhanced uncertainty quantification, would enable effective research in applying machine learning and artificial intelligence concepts to decision making in scalable DMSs.

Significant time is spent searching for data and developing knowledge across the product lifecycle. Leveraging trustworthy artificial intelligence would achieve autonomous decision support, requirements management, and knowledge management across distributed and/or federated systems. This would free up time for labor to focus solely on the activities that require human participation.

5 Summary

The digitalization of manufacturing systems is in full swing thanks to international pursuits in smart manufacturing, I4.0, However, work remains in fully realizing the digital thread in support of scalable DMS. Shifting the design specification paradigm from primarily paper-based twodimensional (2D) drawings to 3D model-based definitions could reduce manufacturing and inspection cycle time by up to 75% on average [2]. Further, managing digital-data streams through models while improving the transmission of digital information, enhancing sensing and monitoring, advancing the use of data analytics, and efficiently communicating information to decision makers to help determine and implement required actions would save United States manufacturers \$57.4 billion annually [53].

2018 - June 22, 2018.

⁵https://www.nist.gov/programs-projects/data-analytics-smartmanufacturing-systems

⁶https://cstools.asme.org/csconnect/CommitteePages.cfm?Committee=101978604

Hedberg, Thomas; Helu, Moneer; Sprock, Timothy. "A Standards and Technology Roadmap for Scalable Distributed Manufacturing Systems." Paper presented at 2018 ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18,

The annual savings could be significantly higher as the figures do not account for reinvestment of savings and other value gains as a result of more manufacturing-system availability and potential positive cash flow.

More importantly, deploying scalable DMS makes the overall system(s) more agile. Engineers would be able to do more with less. Decisions are made today based on a limited set of knowledge and/or selection options. Increasing the agility of manufacturing systems would enable designers and engineers to create on an unprecedented scale since they will now have tools to rapidly test ideas and create new products that will meet customer requirements much better than they do today.

However, barriers to innovation increase the cost of smart manufacturing research and development (R&D) and smallto-medium manufacturers face the highest burden to adopting smart-manufacturing technologies [1]. Therefore, standards must play a key role in the advancement of paradigm-setting technologies. Industry requires standards that support both vertical integrations of manufacturing systems and horizontal integrations of the manufacturing domain with design and other phases of the product lifecycle.

NIST researchers expect the development of a flexible machining cell with industrial partners will illuminate the challenges that must be addressed to enable growth of smart manufacturing and the digital thread in industry. For example, one area identified is the need for standardized interfaces between physical components of manufacturing systems that allow these components to coordinate their activities. Another need is modeling and simulation approaches that can capture the complex interactions in manufacturing systems as well as the lack of observability of certain process and system data. Through this effort, NIST hopes to contribute to the enhancement of standards and technologies where appropriate.

The primary goal of the effort proposed in this paper would be the development of a reference implementation of a flexible, on-demand, pull manufacturing system that may be used by industry to leverage the digital thread within production systems. The digital thread enables the collection, transmission, and sharing of data and information between heterogeneous productlifecycle systems, which are typically silo-ed between different functions and organizations [21]. Such integration allows industry to use data and information to reduce costs, improve productivity, ensure first-pass success, and augment existing capabilities in the workforce [40]. Our proposal also addresses evolving industry challenges due to the increasingly distributed nature of modern manufacturing activities and the growing complexity of manufacturing systems and networks.

Acknowledgments

The authors wish to thank Allison Barnard Feeney (NIST), Michael Brundage (NIST), Vijay Srinivasan (NIST), William Sobel (Vimana), and the peer-reviewers for their comments and input to this paper.

Disclaimers

The work presented in this paper is an official contribution of the National Institute of Standards and Technology (NIST) and not subject to copyright in the United States. Certain commercial systems are identified in this paper. Such identification does not imply recommendation or endorsement by NIST. Nor does it imply that the products identified are necessarily the best available for the purpose.

REFERENCES

- [1] Gallaher, M. P., Oliver, Z. T., Rieth, K. T., and O'Connor, A. C., 2016. Economic analysis of technology infrastructure needs for advanced manufacturing: Smart manufacturing. Report NIST GCR 16-007, RTI International.
- [2] Hedberg Jr, T. D., Lubell, J., Fischer, L., Maggiano, L., and Barnard Feeney, A., 2016. "Testing the digital thread in support of model-based manufacturing and inspection". Journal of Computing and Information Science in Engineering, 16(2), pp. 1-10.
- [3] US Department of Defense, 2013. Standard practice: Technical data packages, 11/1/2009.
- [4] Kraft, E. M., 2016. The Air Force Digital Thread/Digital Twin - Life Cycle Integration and Use of Computational and Experimental Knowledge. AIAA SciTech Forum. American Institute of Aeronautics and Astronautics.
- [5] Wardhani, R., and Xu, X., 2016. "Model-based manufacturing based on step ap242". In 2016 12th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA), pp. 1-5.
- [6] Holzwarth, R., Tuegel, E., and Kobryn, P., 2012. "Airframe digital twin: Creating virtual replicas of every aircraft in the fleet". In Prognostics and Health Management Solutions Conference - PHM: Driving Efficient Operations and Maintenance, MFPT 2012, April 24, 2012 - April 26, 2012, Technical Program for MFPT 2012, The Prognostics and Health Management Solutions Conference - PHM: Driving Efficient Operations and Maintenance, Machine Failure Prevention Technology Society (MFPT).
- [7] Tuegel, E. J., 2012. "The airframe digital Some challenges to realization". twin In 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 2012, April 23, 2012 - April 26, 2012, 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 2012, American Institute of Aeronautics and Astronautics Inc., p. American Institute of Aeronautics and Astronautics (AIAA).

Hedberg, Thomas; Helu, Moneer; Sprock, Timothy. "A Standards and Technology Roadmap for Scalable Distributed Manufacturing Systems." Paper presented at 2018 ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18,

2018 - June 22, 2018.

- [8] Cerrone, A., Hochhalter, J., Heber, G., and Ingraffea, A., 2014. "On the effects of modeling as-manufactured geometry: Toward digital twin". International Journal of Aerospace Engineering, p. 10.
- [9] Rosen, R., von Wichert, G., Lo, G., and Bettenhausen, K. D., 2015. "About the importance of autonomy and digital twins for the future of manufacturing". Ifac Papersonline, 48(3), pp. 567-572.
- [10] The International Society of Automation, 2010. Enterprisecontrol system integration part 1: Models and terminology.
- [11] Johnsson, C., Brandl, D., and Unger, K., 2006. ISA 95 for begineers. Report, WBF.
- [12] The International Society of Automation, 2013. Enterprisecontrol system integration part 3: Activity models of manufacturing operations management.
- [13] Helo, P., Suorsa, M., Hao, Y., and Anussornnitisarn, P., 2014. "Toward a cloud-based manufacturing execution system for distributed manufacturing". Computers in Industry, 65(4), pp. 646-656.
- [14] Chu, Y., and You, F., 2015. "Model-based integration of control and operations: Overview, challenges, advances, and opportunities". Computers & Chemical Engineering, 83(Supplement C), pp. 2-20.
- [15] Leitao, P., Colombo, A. W., and Karnouskos, S., 2016. "Industrial automation based on cyber-physical systems technologies: Prototype implementations and challenges". Computers in Industry, 81(Supplement C), pp. 11-25.
- [16] Barnard Feeney, A., Frechette, S., and Srinivasan, V., 2017. Cyber-Physical Systems Engineering for Manufacturing. Springer International Publishing, Cham, pp. 81-110.
- [17] Sprock, T., Murrenhoff, A., and McGinnis, L. F., 2016. "A hierarchical approach to warehouse design". International Journal of Production Research, pp. 1–13.
- [18] Sprock, T., and Bock, C., 2017. "Incorporating abstraction methods into system-analysis integration methodology for discrete event logistics systems". In Winter Simulation Conference (WSC), 2017, IEEE.
- [19] Regli, W., Rossignac, J., Shapiro, V., and Srinivasan, V., 2016. "The new frontiers in computational modeling of material structures". Computer-Aided Design, 77, pp. 73-85.
- [20] Brandl, D., 2013. "Drowning in data, starved for information". Control Engineering.
- [21] Hedberg Jr, T., Barnard Feeney, A., Helu, M., and Camelio, J. A., 2017. "Towards a lifecycle information framework and technology in manufacturing". Journal of Computing and Information Science in Engineering, 17(2), pp. 021010-021010-13.
- [22] International Standards Organization, 2014. Industrial automation systems and integration - product data representation and exchange - part 242: Application protocol: Managed model-based 3d engineering.

- [23] MTConnect Institute, 2014. Mtconnect standard.
- [24] Dimensional Metrology Standards Consortium, 2014. Part 1: Overview and fundamental principles in quality information framework (QIF) an integrated model for manufacturing quality information.
- [25] Duffie, N. A., 1996. "Heterarchical control of highly distributed manufacturing systems". International Journal of Computer Integrated Manufacturing, 9(4), pp. 270–281.
- [26] Maturana, F. P., and Norrie, D. H., 1996. "Multi-agent mediator architecture for distributed manufacturing". Journal of Intelligent Manufacturing, 7(4), pp. 257-270.
- [27] Shen, W., and Norrie, D. H., 1999. "Agent-based systems for intelligent manufacturing: A state-of-the-art survey". Knowledge and Information Systems, 1(2), pp. 129–156.
- [28] Hedberg Jr, T. D., Hartman, N. W., Rosche, P., and Fischer, K., 2017. "Identified research directions for using manufacturing knowledge earlier in the product life cycle". International Journal of Production Research, 55(3), pp. 819-827.
- [29] Feng, S. C., Bernstein, W. Z., Hedberg, J. T., and Barnard Feeney, A., 2017. "Toward knowledge management for smart manufacturing". Journal of Computing and Information Science in Engineering, 17(3), pp. 031016-031016-9.
- [30] Helu, M., Hedberg Jr, T., and Barnard Feeney, A., 2017. "Reference architecture to integrate heterogeneous manufacturing systems for the digital thread". CIRP Journal of Manufacturing Science and Technology, 19, pp. 191–195.
- [31] Cristea, V., Pop, F., Dobre, C., and Costan, A., 2011. Distributed Architectures for Event-Based Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 11-45.
- [32] Vianden, M., Lichter, H., and Steffens, A. "Experience on a microservice-based reference architecture for measurement systems". In 2014 21st Asia-Pacific Software Engineering Conference, Vol. 1, pp. 183-190.
- [33] Krivic, P., Skocir, P., Kusek, M., and Jezic, G., 2017. Microservices as Agents in IoT Systems. Springer International Publishing, Cham, pp. 22-31.
- [34] Bajaj, M., and Hedberg Jr, T. "System lifecycle handler spinning a digital thread for manufacturing". In 28th Annual INCOSE International Symposium, INCOSE.
- [35] Ruemler, S. P., Zimmerman, K. E., Hartman, N. W., Hedberg, J. T., and Barnard Feeney, A., 2016. "Promoting model-based definition to establish a complete product definition". Journal of Manufacturing Science and Engineering, 139(5), pp. 051008-051008. 10.1115/1.4034625.
- [36] Miller, A. M., Hartman, N. W., Hedberg Jr, T., Barnard Feeney, A., and Zahner, J., 2017. "Towards identifying the elements of a minimum information model for use in a model-based definition". In International Manufacturing Science and Engineering Conference, ASME, p. V003T04A017.
- [37] Trainer, A., Hedberg Jr, T., Barnard Feeney, A., Fischer,

Hedberg, Thomas; Helu, Moneer; Sprock, Timothy. "A Standards and Technology Roadmap for Scalable Distributed Manufacturing Systems." Paper presented at 2018 ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18, 2018 - June 22, 2018.

K., and Rosche, P. "Gaps analysis of integrating product design, manufacturing, and quality data in the supply chain using model-based definition". In 2016 Manufacturing Science and Engineering Conference, American Society of Mechanical Engineers.

- [38] Hedberg Jr, T. D., Krima, S., and Camelio, J. A., 2016. "Embedding x.509 digital certificates in three-dimensional models for authentication, authorization, and traceability of product data". Journal of Computing and Information Science in Engineering, 17(1), pp. 011008-011008-11.
- [39] The Internet Engineering Task Force, 2013. Internet X.509 public key infrastructure certificate and certificate revocation list (CRL) profile.
- [40] Helu, M., and Hedberg Jr, T., 2015. "Enabling smart manufacturing research and development using a product lifecycle test bed". Procedia Manufacturing, 1, pp. 86-97.
- [41] Toncich, D. J., 1993. Data Communications and Networking for Manufacturing Industries. Chrystobel Engineering, Brighton Australia.
- [42] Lynn, R., Contis, D., Hossain, M., Huang, N., Tucker, T., and Kurfess, T., 2016. "Extending access to hpc manufacturability feedback software through hardware-accelerated virtualized workstations". In Flexible Automation (ISFA), International Symposium on, IEEE, pp. 271–277.
- [43] Deming, W. E., 2000. The new economics : for industry, government, education, 2nd ed. MIT Press, Cambridge, Mass.
- [44] Fowler, J. W., and Rose, O., 2004. "Grand challenges in modeling and simulation of complex manufacturing systems". Simulation, 80(9), pp. 469-476.
- [45] Mönch, L., Lendermann, P., McGinnis, L. F., and Schirrmann, A., 2011. "A survey of challenges in modelling and decision-making for discrete event logistics systems". Computers in Industry, 62(6), pp. 557-567.
- [46] Object Management Group, 2017. OMG systems modeling language (OMG SysML) version 1.5.
- [47] Bock, C., Zha, X., Suh, H.-w., and Lee, J.-H., 2010. "Ontological product modeling for collaborative design". Advanced Engineering Informatics, 24(4), pp. 510–524.
- [48] Sprock, T., In Press. Patterns for modeling operational control of discrete event logistics systems (DELS). Springer International Publishing, book section 61.
- [49] Sprock, T., and McGinnis, L. F., 2016. "Simulation optimization in discrete event logistics systems: The challenge of operational control". In Winter Simulation Conference (WSC), 2016, IEEE, pp. 1170-1181.
- [50] Lechevalier, D., Narayanan, A., and Rachuri, S. "Towards a domain-specific framework for predictive analytics in manufacturing". In 2014 IEEE International Conference on Big Data (Big Data), pp. 987-995.
- [51] Narayanan, A., Ak, R., Lee, Y.-T. T., Ghosh, R., and Rachuri, S., 2017. Summary of the symposium on data

analytics for advanced manufacturing. Report AMS 100-7, National Institute of Standards and Technologyu.

- [52] Ramanujan, D., Bernstein, W. Z., Chandrasegaran, S. K., and Ramani, K., 2017. "Visual analytics tools for sustainable lifecycle design: Current status, challenges, and future opportunities". Journal of Mechanical Design, 139(11), pp. 111415-111415-19. 10.1115/1.4037479.
- [53] Anderson, G., 2016. The economic impact of technology infrastructure for smart manufacturing. Report NIST Economic Analysis Briefs 4, National Institute of Standards and Technology, October.

Hedberg, Thomas; Helu, Moneer; Sprock, Timothy. "A Standards and Technology Roadmap for Scalable Distributed Manufacturing Systems." Paper presented at 2018 ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18,

2018 - June 22, 2018.

46th SME North American Manufacturing Research Conference, NAMRC 46, Texas, USA

Force modeling for hybrid manufacturing

Michael Gomez¹, Jarred Heigel², and Tony Schmitz^{1*}

¹UNC Charlotte, Mechanical Engineering and Engineering Science Department, 9201 University City Boulevard, Charlotte, NC 28223, USA ²National Institute of Standards and Technology, Production Systems Group, 100 Bureau Drive, Gaithersburg, MD 20899, USA

* Corresponding author. Tel.: +1-704-687-5086. *E-mail address:* tony.schmitz@uncc.edu

Abstract

This paper investigates the cutting forces during the machining of additively manufactured metals. Two pairs of workpieces were produced by powder bed direct metal selective laser sintering. These workpieces included a 17-4 stainless steel substrate (wrought) and a 17-4 stainless steel additively manufactured rib. After manufacture, one pair was annealed and the other was maintained in the as-produced state. Each was then machined to identify a cutting force model for the various material states. An instantaneous force, nonlinear optimization model was applied to determine the mechanistic cutting force coefficients. Due to the nonlinear dependence on the commanded feed per tooth, a power law fit was applied to the chip thickness-dependent cutting force coefficients. It was determined that there are only minor differences in the cutting force coefficients between the various material states. The strongest effect was annealing, which increased the mean coefficients in almost all cases, particularly for the additively manufactured bulk material.

© 2018 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 4th International Conference on System-Integrated Intelligence.

Keywords: Additive manufacturing; machining; cutting force coefficients

1. Main text

Powder bed fusion (PBF) is an additive manufacturing (AM) technique that enables metal components with complex geometries to be produced. Increased surface roughness and increased dimensional uncertainty are disadvantages of AM as compared to computer numerically-controlled (CNC) machining [1]. To address these limitations, hybrid manufacturing combines additive and subtractive manufacturing strategies. This enables complex metal parts to be produced with the required surface finish and accuracy. An important consideration in hybrid manufacturing is the effect of the additive influence on material removal in the subtractive step.

To determine the influence of the additive step on the corresponding cutting forces, a mechanistic cutting force model is applied here to identify empirical coefficients that relate the required force for material removal to the commanded chip width and thickness. This paper characterizes these cutting force coefficients and explores the effects of the PBF process, as well as follow-on heat treating, on workpiece surface and material characteristics through a collaboration between UNC Charlotte and NIST*.

2. Additive workpieces

A total of four workpieces were used for the cutting tests. Each workpiece consisted of a wrought 17-4 stainless steel substrate (25 mm tall, 50 mm wide, 150 mm long) and a PBF-deposited rib (25 mm tall, 15 mm wide, 140 mm long). The PBF process parameters were held constant and repeated to create a second pair of workpieces. One set of workpieces was heat treated at 650 °C for 1 hr, in an effort to relieve the residual stress generated by the PBF process, while the other set remained in the as-produced condition; see Figure 1.



Fig. 1. One pair of as-produced workpieces mounted on the PBF build platform. Each workpiece has a PBF stainless steel rib (top section) and a wrought substrate (bottom section).

Three material regions of interest were identified: the wrought substrate, the AM surface material, and the AM bulk material. The wrought substrate served as a baseline material for which the AM material characteristics could be directly compared. The AM surface material was defined as the portion of the PBFdeposited material within 1 mm of the vertical faces of the PBF feature. The AM surface (shown by Figure 1) generated by the PBF build process exhibited the characteristic surface roughness. The AM bulk material was the remainder of the PBF feature and did not include the rough PBF surface.

Table 1. Overview of the AM workpieces and material regions of interest

Workpiece	Material region	Heat treat	Effects
1, 3	Wrought substrate	No	As-produced baseline
1, 3	AM surface	No	Rough PBF surface
			Residual stress
1, 3	AM bulk	No	Residual stress
2,4	Wrought substrate	Yes	Heat-treat baseline
2,4	AM surface	Yes	Rough PBF surface
			Residual stress
2, 4	AM bulk	Yes	Residual stress



Fig. 2. Surface profile measurements along the top surface of the PBF stainless steel ribs.

intended to imply that the materials or equipment identified are necessarily the best available for the purpose. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

2018 - June 22, 2018.

^{*} Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it



Fig. 3. Surface profile measurements along the bottom surface of the wrought substrate for each workpiece.

In order to characterize the geometric distortion in the AM workpieces, the top and bottom surfaces of the workpieces were measured with a MarSurf LD 260 stylus profilometer. A trace from a 120 mm section along the top surface of the PBF stainless steel substrates is provided in Figure 2. The high spatial frequency content is characteristic of the surface roughness from PBF processes. This is evident when comparing Figure 2 to Figure 3, which reports the same measurement performed on a 120 mm segment along the bottom of the wrought substrates for each pair of workpieces. From the results reported in Figures 2 and 3, the annealed workpieces experienced a greater amount of out-of-plane distortion.

3. Force modeling

There are a number of strategies available for cutting force modeling. In this paper, a mechanistic force model is applied to the milling process. The mechanistic force model is based on two underlying assumptions [2]: 1) the instantaneous cutting force is directly proportional to the chip thickness and width; and 2) the instantaneous cutting force is independent of other operating parameters. While these are acceptable assumptions for many applications, adjustments can be made to the force model to accommodate other process parameters [2, 3].

The cutting model applied in this study relates the uncut chip dimensions to the cutting force components using Eqs. 1-3 [2]:

$$F_t = k_t bh + k_{te} b \tag{1}$$

$$F_r = k_r bh + k_{re}b \tag{2}$$

$$F_a = k_a b h + k_{ae} b \tag{3}$$

where F_t , F_r , and F_a are the tangential, radial, and axial direction force components, respectively. The coefficients k_t , k_r , and k_a relate the force to chip area, where *b* is the axial depth of cut (chip width) and *h* is the instantaneous chip thickness. The edge coefficients k_{te} , k_{re} , and k_{ae} relate the force to the axial depth only and are associated with rubbing and non-cutting contact [2].

For this analysis, six different force models are established for the material regions of interest identified in Table 1. The empirical cutting force coefficients generated by the force model are then used to make a direct comparison between the asproduced and annealed PBF cases.

A nonlinear optimization algorithm was used to characterize the cutting force coefficients. The leastsquares nonlinear optimization method [3] simulates the cutting forces over one cutter revolution using Eqs. 1-3. It then minimizes the differences between the simulated and measured cutting forces in the time domain. This optimization model takes the form of Eq. 4, where the x, y, and z force components are determined from a projection of the rotating force components (Eqs. 1-3) using the cutter rotation angle [3].

$$f(k_{t},k_{r},k_{a},k_{te},k_{re},k_{ae}) = \sum \left\| \begin{cases} F_{x} \\ F_{y} \\ F_{z} \end{cases} \right\|_{simulated} - \begin{cases} F_{x} \\ F_{y} \\ F_{z} \end{cases} \right\|_{measured} = \left\{ F_{x} \\ F_{y} \\ F_{z} \end{bmatrix}_{measured} \right\|^{2}$$
(4)

4. Experimental setup

The cutting force coefficients were determined experimentally from a series of machining trials performed under stable, down-milling conditions. The axial and radial depths of cut were 3 mm and 1 mm, respectively. The spindle speed was held constant at 4000 rpm and the feed per tooth was varied. The details of the cutting tests are provided in Table 2.

The tests were carried out on a Haas TM-1 CNC milling machine with a maximum spindle speed of 4000 rpm. The workpieces were mounted on a threeaxis force dynamometer (Kistler 9257B). Cutting tests were performed using a 19.05 mm diameter indexable

Gomez, Michael; Heigel, Jarred; Schmitz, Tony. "Force modeling for hybrid manufacturing." Paper presented at 46th SME North American Manufacturing Research Conference, NAMRC 46, College Station, TX, United States. June 18, 2018 - June 22, 2018.

end mill with a single cutting insert attached (i.e., a single tooth tool). This configuration eliminated potential runout effects. Additionally, a new PVDcoated micro-grain carbide cutting insert (Sandvik Coromant 390R-070204E-MM S30T) was used for each test to mitigate the effects of tool wear. The machining setup for cutting force measurements is shown in Figure 4.

Table 2. Machining parameters used for cutting force measurements

Feed per tooth [mm]	Spindle speed [rpm]	Milling direction	Axial depth of cut [mm]	Radial depth of cut [mm]
0.025-0.125	4000	Down	3	1



Fig. 4. Machining experimental setup.

Insert	Diameter difference [µm]
1	0
2	152
3	51
4	25
5	76
6	127
7	0
8	25
9	-25
10	102

Tool cutting diameter measurements were performed as each insert was replaced between tests. The results are displayed in Table 3, where the measurements were performed in the machine spindle using a Mitutoyo dial indicator and a Noga dial gage holder. It was concluded that this effect could be neglected in the subsequent force modeling.

The cutting force measurements were repeated two times for workpieces 1 and 2 and three times for workpieces 3 and 4 in order to establish a statistical mean and confidence intervals for the cutting force coefficients between and workpiece tool combinations.

5. Experimental results

In this section experimental results are presented. The cutting force coefficients calculated with the nonlinear optimization method over a range of feed per tooth values are compared for the different material states. Figures 5 and 6 show the mean values for tangential and radial cutting force coefficients at the commanded feed per tooth values including 95% confidence intervals for the repeated tests. There does not appear to be an appreciable difference between the pair of as-produced workpieces.



Fig. 5. Tangential cutting force coefficient mean values and 95% confidence intervals for the pair of as-produced workpieces.

Next, repeatability tests were carried out for the pair of annealed workpieces. The tangential and radial cutting force coefficients are provided in Figures 7 and 8, respectively. Once again, there is no significant difference between the tangential cutting force coefficients between the two workpieces. While the

Gomez, Michael; Heigel, Jarred; Schmitz, Tony. "Force modeling for hybrid manufacturing." Paper presented at 46th SME North American Manufacturing Research Conference, NAMRC 46, College Station, TX, United States. June 18, 2018 - June 22, 2018.

confidence intervals for the radial cutting force coefficients do not overlap in every instance, there is a minimal observable difference between the material regions of interest.



Fig. 6. Radial cutting force coefficient mean values and 95% confidence intervals for the pair of as-produced workpieces.



Fig. 7. Tangential cutting force coefficient mean values and 95% confidence intervals for the pair of annealed workpieces.

The results of the repeatability tests displayed in Figures 4-8 for the as-produced and annealed workpieces were next grouped for direct comparison. A total of six cutting force models for the material regions of interest are established. The results are presented in Figures 9 and 10. In nearly every instance, annealing increases the mean values in both the tangential and radial cutting force coefficients. The AM bulk material exhibits the most notable change in cutting force coefficients.



Fig. 8. Radial cutting force coefficient mean values and 95% confidence intervals for the pair of annealed workpieces.



Fig. 9. Comparison of the tangential cutting force coefficients for the as-produced and annealed workpieces. The error bars identify 95% confidence intervals about the mean.

6. Power law fit

It is evident in Figures 9 and 10 that both the tangential and radial cutting force coefficients decrease with increasing feed per tooth (chip thickness). This results in a nonlinear relationship between the cutting force coefficients and feed per tooth [2]. This phenomena occurs when the chip thickness approaches the size of the cutting edge radius [3-4]. As the commanded feed per tooth

794

approaches the same order of magnitude as the cutting radius for the milling insert, the effective rake angle becomes negative. This negative rake angle serves to increase the cutting force coefficients.



Fig. 10. Comparison of the radial cutting force coefficients for the as-produced and annealed workpieces. The error bars identify 95% confidence intervals about the mean.

A cutting insert was sectioned using wire electrical discharge machining (WEDM) to expose the cutting radius. The sectioned insert was then placed in a scanning electron microscope (SEM) to obtain an approximate cutting edge radius of 24.7 µm. An SEM image is displayed in Figure 11.

Based on Figures 9-11, the cutting force model was modified to account for the edge radius effect. In this case, a power law fit can be applied [6]:

$$F_t = A_t b h^{1-p_t} \tag{5}$$

$$F_r = A_r b h^{1-p_r} \tag{6}$$

$$F_a = A_a b h^{1-p_a} \tag{7}$$

where there is a power law relationship on the chip thickness through the p coefficients. The A and pcoefficients are identified by a nonlinear least squares fit to the feed-dependent cutting force coefficients obtained from the nonlinear least squares optimization. Equations 8-10 define the relationship between the original, feed dependent nonlinear optimization model cutting force coefficients and the new A and p coefficients:

$$k_t = A_t f_t^{-p_t} \tag{8}$$

$$k_r = A_r f_t^{-p_r} \tag{9}$$

$$k_a = A_a f_t^{-p_a} \tag{10}$$

where k_t , k_r , and k_a , are the cutting force coefficients obtained from the nonlinear optimization model, A_b A_r , and, A_a are the new coefficients obtained from the power law fit and p_t , p_r , and p_a are positive dimensionless constants with a value less than one.



Fig. 11. SEM image of the sectioned cutting insert (top) and magnified SEM image of the cutting edge radius (bottom).

To determine the uncertainty in the power law fit, a Monte Carlo simulation was completed in which a random value was generated from normal distributions defined by the mean values and uncertainties from the cutting force coefficients. At each simulation iteration, the power law was fit to the sampled cutting force coefficients to determine the A and p fitting parameters. After 1×10^4 iterations, the mean and

Gomez, Michael; Heigel, Jarred; Schmitz, Tony. "Force modeling for hybrid manufacturing." Paper presented at 46th SME North American Manufacturing Research Conference, NAMRC 46, College Station, TX, United States. June 18, 2018 - June 22, 2018.

795

standard deviation in the distribution of fit parameters was calculated. The results of this exercise are displayed in Tables 4 and 5, where the uncertainties correspond to the Monte Carlo simulation standard deviations.

Table 4. Power law coefficients generated from the Monte Carlo simulation for the as-produced workpieces.

As-produced material region	A_t [N/mm ³]	p_t	A_r [N/mm ³]	p _r
AM surface	911.9	0.3734	315.8	0.6627
	±138.6	±0.053	±42.2	±0.046
AM bulk	964.4 ±68.1	0.3574 ±0.021	284.9 ±26.7	0.6872 ± 0.028
Wrought	925.8	0.3716	451.0	0.5819
	±94.4	±0.036	±71.0	±0.049

Table 5. Power law coefficients generated from the Monte Carlo simulation for the annealed workpieces.

Annealed material region	A_t [N/mm ³]	p_t	A _r [N/mm ³]	<i>p</i> _r
AM surface	804.9 ±105.3	0.4297 ± 0.047	308.9 ±40.6	0.6984 ±0.045
AM bulk	1012.6 ±102.2	$\begin{array}{c} 0.3773 \\ \pm 0.038 \end{array}$	407.8 ±79.6	0.6277 ±0.062
Wrought	1039.3 ±161.3	$\begin{array}{c} 0.3562 \\ \pm 0.051 \end{array}$	519.8 ±132.8	0.5520 ±0.077



Fig. 12. Mean values and standard deviations of the coefficient A for as-produced and annealed states grouped by regions of interest.

The power law fit consolidates the feeddependent specific cutting force coefficients to a pair of fit parameters for each material region. Furthermore, by using the fit parameters and Eqs. 8-10, the cutting force coefficient can be estimated based on the selection of the commanded feed per tooth. A graphical representation of the information in Tables 4-5 is displayed in Figures 12 and 13.



Fig. 13. Mean values and standard deviations of the coefficient p for as-produced and annealed states grouped by regions of interest.

7. Conclusions

The purpose of this project was to investigate the cutting forces generated during the machining of additively manufactured metals. Two conditions were considered: as-produced and annealed. In both material conditions, there were various material states of interest and, in each case, an instantaneous force nonlinear optimization method was used determine the cutting force coefficients.

The results showed only minor differences in the coefficients between the various material states. The most significant effect was annealing which increased the mean cutting force coefficients in nearly all instances.

A nonlinear trend of decreasing cutting force coefficients as a function of increasing feed per tooth was identified. Therefore, a power law fit was applied which incorporated the effect of the change in rake angle at low feed per tooth values. New power law fit coefficients were calculated which, again, did not

796

Gomez, Michael; Heigel, Jarred; Schmitz, Tony. "Force modeling for hybrid manufacturing." Paper presented at 46th SME North American Manufacturing Research Conference, NAMRC 46, College Station, TX, United States. June 18, 2018 - June 22, 2018.

exhibit significant difference between the various material states.

To continue this research effort, a tool wear study will be carried out to investigate the possible changes in tool life when machining the different material regions of interest. Additionally, surface profile measurements will be completed to establish the change in surface finish as the tool wear progresses. Finally, AM samples will be polished and etched to investigate the possible change in grain structure due to the annealing process.

Acknowledgements

The authors gratefully acknowledge partial financial support from the UNC Charlotte Center for Precision Metrology Affiliates Program.

References

- Cooke, A.L. and Soons, J.A., 2010, Variability in the geometric accuracy of additively manufactured test parts, 21st Annual International Solid Freeform Fabrication Symposium, Austin, TX.
- [2] Schmitz, T. and Smith, K.S., 2008, Machining Dynamics: Frequency Response to Improved Productivity, Springer, New York, NY.
- [3] Rubeo, M. and Schmitz, T., 2016, Mechanistic force model coefficients: A comparison of linear regression and nonlinear optimization, Precision Engineering, 45: 311-321.
- [4] Coelho, R.T., Braghini Jr., A., Valente, C.M.O. and Medalha, G.C., 2003, Experimental evaluation of cutting force parameters applying mechanistic model in orthogonal milling, Brazilian Society of Mechanical Science & Engineering, 247-248.
- [5] Kienzle, O., 1952, Die Bestimmung von Kräften und Leistungen an spanenden Werkzeugen und Werkzeugmaschinen, In: Zeitschrift des Vereins deutscher Ingenieure, 657-662.

Thermal Performance Evaluation of a Small Duct High Velocity System Using NIST's Net- Zero Energy Residential Test Facility

Khiem Nguyen¹ Kim Hyojin¹ Lisa Ng² W. Vance Payne II²

¹The Catholic University of America, Washington, DC ²Engineering Laboratory, National Institute of Standards and Technology 100 Bureau Drive Gaithersburg, MD 20899

> Content submitted to and published by: Proceedings of the 15th Conference of the International Society of Indoor Air Quality & Climate (ISIAQ) Location: Philadelphia, PA Date: July 22-27, 2018

> > U.S. Department of Commerce Wilbur Ross, Secretary of Commerce



National Institute of Standards and Technology Walter Copan, Director

NGST National Institute of Standards and Technology • U.S. Department of Commerce

DISCLAIMERS

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

Any link(s) to website(s) in this document have been provided because they may have information of interest to our readers. NIST does not necessarily endorse the views expressed or the facts presented on these sites. Further, NIST does not endorse any commercial products that may be advertised or available on these sites.

Thermal Performance Evaluation of a Small Duct High Velocity System in NIST's Net-Zero Energy Residential Test Facility

Khiem Nguyen¹, Hyojin Kim^{1*}, Lisa Ng², Vance Payne²

¹The Catholic University of America, Washington, DC

²National Institute of Standards and Technology, Gaithersburg, MD

**Corresponding email: kiml@cua.edu*

SUMMARY

This study aims to evaluate the performance of a Small Duct High Velocity (SDHV) heat pump relative to a Conventionally-Ducted Heat Pump (CDHP) using one-year of detailed data from the two-story Net-Zero Energy Residential Test Facility (NZERTF) on the campus of the National Institute of Standards and Technology (NIST). The NZERTF is a single-family home that serves as a laboratory with simulated occupancy and scheduled internal loads. During the analysis period, the SDHV and the CDHP were alternately operated every other week in order to compare the two systems under similar weather conditions. The preliminary analysis indicated that the SDHV system might not always provide thermal comfort comparable with the CDHP. Potential reasons are lower supply airflow rates from the SDHV during the heating season in this low-load house, and a tight thermostat deadband range for the CDHP, which resulted in comparatively more frequent on-cycles.

KEYWORDS

Measurement and Verification; Thermal comfort; IEQ; Low-energy building technology

1 INTRODUCTION

Small Duct High Velocity (SDHV) systems have been recognized as a potential retrofit solution for older residential or historic buildings that were not originally designed for central air distribution systems because of their space savings and easier modular installation, requiring less demolition. Another potential benefit is improved thermal comfort due to enhanced dehumidification and air mixing. However, few studies (Baskin and Vineyard, 2003; Poerschke, 2017) have measured and verified SDHV system thermal performance in residential buildings. The existing studies are also based on short-term measurements or focused solely on SDHV performance verification rather than a performance comparison between SDHV and conventionally-ducted systems using the same test house. Therefore, this study evaluated the performance of a SDHV heat pump by comparing it with a Conventionally-Ducted Heat Pump (CDHP) using one-year of detailed building and system data collected from the NIST's Net-Zero Energy Residential Test Facility (NZERTF).

2 METHODS

This study collected one-year of high resolution (i.e., 10-sec or 1-min) data throughout the NZERTF in Gaithersburg, MD from September 2016 to August 2017 along with coincident outdoor weather data. The NZERTF provides a unique platform for comprehensive, accurate measurements to explore various designs, technologies, and control strategies to achieve netzero energy performance (Fanney et al., 2015). The NZERTF is a single-family home that serves as a laboratory with simulated occupancy and scheduled internal loads. During the analysis period, the SDHV and the CDHP were operated alternately every other week in order to compare the two systems under similar weather conditions. This analysis closely followed the methods outlined in the ASHRAE PMP (2010) and Kim and Haberl (2014).

Kim, Hyojin; Ng, Lisa; Nguyen, Khiem; Payne, William. "Thermal Performance Evaluation of a Small Duct High Velocity System Using NIST's Net-Zero Energy Residential Test Facility." Paper presented at Indoor Air 2018, Philadelphia, PA, United States. July 22, 2018 - July 27, 2018.

3 RESULTS AND DISCUSSION

The preliminary analysis of the data revealed that on average, both the SDHV and CDHP systems reasonably maintained the house close to the thermostat set-point temperatures (i.e., 21.1°C in the heating season and 23.9°C in the cooling season). The CDHP system maintained room temperatures in a tighter range than the SDHV system during the cooling season, which was mainly due to a smaller thermostat deadband for the CDHP, i.e., $\pm 0.1^{\circ}$ C. This result was more obvious for the rooms on the first floor where the thermostat was located. It was also found that there were several occasions when measured humidity levels exceeded the upper humidity limit recommended in the ASHRAE Standard 55 (2013) during the cooling season, which was more noticeable when the SDHV system was on standby (scheduled to run but inactive).

When floor- and room-level data were examined, the SDHV system maintained smaller temperature differences during the cooling season. The average floor-to-floor and room-toroom temperature differences in the SDHV operation mode was 1.0°C during the cooling season, which was about half of the temperature differences measured when the CDHP was operating. This meant that the bedrooms on the second floor were consistently warmer than the set-point temperature during the cooling season, which indicates the potential for discomfort in the warm range on the second floor in the CDHP operation mode.

The vertical and horizontal temperature stratification within a single room was also analyzed using the data collected from a 3 x 3 x 3 grid measurement system in a second floor bedroom. The measured vertical temperature stratification in this bedroom was 0.54°C on average in the SDHV operation mode during the heating season, which was higher than the measured vertical temperature stratification when the CDHP was operating (i.e., 0.20°C). Because the SDHV system has a variable stage compressor and variable speed fan, it delivers lower supply airflow rates in this low-load house during the heating season. An unintentional consequence of lower supply airflow rates was slightly poorer air mixing than with the CDHP's intermittent but higher supply airflow rates.

4 CONCLUSIONS

Preliminary analysis indicated that the SDHV system at the NZERTF might not always provide thermal comfort comparable with a CDHP system. Potential reasons are lower air supply airflow rates from the SDHV during the heating season in this low-load house, and a tight thermostat deadband for the CDHP. Additional analyses are planned to confirm the findings.

ACKNOWLEDGEMENT

The authors acknowledge Gabriel Grajewski and the NIST NZERTF team for their assistance and support with data collection.

5 REFERENCES

ASHRAE. 2010. ASHRAE/CIBSE/USGBC Performance Measurement Protocols for Commercial Buildings. Atlanta: ASHRAE, Inc.

ASHRAE. 2013. ANSI/ASHRAE Standard 55. Atlanta: ASHRAE, Inc.

Baskin E, Vineyard E. 2003. In: ASHRAE Transactions 109(1):513-522.

Fanney A. et al. 2015. In: Energy and Buildings 101(0):95-109.

Kim H, Haberl J. 2014. In: Proceedings of the 13th International Conference on Indoor Air Quality and Climate – Indoor Air 2014, Hong Kong, Vol.V:181-188.

Poerschke A. 2017. Building America Subcontract Report DOE/GO-102017-4664. Pittsburgh: IBACOS. Inc.

Kim, Hyojin; Ng, Lisa; Nguyen, Khiem; Payne, William. "Thermal Performance Evaluation of a Small Duct High Velocity System Using NIST's Net-Zero Energy Residential Test Facility." Paper presented at Indoor Air 2018, Philadelphia, PA, United States. July 22, 2018 - July 27, 2018.

A Gateway to Easily Integrate Simulation Platforms for Co-Simulation of Cyber-Physical Systems

Thomas Roth and Martin Burns Smart Grid and Cyber-Physical Systems Program Office National Institute of Standards and Technology Gaithersburg, Maryland

Abstract-Cyber-physical systems (CPS) research leverages the expertise of researchers from multiple domains to engineer complex systems of interacting physical and computational components. An approach called co-simulation is often used in CPS conceptual design to integrate the specialized tools and simulators from each of these domains into a joint simulation for the evaluation of design decisions. Many co-simulation platforms are being developed to expedite CPS conceptualization and realization, but most use intrusive modeling and communication libraries that require researchers to either abandon their existing models or spend considerable effort to integrate them into the platform. A significant number of these co-simulation platforms use the High Level Architecture (HLA) standard that provides a rich set of services to facilitate distributed simulation. This paper introduces a simple gateway that can be readily implemented without co-simulation expertise to adapt existing models and research infrastructure for use in HLA. An open-source implementation of the gateway has been developed for the National Institute of Standards and Technology (NIST) cosimulation platform called the Universal CPS Environment for Federation (UCEF).

Keywords-co-simulation; cyber-physical systems; functional mock-up interface; high level architecture

I. INTRODUCTION

Cyber-physical systems (CPS) consist of interconnected devices that use logical computation to actuate changes on the physical world based on measurements of their environment. These systems lead to improved quality of life and technological advances in critical infrastructures such as personalized health care, smart manufacturing, autonomous vehicles, and the smart grid [1]. A strong assurance case must be made to validate each CPS prior to its deployment because failure has the potential for severe physical consequences that will impact humans and society. However, due to the scale and cost of these systems, it is impractical to prototype each design decision and validation relies heavily on computer simulations. An accurate simulation of a CPS requires the integration of simulators from multiple domains. For example, the common approach to evaluate new control algorithms for the smart grid is to integrate a network simulator with a power system simulator to enable researchers to measure the impact of control algorithms on the physical grid under realistic network conditions [2]. This integration of multiple simulators into a joint simulation is a technique called co-simulation.

IEEE 1516-2010 High Level Architecture (HLA) is one standard for the co-simulation of distributed processes [3]. HLA defines the heterogenous, domain-specific simulators as *federates*, and the collection of federates that comprise the joint simulation as a *federation*. The federates communicate and coordinate using software called the runtime infrastructure (RTI) that implements the common set of services described in the HLA federate interface specification. The most useful service that HLA provides is an abstraction of logical time that allows for time synchronization between simulators with different time semantics.

HLA is a comprehensive standard for distributed simulation that defines a rich application programming interface (API) containing over 168 services [4]. However, HLA is difficult to both learn and use because a conscious decision must be made on which HLA services to implement for each federate. This usability problem is compounded in CPS where often decades of research have gone into models and simulations of various sectors such as smart grid in well-established domain-specific testbeds. Almost all this research infrastructure was built without the vision of a co-simulation standard such as HLA, and it is difficult to imagine these testbeds are readily compatible with the HLA federate interface specification.

This paper introduces a gateway federate that defines a reduced API to perform the minimum services required to function as a HLA federate. It retains the time synchronization and data exchange services from HLA while reducing the API from 168 services to under a dozen methods. An open-source implementation of the gateway was developed in Java using the Portico¹ RTI, but the architecture described in this paper is applicable to other languages and for other RTI. Using this gateway federate, a researcher can readily integrate their existing research infrastructure into HLA for co-simulation of complex CPS with little to no knowledge of the standard.

Section II provides an overview of other co-simulation methodologies and the related work in this area. Section III discusses the gateway federate architecture, and Section IV summarizes which HLA services the gateway federate implements. The gateway life cycle is presented in Section V, and the results of the work are concluded in Section VI.

Burns, Martin; Roth, Thomas. "A Gateway to Easily Integrate Simulation Platforms for Co-Simulation of Cyber-Physical Systems." Paper presented at 2018 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems, Porto, Portugal. April 10, 2018 - April 10,

¹ Portico available at https://github.com/openlvc/portico

II. RELATED WORK

There are two popular co-simulation standards: HLA and Functional Mock-up Interface (FMI) [5]. Unlike HLA where federates are standalone processes that are peers in distributed system, FMI uses a master-slave architecture where the master algorithm imports each simulator as a shared library and makes direct function calls into the slave code. A simulation library in FMI is called a Functional Mock-up Unit (FMU). Each FMU is stored in a ZIP archive that contains either source code or a compiled shared library, an XML file that describes the model inputs and outputs, and other documentation. The FMI standard for co-simulation prescribes the function definitions that each slave must implement to be interoperable with the master. The advantages of FMI over HLA are that it has an intuitive API and FMU are not dependent on a specific RTI implementation. However, FMI is less comprehensive than HLA and is more difficult to use in distributed simulations that require multiple networked computers.

One challenge with using FMI for CPS co-simulation is that it does not provide a re-usable implementation of the master algorithm. A small body of research has investigated how to bridge the two standards by using HLA as the FMI master algorithm, which is equivalent to a federation of FMU federates [6] [7] [8]. This integration makes FMI more viable for CPS co-simulation since the master algorithm would not change between different sectors such as transportation and smart grid. However, FMI still uses shared libraries for the slaves in its architecture and it remains a challenge to use FMI for experiments that involve processes distributed across multiple computers. The distributed process approach of HLA is easier to use for the co-simulation of CPS which can require multiple networked computers for the cases of hardware-inthe-loop simulations and federation between remote testbeds.

There are several ongoing development efforts working on environments that use a combination of graphical domainspecific languages and code generation to ease the production of HLA federates. Vanderbilt University has evolved the Command and Control Wind Tunnel (C2WT) over several major releases to aid in federate development for Java, C++, and multiple simulation engines using a graphical language designed in their Generic Modeling Environment (GME) [9]. The National Institute of Standards and Technology (NIST) released an image of a virtual machine for a variant of C2WT called the Universal CPS Environment for Federation (UCEF) to make federate development more portable and accessible for CPS researchers [10]. A parallel effort at the University of Rome Tor Vergata uses an HLA profile for the Systems Modeling Language (SysML) to go from requirements analysis to co-simulation in a single modeling environment [11]. A more recent effort uses an Eclipse plug-in to design HLA federates in the Parallel Object-Oriented Specification Language (POOSL) [12]. It is worth noting that each of these approaches uses a different RTI implementation and federates produced from one environment will be incompatible with the others. All these platforms focus on the creation of new federates for their supported simulation engines, and none of them provide guidance on how to integrate new simulation engines or research infrastructure into the workflow.

The HLA Development Kit takes a different approach and provides a Java library with a simplified API rather than using a modeling language to ease the development of federates [13]. It uses Java annotations in classes that resemble JavaBeans to define the data structures used for data exchange, and contains libraries and documentation useful for the development of Java federates. However, the requirement to implement classes for each message type is too restrictive for a gateway federate that might not be aware of the federation data model until runtime. This paper adopts a similar approach to the HLA Development Kit and defines the interface and life cycle for a gateway federate made available as a Java library to ease the integration of existing infrastructure into HLA. The gateway described in this paper exposes a small API that resembles a FMU from the FMI standard for co-simulation, but retains the richness and flexible architecture of HLA.

III. GATEWAY ARCHITECTURE

This section describes the architecture of a HLA gateway that exposes an existing model as a federate with an opensource implementation for UCEF². The UCEF Gateway is written in Java for Portico and as such is not available in other languages or for other RTI. However, this paper mentions few implementation-specific details and the presented architecture could be used for the design for another gateway in a different language for a different RTI. The gateway federate was designed to satisfy the following set of requirements:

- Usable without HLA expertise: a significant barrier to the use of HLA for co-simulation is that the standard is complex with limited documentation. A gateway federate must not presume knowledge of the HLA standard to be useful for researchers.
- Easy to integrate new things: a significant gap in the current approaches towards accessible co-simulation is the assumption that all federates will be created within a dedicated platform. This assumption is not true for CPS where significant testbeds already exist in multiple domains such as smart grid that were developed without using co-simulation platforms. A gateway must be able to integrate existing research infrastructure into HLA with minimal effort.
- Agnostic to the federation data model: several gateway applications such as protocol conversions are independent of the data exchanged at runtime. A gateway must not need to be recompiled each time there is a change in the data model that doesn't affect the logic of its execution.

An instance of the gateway federate is a single-threaded Java application that consists of two parts: the gateway library that defines the program control flow and interacts with the RTI, and a user application that implements the set of callback methods defined in the gateway library. Each user application must determine how best to implement these callbacks to interact with the simulator, model, or research infrastructure

² UCEF Gateway available at https://github.com/usnistgov/ucef-gateway

2018.

that should be integrated as a HLA federate. The gateway library yields control to the user application during specific points in the federate life cycle by invoking the callbacks, during which the user application uses the gateway library to perform basic HLA services. Fig. 1 shows the structure of the gateway federate where the shaded areas represent the user application.

The lower dotted line in Fig. 1 represents the service set defined in the HLA standard, while the upper dotted line represents the reduced number of services that are exposed to the user application. Section IV discusses the services exposed by the gateway library, almost all of which relate to the data exchange within a federation. However, although the gateway library eliminates the need to consider most HLA services, the gateway federate is still HLA compliant and can participate in co-simulations with federates produced from the platforms described in Section II. Also, while a developer is shielded from the complexity of the HLA standard when using the gateway library, the user application could still choose to implement the other HLA services if the additional complexity is desired. This architecture satisfies the first requirement that the gateway be usable without extensive knowledge of the HLA standard because the user application only needs to touch the well-documented gateway library.

The minimum requirement for a user application is the implementation of the callback methods-6 in total-defined in the gateway library. These callback implementations can be as complex or simple as required to connect the new hardware or simulation with the gateway library to turn it into a HLA federate. The list of callback methods, as well as when each will be invoked during the gateway life cycle, are discussed in Section V. This addresses the second requirement on ease of integration because there are only 6 integration points where the hardware or simulation interacts with the HLA federation.

To satisfy the third requirement on independence from the data exchanged at runtime, the message structures used in the gateway library are represented entirely by string name-value pairs. A HLA message in the gateway library is structured as a string identifier and a map that points the property names to their byte array values. Data conversion, when necessary, must be performed in the user application through queries to the gateway library about the current object model. Several gateway applications are simple protocol conversions from the HLA message structure to another format such as a JSON string used in a representational state transfer (REST) API. This design decision allows the gateway federate to be data driven through configuration files as opposed to instancespecific compiled code.

This architecture is supported by two configuration files. A JSON configuration file is used to set the basic configuration options related to how to join the HLA federation and how to progress logical time. An extensible markup language (XML) configuration file specifies the structure for the messages that the gateway plans to send and receive during runtime. This XML file is compliant with a format prescribed by the HLA standard called the Federate Object Model (FOM), and is loaded into the gateway federate memory at runtime. Once a gateway federate has been developed through implementation



Fig. 1. The gateway federate architecture where the shaded rectangles must be implemented to integrate existing research infrastructure

of a specific user application, the same executable can be reused without code modification across multiple federations solely through modifications of these two text files.

IV. GATEWAY SERVICES

This section is a primer on HLA that describes the services relevant to the gateway federate. The HLA federate interface specification defines a set of services that are categorized into service categories. The Federation Management services relate to the creation, destruction, and synchronization of federation executions. The Declaration Management services declare publications and subscriptions to different HLA message types, while the Object Management services send and receive specific message instances. The Ownership Management services transfer the right to update data from one federate to another, and the Time Management services relate to the progression of logical time at individual federates. The Data Distribution Management services apply filters to restrict who receives specific messages to reduce the amount of network traffic in the federation. The remainder of this section will describe the services implemented by the gateway federate.

A. Federation Management

A federation is a named group of federates that can be joined using a specific network address. Federates can join and leave an existing federation at any time during its execution, and all joined federates have access to the same set of HLA services. However, one federate must choose to create the federation, and only this federate starts with the permission to delete the federation. The group of federates in the federation execution can be synchronized at named points along the logical time axis called synchronization points. Whenever a federate tries to synchronize on a registered synchronization point, its execution is blocked until all federates in the federation reach the same synchronization state.

The gateway federate has a limited role in the federation management services. It assumes the existence of a single federate designated as the federation manager whose responsibilities include creation of the federation, registration of the critical synchronization points, and determining when an experiment starts by being the last federate to synchronize. As such, the gateway federate cannot create federations and has a largely reactive role in federation management. It assumes the existence of three synchronization points, and will not run unless these synchronization points are registered by the federation manager. The services for joining, leaving, and synchronization are automated by the gateway library, and the user application can choose to ignore the federation management services.

B. Declaration Management

A federation uses the publish-subscribe message pattern for data exchange between federates. Each federate must declare its interests in terms of publications and subscriptions before it can send or receive data. These publications and subscriptions are restricted to the message classes in the current object model of the federation. The gateway federate requires an XML configuration file that describes its publication and subscription interests in the format defined by the HLA Object Model Template (OMT). At initialization, the gateway library will parse this configuration file to automate the declaration management service calls based on which message classes are annotated with publish and subscribe. The gateway configuration will not be passed to the federation for dynamic configuration of the object model, and the gateway federate assumes that the federation was created to include the message classes defined in its configuration file. The user application has no role in declaration management other than loading the gateway library with a configuration file that has the desired publication and subscription interests.

C. Object Management

HLA has two primitive message types: object classes, and interaction classes. An object is a persistent entity that contains one or more variables that change with time. For instance, a transportation simulation might define multiple vehicle objects which have attributes for current speed and position. An interaction is a one-time event with no continuity between two instances of the same interaction class. For instance, a transportation simulation that contains sensors that trigger an alert when a vehicle exceeds a certain speed limit might define an alert interaction. Objects have a life cycle where individual, named instances must be created and maintained over the course of a federation execution, while interactions are ad hoc and can be sent without preparation. Both object updates and interactions can be sent with an optional logical time stamp to ensure the causal ordering of messages given certain time management constraints.

The gateway federate has full support for the sending and receiving of both objects and interactions, either with or without logical time stamps. The gateway library exposes all the services related to object life cycle in its API, and expects the user application to manage its own object instances.

D. Time Management

The time management services control how an individual federate uses logical time within the federation. A federate can set two different flags to control its logical time progression: time constrained, and time regulating. A time constrained federate cannot advance its logical time beyond the current federation logical time. A time regulating federate dictates the pace of logical time progression, and the federation logical time will advance at the rate of the slowest regulating federate. A federate can use all four permutations of time constrained and time regulating, with a federate that is both constrained and regulating operating in lockstep with respect to federation logical time.

HLA defines two services for federates to advance logical time: time advance request, and next event request. A federate cannot advance its local logical time until it receives a time grant from the federation. When a federate receives the time grant depends on whether it is time constrained, and the current state of the set of time regulating federates. A time advance request will step the federate logical time by a fixed value, while a next event request will act as an interrupt signal and update the logical time by the smallest increment necessary to receive the next event in the federation.

The gateway federate is both time constrained and time regulating to operate in lockstep with the federation logical time, and this behavior cannot be reconfigured. It uses time advance requests with a step size that is static for the duration of the federation execution, and does not support the next event request service. All time management services are automated by the gateway library and hidden from the user application, but the user application configures the logical time step size.

V. GATEWAY LIFE CYCLE

The gateway library executes a well-defined life cycle with callbacks to the user application that can be implemented to perform custom functions. During these callbacks, the library exposes a small number of methods that can be invoked by the user application for sending data and querying the federation object model. As described in Section IV, the services relevant to the user application are primarily concerned with object management. Fig. 2 shows the complete gateway life cycle, and Table 1 describes the gateway callback methods.

The life cycle is bounded by three synchronization points named readyToPopulate, readyToRun, and readyToResign. At each of these synchronization points, the gateway federate will block its execution until the federation synchronizes on the listed name. These synchronization points divide the gateway life cycle into three separate stages of initialization, logical time progression, and termination. This section describes the behavior of the gateway federate during each stage.

A. Initialization

Each federation has a set of expected federates, without which the purpose of the federation-often the execution of some experiment-cannot be realized. It is assumed that the federate designated as the federation manager keeps track of the names and current states of all expected federates. When



Fig. 2. The gateway federate life cycle where the shaded rectangles are callbacks that must be implemented by the user application

the federation manager detects that all expected federates have joined the federation, it will achieve the synchronization point *readyToPopulate*. The federates will then initialize themselves against their peers in an initialization stage that proceeds until all federates have reached *readyToRun*. The initialization stage of the life cycle covers the period from when the gateway federate is instantiated until the federation execution achieves the synchronization point for *readyToRun*.

The gateway library does most of its automated functions when it joins the federation. It parses the FOM configuration file to build its internal object model, calls all the declaration management services to create the required publications and subscriptions, and invokes the time management services to set itself as time constrained and time regulating. It then yields control to the user application for two different phases of initialization. The first phase, self initialization, allows the user application to initialize or connect to the research infrastructure being integrated into HLA, and to register any object instances that are needed to expose the infrastructure's internal state to the federation. The second phase, peer initialization, lets the user application exchange initial values with its peers in the federation before the first logical time step.

B. Time Progression

After the *readyToRun* synchronization point, the gateway federate enters the time progression stage of its life cycle. This stage will continue until either the federation manager decides to destroy the federation execution, or the user application requests to leave the federation. The only automated function of the gateway library during this stage is synchronization of the *doTimeStep* callback to the logical time progression of the LA federate. However, the user application must implement almost all the object management services during this stage.

The interactions and object updates received during a time step will be sent to the user application one at a time using the gateway callback methods. While Fig. 2 represents receiving data as a single state, different callbacks are used for receiving interactions and for receiving objects as listed in Table 1. All received interactions and object updates will be handled at the start of the logical time step. Then the *doTimeStep* callback will be invoked with the current logical time to allow the user application to implement how a time step affects the research infrastructure being integrated using the gateway federate. If the user application expects to receive a message during the middle of its time step—which is possible if the message was sent without a time stamp—the gateway library exposes a tick method which can be used to poll for new messages.

C. Termination

The termination stage of the life cycle starts when either the user application requests to exit, or the federation manager decides to destroy the federation. If the federation manager causes the termination, then the gateway federate will wait for the readyToResign synchronization point which is achieved when all federates have transitioned into a terminating state. Otherwise, the readyToResign synchronization point will be skipped under the assumption that the federation will continue to execute after a termination request initiated by the user application. Both cases will invoke one final callback method for the user application before the Java application exits.

The gateway library automates the process of cleaning up all registered object instances and leaving the federation. When the terminate callback is invoked, the gateway federate will no longer be a member of a HLA federation, and none of the methods defined in the gateway library will be usable. The purpose of the terminate callback is for the user application to perform any necessary cleanup functions, such as disconnect or shutdown procedures required by the connected research infrastructure, before the Java application exits.

VI. CONCLUSION

CPS research and development activities use co-simulation as one technique to validate increasingly complex systems that are integrating more and more domains. HLA is one co-simulation standard that considers federates as distributed, peer processes rather than shared libraries. However, HLA is a comprehensive standard that can be difficult to integrate with existing research infrastructure such as testbeds due to its rich service set. This paper proposes the architecture for a gateway federate with a simple interface that resembles FMI to address this integration challenge. It identifies the minimum set of

TABLE I. GATEWAY LIFE CYCLE CALLBACKS

callback	description
initializeSelf	Perform any initialization that depends on the existence of a joined federation, such as registering new object instances.
initializeWithPeers	Perform any initialization that depends on communication with other federates, such as the exchange of initial values.
receiveInteraction	Handle a single received interaction.
receiveObject	Handle a single received object reflection.
doTimeStep	Perform the logic that must be executed each logical time step, such as updating the states of registered objects.
terminate	Perform any cleanup, such as closing output files and stopping simulations.

"A Gateway to Easily Integrate Simulation Platforms for Co-Simulation of Cyber-Physical Systems." Paper presented at 2018 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems, Porto, Portugal. April 10, 2018 - April 10,

2018

services required to participate in a HLA co-simulation, and proposes a federate life cycle based on the assumption of a designated federation manager. An instance of the gateway federate for UCEF was discussed which uses a Java library with the minimum set of methods required to participate and exchange data in a federation. Through implementation of the callback functions defined in the gateway library, new simulators and new physical resources such as domain-specific testbeds can be integrated into HLA federations.

An implementation of the gateway federate for a specific user application requires a smaller number of integration activities-and significantly less co-simulation expertise-than a full implementation of the HLA standard. Several common activities for the integration of a new simulator are:

- Define how to represent the state of the simulation model as a set of HLA object and interaction classes;
- Determine how to progress simulation time during the . fixed logical time steps of the gateway federate;
- Launch or connect to the simulation model during the initialization stage of the gateway federate life cycle;
- Implement the gateway federate callback methods to exchange inputs and outputs with the model;
- Terminate or disconnect from the simulation model during the termination stage of the life cycle.

The open-source implementation of the UCEF Gateway federate is being continuously refined to support the integration of new applications. Future development efforts for the UCEF Gateway will focus on how to best integrate both real-time hardware-in-the-loop and entire research testbeds to make it easier to run federations between geographically remote facilities. The UCEF Gateway will also be updated to support both the 2010 Evolved version of the HLA standard, and the upcoming release of Portico v2.2.0 to take advantage of several new quality of life features.

NIST also has multiple ongoing research activities that are using the UCEF Gateway to integrate different simulation engines into the Portico RTI. These simulation engines include the EnergyPlus[™] building energy simulator, the GridLAB-D[™] power distribution system simulator, and the LabVIEW design and development environment. NIST has additional activities to incorporate both the Cucumber[™] testing environment and an experiment scripting language called courses of actions into HLA as implementations of the gateway federate. All these activities are based on the architecture described in this paper, with most choosing to implement the user application as a TCP/IP server using a custom socket protocol.

ACKNOWLEDGMENT

Official contribution of the National Institute of Standards and Technology; not subject to copyright in the United States.

Certain commercial products are identified in order to adequately specify the procedure; this does not imply endorsement or recommendation by NIST, nor does it imply that such products are necessarily the best available for the purpose.

REFERENCES

- [1] E. Griffor, C. Greer, D. Wollman and M. Burns, "Framework for cyberphysical systems: volume 1, overview," National Institute of Standards and Technology, Gaithersburg, MD, Special Publication. NIST-SP-1500-201, June 26, 2017, doi: 10.6028/NIST.SP.1500-201
- [2] IEEE Task Force on Interfacing Techniques for Simulation Tools, "Interfacing power system and ICT simulators: challenges, state-of-theart, and case studies," in IEEE Transactions on Smart Grid, vol. 9, no. 1, pp. 14-24, Jan. 2018. doi: 10.1109/TSG.2016.2542824
- IEEE Standard for Modeling and Simulation (M&S) High Level [3] Architecture (HLA)-- Framework and Rules, IEEE Std 1516-2010, Aug. 18 2010. doi: 10.1109/IEEESTD.2010.5553440
- [4] IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)-- Federate Interface Specification, IEEE Std. 1516.1-2010, Aug. 18 2010. doi: 10.1109/IEEESTD.2010.5557728
- Functional Mock-up Interface for Model Exchange and Co-Simulation [5] 2.0, July 2014. [Online] Available: http://fmi-standard.org
- M. U. Awais, P. Palensky, A. Elsheikh, E. Widl and S. Matthias, "The high level architecture RTI as a master to the functional mock-up interface components," 2013 International Conference on Computing, Networking and Communications (ICNC), San Diego, CA, 2013, pp. 315-320. doi: 10.1109/ICCNC.2013.6504102
- A. Garro and A. Falcone, "On the integration of HLA and FMI for [7] supporting interoperability and reusability in distributed simulation," in Proceedings of the Symposium on Theory of Modeling & Simulation: DEVS Integrative M&S Symposium (DEVS '15), Alexandria, VA, 2015, pp. 9 - 16.
- F. Yılmaz, U. Durak, K. Taylan, and H. Oğuztüzün, "Adapting [8] functional mockup units for HLA-compliant distributed simulation," in Proceedings of the 10th International Modelica Conference, Lund, Sweden, Mar. 2014, pp. 247 - 257. doi 10.13140/2.1.4189.7765
- G. Hemingway, H. Neema, H. Nine, J. Sztipanovits and G. Karsai, [9] "Rapid synthesis of high-level architecture-based heterogeneous simulation: a model-based integration approach," in Simulation, vol. 88, no. 2, pp. 217 - 232, Feb. 2012. doi: 10.1177/0037549711401950
- [10] T. Roth, E. Song, M. Burns, H. Neema, W. Emfinger and J. Sztipanovits, "Cyber-physical system development environment for energy applications," in ASME 11th International Conference on Energy Sustainability, Charlotte, NC, 2017. doi 10.1115/ES2017-3589
- [11] P. Bocciarelli, A. D'Ambrogio, A. Giglio and E. Paglia, "A modeldriven framework for distributed simulation of autonomous systems," in Proceedings of the Symposium on Theory of Modeling & Simulation: DEVS Integrative M&S Symposium (DEVS '15), Alexandria, VA, 2015, pp. 213-220.
- [12] T. Nägele and J. Hooman, "Rapid construction of co-simulations of cyber-physical systems in HLA using a DSL," 2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Vienna, 2017, pp. 247-251. doi: 10.1109/SEAA.2017.29
- [13] A. Falcone, A. Garro, A. Anagnostou, N. R. Chaudhry, O. A. Salah and S. J. E. Taylor, "Easing the development of HLA federates: the HLA development kit and its exploitation in the SEE project," IEEE/ACM 19th International Symposium on Distributed Simulation and Real Time Applications (DS-RT), Chengdu, 2015, pp. 50-57. doi: 10.1109/DS-RT.2015.18

Burns, Martin; Roth, Thomas. "A Gateway to Easily Integrate Simulation Platforms for Co-Simulation of Cyber-Physical Systems." Paper presented at 2018 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems, Porto, Portugal. April 10, 2018 - April 10,

Spring Technical Meeting Eastern States Section of the Combustion Institute March 4-7, 2018 State College, Pennsylvania

Predicted Burning Velocities of C₁ and C₂ Hydrofluorocarbon Refrigerant Flames with Air

*Gregory Linteris*¹, *Valeri Babushok*¹

¹Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, 20899

Abstract: Due to their high global warming potentials, many existing working fluids for heating, cooling and refrigeration equipment are being phased out. Their replacements will often be flammable or slightly flammable, and the burning velocity of refrigerant-air mixtures is being used as a metric to rank their flammability. To allow industry to estimate the flammability of new blends of agents, predictive tools for the burning velocity of refrigerants are being developed, which requires a kinetic mechanism. The NIST hydrofluorocarbon HFC mechanism was developed 20 years ago to describe hydrocarbon-air flames with added trace amounts of hydrofluorocarbon fire retardants (primarily CH₂F₂, CF₃H, CF₄, C₂H₂F₄, C₂HF₅, and C₂F₆). The mechanism has been modified to include new HFC compounds, more recent rate data, and rate data for new species. The modified mechanism is used to predict steady, planar, 1D, unstretched burning velocities for mixtures of each refrigerant with air, over a range of fuel-air equivalence ratio ϕ , for comparison with the available experimental data. Agreement is good, and major kinetic pathways and radical populations are explored to uncover the general reaction properties of these new flames.

Keywords: Refrigerant flammability, burning velocity, low-GWP refrigerants, hydrofluorocarbons

1. Introduction

Existing refrigerant working fluids in vapor-compression heating/cooling equipment have high global warming potential (GWP), and are being phased through international treaties (i.e., the Kigali Agreement, an addendum to the Montreal Protocol.). Low-GWP replacements have been developed, primarily by adding double bonds or hydrogens to the molecules, which makes them break down in the troposphere. Unfortunately, these properties also make them more flammable. Flammable refrigerants are a new challenge for the Heating, Ventilation, and Air-Conditioning industry, and new building standards are required for the safe use of the new compounds. Burning velocity has been adopted as part of the standard to characterize the new refrigerants.

To meet the challenges of high efficiency, good volumetric capacity, low toxicity, zero ozone depletion potential, low GWP, and low flammability, industry will use blends of compounds. Analytical methods exist for optimizing the blends for all these properties except for flammability. To allow industry to estimate the flammability of new blends of agents, predictive tools for the burning velocity of refrigerants are being developed, which requires a kinetic mechanism. The NIST hydrofluorocarbon (HFC) mechanism was developed 20 years ago to describe the addition of HFC fire suppressants to hydrocarbon-air flames. It is currently being updated and extended to apply to new refrigerants added at high concentrations in air, and will likely require additional

reactions and species. As a first step in this process, the existing NIST HFC mechanism was applied to predict burning velocities of some pure C_1 and C_2 HFC compounds in air. Agreement was initially good for some compounds and poor for others. Consequently, some improvements were made, including addition of new HFC intermediates and their reaction rates, more recent rate data, and updated thermodynamic data. The modified mechanism is then used to predict steady, adiabatic, planar, 1D, unstretched burning velocities for mixtures of each refrigerant with air, over a range of fuel-air equivalence ratio ϕ , for comparison with experimental data in the literature. The compounds of interest are the saturated C_1 and C_2 HFC compounds with varying fluorine to hydrogen atom ratio in the reactant stream, which is characterized by F/(F+H), or the (moles F)/(mole F + moles H).

2. Kinetic Modeling

The starting kinetic model is from the NIST C_1 - C_2 HFC model [1, 2], which was combined with GRI2.0 [3]. The mechanism has been updated and expanded to include larger HFC's, $C_3H_2F_3Br$, and $C_6F_{12}O$, as described in Ref. [4]. Since the original work, a rather large amount of new kinetic data on the reactions of fluorine containing species has been published. The kinetic model has been modified to include new species and recent reaction rate data. The thermodynamic data for fluorine-containing species has been also updated using the data of Burcat et al. [5]. The CHEMKIN and Cantera packages were used for combustion equilibrium calculations and for modeling of laminar flame propagation in mixtures of the C_1 - C_2 -HFCs in dry air. It should be noted that the present kinetic model should be considered as a starting point for further development and refinement. Numerous changes to both the rates and the reactions may be made once a variety of experimental data and theoretical results are available for testing the mechanism.

3. Results and Discussion

Figure 1 and Figure 2 show, for a range of stoichiometries, the adiabatic flame temperature T_{aft} (upper curves), and the steady, adiabatic, 1D, planar, laminar, unstretched burning velocities S_u^0 (lower curves) calculated with the Sandia EQUIL and PREMIX codes. The figures also show the experimental data (points) of Takizawa and co-workers from outwardly propagating spherical flames in constant volume and constant pressure experiments [6-8]. Table 1 summarizes for each compound the peak T_{aft} and S_u^0 from the experiments and simulations. Also shown are the stoichiometric volume fraction of each compound and the ratio of fluorine to hydrogen atoms, expressed as F/(F+H).

The stoichiometric concentration of the agents is generally higher than that for hydrocarbons, and increases as the fluorine loading in the molecule increases. The temperatures are similar for the flammable HFCs (R161, R41, R152, R152a, R143, R143a), with the peak flame temperatures in the range 2100 K <= T_{aft} <= 2300 K, which is comparable, sometimes higher, than that of hydrocarbons. For the non-flammable refrigerants (R23, R134a, R134, and R125), the peak T_{aft} is somewhat lower (< 1960 K). As illustrated, the agreement in the measured and predicted burning velocity is reasonable for most refrigerants. For R143 and R143a, the simulations predict the peak burning velocity fairly well, although they show a peak value at leaner values of ϕ than measured in the experiments. It should be noted that agreement between predictions and measurements for fluoromethane (CH₃F) and fluoroethane (C₂H₅F) using the original NIST model was poor. Modification of the enthalpy of formation for these compounds and kinetic data for several reactions was required to improve the predictions. Nonetheless, it should also be noted that stretch

and radiation have not been included in the simulations or in the reduction of the experimental data, and these might affect the results [9].



CHxFy, volume fraction

Figure 1: Burning velocity (left scale) and adiabatic flame temperature (right scale) for C₁ hydrofluorocarbons in air.



Figure 2: Burning velocity (left scale) and adiabatic flame temperature (right scale) for C₂ hydrofluorocarbons in air.

In Table 1, burning velocities and adiabatic flame temperatures are listed from highest value of S_u^0 to lowest; both T_{aft} and S_u^0 decrease with increasing fluorine loading in the refrigerant. This is also shown in Figure 3, which presents T_{aft} and S_u^0 as a function of the fluorine loading the system. Also, the asymmetrical isomers tend to be less flammable than the symmetrical ones, having both lower T_{aft} and S_u^0 .

Using the calculated flame structures, the variation in radical concentrations with fluorine loading is determined. Figure 4 shows the sum of the peak volume fraction of chain-carrying radicals (H, O, and OH) as well as that for F-containing radicals, as a function of the fluorine loading. As illustrated, the former drops off rapidly as the number of H atoms in the system becomes close to that of F atoms F/(F+H)=0.5, while the F radicals steadily increase. Hence, at higher fluorine loading, the chemistry becomes dominated by fluorine-containing radicals. Figure 5 shows that with increased fluorine loading, the volume fraction of both F-atoms (at the point of maximum OH volume fraction) and sum of the peak for F-containing radicals increases steadily, as does the equilibrium F atom volume fraction, which becomes higher than the value in the flame zone. For these flames equilibrium F atom volume fraction can be on the order of 1 %.

Refrigerant	Formula	T _{aft} K	<i>S^{u⁰}</i> max (Expt.) cm/s	Su ⁰ max (Calc.) cm/s	X _{stoic.} %	F/(F+H)
<u>Flammable:</u>						
R-170	C_2H_6	2265	40.9	43.1	5.66	0.00
R161	C_2H_5F	2265	38.3	41	6.54	0.17
R-50	CH4	2230	36.5	38.6	9.5	0.00
R-152	CH_2F-CH_2F	2278	30.1	32.1	7.75	0.33
R-41	CH₃F	2273	28.3	27.2	12.3	0.25
R-152a	CH ₃ -CHF ₂	2220	23.6	24.9	7.75	0.33
R-143	CH_2F-CHF_2	2248	13.1	13.7	9.5	0.50
R-32	CH_2F_2	2207	6.7	7.3	17.4	0.50
R-143a	CH_3 - CF_3	2115	7.1	6.1	9.5	0.50
<u>Non-Flammable:</u>						
R-134	CHF ₂ -CHF ₂	1958		4.6	12.3	0.67
R134a	CH_2F-CF_3	1927		1.8	12.3	0.67
R-125	CHF_2-CF_3	1793		1.56 (at 400K)	17.4	0.83
R-23	CHF₃	1713		0.57 (at 400K)	29.6	0.75
R-116	C2F6	1389		0	29.58	1.00

Table 1. Burning velocities and adiabatic combustion temperatures for stoichiometric refrigerant	[-
air mixtures (initial temperature 298 K, 1 bar).	

4

Babushok, Valeri; Linteris, Gregory. "Predicted Burning Velocities of C1 and C2 Hydrofluorocarbon Refrigerant Flames with Air." Paper presented at Spring Technical Meeting? Eastern States Section of the Combustion Institute, Station College, PA, United States. March 4, 2018 - March 7, 2018.



Figure 3: Burning velocity (left scale) and adiabatic flame temperature (right scale) as a function of the ratio of F atoms to F plus H atoms in the initial refrigerant-air mixture.



Figure 4: Volume fraction of total chainbranching radicals (H+O+OH) and F atoms as a function of the ratio of F atoms to F plus H atoms.

Figure 5: Volume fraction for fluorine radicals (F-radicals) and F atoms in the flame reaction zone, and for F atoms at equilibrium as a function of the ratio of F to F plus H atoms.

Babushok, Valeri; Linteris, Gregory. "Predicted Burning Velocities of C1 and C2 Hydrofluorocarbon Refrigerant Flames with Air." Paper presented at Spring Technical Meeting? Eastern States Section of the Combustion Institute, Station College, PA, United States. March 4, 2018 - March 7, 2018.

4. Conclusions

The laminar, planar, 1D, adiabatic burning velocity of saturated C_1 and C_2 HFC refrigerants (R41, R32, R161, R152, R152a, R143, R143a, R134, R134a were calculated using the NIST HFC mechanism available in the literature and compared to existing experimental data. The predictions were in significant disagreement for CH_3F and C_2H_5F , and mild disagreement for other compounds. Consequently, the NIST HFC mechanism was modified with additional reactions, using more recent rate data in the literature, and with updated thermodynamic properties. After the changes, the agreement for these refrigerants, over a range of equivalence ratio, was good. The mechanism was then used to examine the properties of the refrigerant-air flames. Adiabatic temperatures of the refrigerant-air flames were comparable to, and sometimes higher than, similar hydrocarbons, and were lower for the non-flammable refrigerants. Burning velocity and flame temperature decreased as the fluorine to hydrogen ratio in the reactants increased. The symmetrical isomers of the fluoroethanes (R152, R143, R134) have higher adiabatic flame temperature and laminar burning velocity than the asymmetrical isomers (R152a, R143a, R134a), consistent with differences in enthalpies of formation. Analysis of the flame structures revealed that with increasing fluorine to hydrogen ratio, the chain-branching radical concentrations in the flame decrease, and fluorine-containing radicals, particularly F atom, increase. At high enough F/H ratio, the F atom equilibrium values are even higher than those in the flame zone.

5. Acknowledgements

This work was supported by the Buildings Technologies Office of the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy under contract no. DE-EE0007615 with Antonio Bouza serving as Project Manager.

6. References

- [1] D. R. Burgess, M. R. Zachariah, W. Tsang, P. R. Westmoreland, Thermochemical and chemical kinetic data for fluorinated hydrocarbons, Prog. Energy Combust. Sci. 21 (1995) 453-529.
- [2] D. Burgess, M. R. Zachariah, W. Tsang, P. R. Westmoreland, Thermochemical and Chemical Kinetic Data for Fluorinated Hydrocarbons, Report No. NIST Technical Note 1412, Gaithersburg, MD, 1995.
- [3] G. T. Linteris, L. Truett, Inhibition of premixed methane-air flames by fluoromethanes, Combust. Flame 105 (1996) 15-27.
- [4] V. I. Babushok, G. T. Linteris, D. R. Burgess, P. T. Baker, Hydrocarbon flame inhibition by C3H2F3Br (2-BTP), Combust. Flame 162 (2015) 1104-1112.
- [5] E. Goos, A. Burcat, B. Ruscic, Extended third millennium thermodynamic database for combustion and air-pollution use with updates from active thermochemical tables. <<u>Ftp://ftp.technion.ac.il/pub/supported/aetdd/thermodynamics/BURCAT.THR</u> > (August 2012).
- [6] K. Takizawa, A. Takahashi, K. Tokuhashi, S. Kondo, A. Sekiya, Burning velocity measurement of fluorinated compounds by the spherical-vessel method, Combust. Flame 141 (2005) 298-307.
- [7] K. Takizawa, A. Takahashi, K. Tokuhashi, S. Kondo, A. Sekiya, Burning velocity measurement of HFC-41, HFC-152, and HFC-161 by the spherical-vessel method, J Fluorine Chem 127 (2006) 1547-1553.
- [8] K. Takizawa, S. Takagi, K. Tokuhashi, S. Kondo, M. Mamiya, H. Nagai, Assessment of Burning Velocity Test Methods for Mildly Flammable Refrigerants, Part 1: Closed-Vessel Method, ASHRAE Trans. 119 (2013) 243-254.
- [9] R. Burrell, J. L. Pagliaro, G. T. Linteris, Effects of stretch and thermal radiation on difluoromethane-air burning velocity measurements in constant volume spherically expanding flames, Proc. Combust. Inst. 37 (2018) submitted.

Babushok, Valeri; Linteris, Gregory. "Predicted Burning Velocities of C1 and C2 Hydrofluorocarbon Refrigerant Flames with Air." Paper presented at Spring Technical Meeting? Eastern States Section of the Combustion Institute, Station College, PA, United States. March 4,



FProceedings of the Annual Stability Conference Structural Stability Research Council Baltimore, Maryland, April 10-13, 2018

Stability of Steel Structures at Elevated Temperatures: A Hybrid Fire Testing Approach

Ana Sauca¹, Chao Zhang², Mina Seif³

Abstract

The response of structural systems to fire loads is typically assessed through performing 'standard' tests under constant mechanical boundary conditions. Such tests are usually performed on the individual elements. Full scale tests showed differences in their behavior compared to the behavior of the individual members tested in standard conditions. Full-scale tests of structural systems remain impractical due to physical, economic, and time constraints. The goal remains to have the capability to predict the behavior of a full-scale test though experimentally testing individual structural members. A promising approach to that problem is "Hybrid Fire Testing (HFT)" where a subset of the structural system (Physical Substructure PS), is physically tested, while the remaining structure (Numerical Substructure NS), is simultaneously numerically analyzed. PS represent the parts of the structure with higher behavioral uncertainty, while the NS represent the parts which can be numerically modelled with high confidence. During the test, the mechanical boundary conditions on the PS and NS are continuously updated. Certain challenges are unique to HFT due to the continuous fire exposure which induces continuous thermal expansions. This paper describes a recent virtual HFT study performed on a ten-story multi-span steel frame assembly, with the focus on the structural stability and interface equilibrium and compatibility between the substructures. In the early stage of this work, the HFT study is done in a virtual environment, where the PS is also modeled numerically as a proof of concept. As part of the study, a traveling fire analysis was performed on the building and results highlighted the importance of considering the performance of the structure as a whole assembly, and showed that individual member standard testing could be unsafe. The paper also describes some of the challenges that are unique to the HFT, and approaches to overcome them.

1. Introduction

The structural engineering industry is rapidly changing. Structures with unique functionalities and scales are being designed, and new materials and technologies are constantly emerging. This is leading the industry into transitioning towards performance-based design methodologies, which require structural engineers to assess the performance of such structures, especially under

Sauca, Ana; Seif, Mina; Zhang, Chao. "Stability of Steel Structures at Elevated Temperatures: A Hybrid Fire Testing Approach." Paper presented at Annual Stability Conference / NASCC: The Steel Conference, Baltimore, MD, United States. April 10, 2018 - April 13, 2018.

¹ Guest Researcher, National Institute of Standards and Technology NIST, <ana.sauca@nist.gov>

² Guest Researcher, National Institute of Standards and Technology NIST, < chao.zhang@nist.gov >

³ Research Engineer, National Institute of Standards and Technology NIST, < mina.seif@nist.gov>

extreme loading conditions. Such structures and materials require testing under these extreme loads in order to validate and calibrate their design modeling approaches.

Fire tests are performed to understand the behavior of structures exposed to fire conditions. Generally, the fire tests are performed on individual elements, with no restraint, neglecting the restraining effect of the rest of the structure, e.g., Tondini et al. (2013). Entire structures have been tested as presented by Lennon (2003), Newman et al. (2000), Armer et al. (1994), but the high cost of such tests limits the ability to conduct them. Full-scale tests of structural systems remain impractical due to physical, economic, and time constraints. However, the goal remains to have the capability to predict the behavior of a full-scale structure though experimentally testing individual structural members. Maintaining the practicality of testing only parts of the structure, but at the same time considering the overall global behavior of the structure as a whole system, Hybrid Fire Testing (HFT) methodology appears to be a promising approach.

Hybrid testing in general is an approach based on a sub-structuring procedure, where a subset of the structural system (Physical Substructure (PS)), is physically tested, while the remaining structure (Numerical Substructure (NS)), is simultaneously numerically analyzed. PS represents the parts of the structure with higher behavioral uncertainty, while the NS represents the parts which can be numerically modelled with high confidence. During the test, the mechanical boundary conditions on the PS and NS are continuously updated.

Hybrid testing can potentially be applied under different loading conditions such as wind, blast, impact, waves, fire, traffic and seismic events. Although it can potentially be applied for such a variety of loading conditions, it has only been widely implemented in the seismic field since early 1970s. The implementation of the method developed from the seismic field to the fire field remains a challenge and only a few HFT have ever been performed.

Hybrid testing under fire conditions in particular has proven to be necessary for many reasons. Globally, the thermal expansion of the structural elements exposed to fire induces major forces that affect the structural behavior during the fire test. The forces are dependent on the boundary conditions adopted in the test. Locally, specific phenomenon occurs for some materials when exposed to fire such as the spalling of the concrete. This phenomenon is very challenging to be numerically modeled. Therefore, fire tests in the realistic conditions are required to better understand the overall structural behavior, and HFT has the potential to provide insight into behavior of structures exposed to fire, under their real boundary conditions.

Usually, standard fire exposure (ASTM E119-16a (2016)) is adopted when studying structural behavior under fire loads. More research is currently focused on studying the behavior of structures exposed to natural fires (Sauca et al. 2015). Zhang et al. (2015, 2013a, 2013b) investigated the behavior of a steel beam under localized fire exposure. These studies showed that under the effect of temperature gradient, the behavior of steel members when exposed to localized fires may be very different from when exposed to the standard fire. Moreover, the failure temperature of steel members exposed to localized fires may be much lower than when exposed to the standard fire. Rackauskaite et al. (2017a, 2017b) presented the differences in the behavior of a multi-story steel frame exposed to travelling fires and Eurocode fires need to be

considered to ensure a safe building design. Different structural responses for the same structure might be triggered depending on the type of the considered fire exposure. In case of uniform fires, higher axial forces were observed compared with the case of traveling fires, where larger displacements at different times and locations were observed. The study also showed that the highest stresses developed in the fire floors adjacent to the unheated floors while in the intermediate fire floors, the stresses were significantly lower. Peak compressive and tensile axial forces developed in the fire floors adjacent to the cooler floors, and in the cooler floors, respectively. This shows that considering designed (realistic) fire scenarios when analyzing the behavior of structures is necessary for safe design. The study also highlighted the effect of the unheated part of the structure on the performance of the remaining structure exposed to fire.

Therefore, in order to predict the real behavior of the tested structural member under specific conditions, the effect of the remaining structure should be considered, which in turn could be exposed to fire. Thus, hybrid fire testing is a promising approach to reach this objective, i.e., perform tests on individual structural members while simultaneously considering the effect of the surrounding members. However, despite the large amount of existing information related to hybrid testing in other fields, the application in the fire field is not straightforward, and has its unique challenges. This is due to the main characteristics of HFT which require the tests to be performed in real time. The effect of the fire on the materials is immediate and continuous and the thermal expansion requires modification of the interface boundary conditions in real time, as discussed in the following section.

2. Background and specific HFT challenges

The main concept of hybrid simulation is to combine the advantages of individual member testing and full-scale testing, meaning that an individual member(s) or sub-structure will be tested while accounting for the remaining structure's contribution. The surrounding can be modelled aside, via nonlinear numerical software or a predetermined matrix, if elastic behavior of the surrounding is expected (see Sauca 2017) for a detailed discussion). The tested substructure is referred to as the physical substructure (PS), while the surrounding (remaining) structure is referred to as the numerical substructure (NS). The approach ensures equilibrium and compatibility between the PS and the NS over the duration of the test. At each time step, data (displacements or forces) are measured at the substructure interfaces. Due to the fire exposure of the PS, equilibrium and compatibility at the interface are generally no longer satisfied at the end of the time step. To restore the equilibrium and compatibility at the interface, new data are computed (forces or displacements) and are imposed to the substructures, based on the measured data from the previous time step. At frequent intervals (time step Δt), the displacements or the forces at the interface are measured from the PS and computed for the NS. There may be an additional delay of time, Δt_{P} , required for the calculation of the NS reaction and for application of the reaction to the PS. The procedure is either called Force Control Procedure (FCP) or Displacement Control Procedure (DCP), when reaction forces or displacements are sent back to the PS.

A specific challenge of the HFT is the continuous fire exposure of the PS (and the NS in some cases). The consequence of the continuous fire exposure is the continuous change of the registered displacements and reaction forces at the boundaries (interface) of the PS and the NS. Therefore, to ensure equilibrium and compatibility between the two substructures, HFT must be

performed in real-time. This is unlike the seismic field (where hybrid testing evolved), in which slow tests, fast tests, real time tests, and smart shaking table tests, are all possible, only requiring a specific algorithm to solve the dynamic equation for each case.

Very few HFT are reported in the literature. Kiel et al. (1989) performed the first HFT, but the publications are not publicly available. The first reported HFT was performed by Korzen et al. (1999), where a column specimen was experimentally tested as part of a simulated building environment. The mode of action between both parts was exemplified on a single degree-offreedom (DoF) basis, i.e., the axial column force is measured and adjusted continuously to the model force, which is represented through a - not necessarily constant – stiffness, in displacement control. Another hybrid fire test was performed by Robert et al. (2009, 2010). In the latter, the PS consisted of a concrete slab whereas the NS was a surrounding one floor concrete building. One axial DoF and two rotational DoFs were controlled. A force-controlled procedure was employed. The behavior of the NS was modelled by a constant predetermined matrix, which had been calculated before the test. Mostafaei (2013a, 2013b) presented the hybrid fire test of the first-floor central column part of a 3D concrete frame. The column was tested in a furnace (PS) while the surrounding was numerically modelled in the non-linear finite element software SAFIR® (Franssen 2005, Franssen and Gernay 2017). For each time step, the interaction between the PS and NS was manually enforced, and axial force in the column was controlled. Part of the NS was also exposed to fire. Whyte et al. (2016) presented a HFT performed on a small-scale steel coupon. The objective was to extend the mechanical hybrid simulation of OpenFresco (2016) and OpenSees (2016) (software commonly used for seismic hybrid simulation) by introducing the temperature DoF and loads. The NS was modeled in OpenSees while OpenFresco was the framework used for the experimental setup and control. Tondini et al. (2016) presented a static solver for HFT, a method based on the finite element tearing and interconnecting (FETI) algorithm. The validation of the method was done in a numerical environment considering a moment-resisting steel frame. Schulthess et al. (2016) present a hybrid fire test performed on a small-scale specimen, i.e., steel coupon specimen, tested in a universal testing machine inside an electric furnace. The NS was analyzed in ABAOUS (2011). Therefore, the only hybrid fire tests performed on full size members (of which the authors are aware) remain the tests reported by Korzen (1999), Robert (2009, 2010) and Mostafaei (2013a, 2013b) and the same methodology was used to perform each of those three tests. In this paper, the method used in those tests will be referred to as the "first generation method".

Stability issues have been observed in the former methodology, i.e., first generation method, as presented by Sauca et al. (2016a) meaning that the methodology was only stable for certain ratios between the NS stiffness and PS stiffness. A new method has been proposed by Sauca et al. (2016b, 2017a, 2017b) and Sauca (2017) which is stable independently from the stiffness ratio between the NS and PS. The stability of the method was numerically proven and validated.

3. A novel HFT approach

In the HFT presented by Korzen (1999), Robert (2009, 2010) and Mostafaei (2013a, 2013b), data (displacements or forces) were measured from the PS and sent to the NS at every time step Δt . The reactions (forces or displacements) of the NS at the interface were then calculated considering only the characteristics of the NS (i.e., disregarding the characteristics of the PS). This calculation can be done using a numerical model or a predetermined matrix for the NS

Sauca (2017). Then, the reactions are sent back to the PS and applied at the interface (boundary) to restore equilibrium and compatibility. There may be an additional delay of time Δt_p requested for the calculation of the NS reactions and for the application of the reactions to the PS. When reaction forces or displacements are sent back to the PS, the procedure is either called FCP or DCP, respectively. Sauca et al. (2016a, 2017b) showed that the later methodology was not always stable. The stiffness ratio between the NS and PS dictated the type of procedure to be used and stability was ensured only for some values of this stiffness ratio. The instability was caused because the stiffness of the PS was not measured during the test and thus was not considered in the calculation of every time step's boundary conditions.

Thus, in the novel method (Sauca et al. 2016b, 2017a, 2017b), the stiffness of the PS is considered in the calculation process and thereby the stability of the solution is ensured regardless of the stiffness ratio between the NS and PS.

A step by step summary of the method for the DCP case is presented below (see Sauca et al. (2017b) for a detailed description).

- a. The interface (boundary) forces and displacements are determined before the start of the test from the analysis of the entire structure.
- b. The PS is placed in the heating condition (direct fire or furnace), and loaded with the exterior loads and interface displacements, while the NS is numerically modeled aside.
- c. Heating of the PS starts. The interface displacements of the PS are blocked for the duration of a time step (DCP) and the reaction forces are measured.
- d. Meanwhile, the corresponding displacements are blocked at the interface of the NS and the reaction forces of the NS are computed.
- e. The measured reaction forces of the PS are compared with the computed reaction forces of the NS. Generally, the equilibrium is not ensured due to the fire effect. The same displacements are applied at the interface of the PS and NS, thus the compatibility is enforced.
- f. To restore the equilibrium, a correction of displacement vector $d\mathbf{u}$ is calculated and applied at the interface of the PS and NS. The calculation includes the vector of out-ofbalance forces $d\mathbf{F}$ (from step *e*). In this step, the stiffnesses of both the PS and the NS are accounted for, according to Eq. (1). This is the main difference from the first generation method (in which only the stiffness of the NS was considered) and the most important contribution that allows ensuring stability of the method.

$$d\mathbf{u}(t_n) = (\mathbf{K}_N + \mathbf{K}_P)^{-1} \, \mathrm{d}\mathbf{F}(t_n) \tag{1}$$

- g. The new calculated displacements are imposed on both the PS and NS. Some time is required to perform the above computations and to adjust the new displacements in the actuators representing the structural interface.
- h. The new imposed displacements will generate new reaction forces in the PS and NS.

5

Steps e-g are repeated until the end of the fire test. As shown in Eq. (1), the stiffness of the PS is used for the correction of displacements. As this stiffness is generally unknown during the HFT, a constant value can be considered and several iterations would normally be needed at each time step to converge to the correct solution. In a fire test, the evolution of temperatures in the PS cannot be held constant during the period requested to perform the iterations at every time step. During the time needed to perform the calculations and for the testing equipment to apply the corrections of displacements, the temperatures are still changing, which in turn changes the stiffness of the PS, the restraint forces, etc. Hence, the convergence process tends to achieve an equilibrium that is constantly changing. As a result, it is not relevant to distinguish between iterations and time steps. Instead, the test can be performed by continuously applying Eq. (1), with a cycling frequency that is as high as possible, which requires computing techniques and testing equipment that has a short response time. Note that the compatibility is continuously achieved, as the same displacements are imposed both on the PS and on the NS at their interface. The purpose of the methodology is thus to constantly adapt these displacements to satisfy equilibrium between the substructures throughout the entire test duration.

4. Analysis

This section describes a virtual HFT study performed on a ten-story multi-span steel frame assembly, with the focus on the structural stability and interface equilibrium and compatibility between the substructures. The HFT study is done in a virtual environment, where the PS is also modeled numerically as a proof of concept. As presented in the previous sections, the HFT found in the literature were performed using the standard fire exposure. This study focuses on applicability of the HFT to real fire scenario exposure cases. Therefore, in the first part of the study, a traveling fire analysis was performed on the ten-story building.

4.1. Fire-structural Analysis

4.1.1 Prototype building

A 2D analysis was performed on a prototype structure that is a 10-story moment resisting steel frame (Fig. 1), which has been designed in accordance to the American Society of Civil Engineers (ASCE 7-02) standard by Sadek et al. (2010). The floor is 45.5 m x 30.5 m (5 bays in the longitudinal and transversal directions). The ground floor columns are 5.3 m high while the rest of the columns are 4.2 m high. The lightweight concrete floor slab is supported by the steel beams. Fig. 1 shows the details of the building. It is noted that this building was studied under traveling and traditional design fire conditions by Rackauskaite et al. (2017a). The composite action between the beams and concrete floor is neglected in the 2D analysis. It was shown in the literature that the tensile membrane action has a beneficial effect to the structural response during fire (Bisby et al., 2013), which if neglected will result in slightly more conservative results.



Figure 1. Plan layout and elevation of the steel frame building (Sadek et al., 2010)

Beam sections of the 2D frame (highlighted in Fig. 1) were W14x22 for all the floors. Column sections of the 2D frame are W18x119, W18x97, and W18x55, for floors 0 through 3, floors 4 through 6, and floors 7 through 9, respectively. ASTM A992 structural steel with yield strength (F_y) of 344.8 MPA is used for all the columns and beams. The dead load on the floor beams and roof beams was 3.64 kN/m², and 2.68 kN/m², respectively. Their corresponding live load was 4.79 N/m² and 0.96 kN/m². The considered load combination is consistent with equation A-4-1 of Appendix 4 presented in the American National Standard ANSI/AISC 360-16. The compartments are labeled as Ci-j, where "i" refers to the floor number and "j" to the compartment number.

4.1.2 Fire scenarios

Two fire scenarios were considered in this case study. Figure 2 shows: (a) the heat release rate curve for each compartment fire; (b) simulated fire curves for fire scenario 1; (c) CFAST model for fire scenario 1; (d) CFAST model for fire scenario 2. The Consolidated Model of Fire and Smoke Transport CFAST (Peacock (2017)) is a two-zone fire model capable of predicting the environment in a multi-compartment structure subjected to a fire. 1D heat transfer finite element (FE) model is used to determine fire spreading.

Fire scenario 1 initiates in a corner compartment and Fire scenario 2 starts in a middle compartment. For each compartment, zone model was used to simulate the fire environment. Fire spread from one compartment to an adjacent compartment occurs when the temperature of the unexposed surface of the connection wall, ceiling, or floor reaches a critical temperature of 139 °C (as per ASTM E119-16a (2016)). Fire in a compartment ignites immediately when temperature of the inner surface of any of the compartment boundaries (walls, ceiling, and floor) reaches 139 °C.


4.1.3 SAFIR model

Analysis of the building exposed to fire scenarios 1 and 2 was carried in SAFIR[®]. To determine the evolution of temperature in the structural members, the thermal analysis was performed prior the structural analysis and then used in the latter.

Thermal analysis of the structural members

Table 601 in the International Building Code (IBC (2015)) specifies a fire resistance rating of 3 hours for the Type IA primary structural frame analyzed here. The fire rating of the floor construction and associated secondary members is 2 hours as presented in the same table. Steel insulation properties are considered as for the high-density perlite (Buchanan, (2001)) with the following characteristics: thermal conductivity of 0.12 W/mK, density of 550 kg/m³ and specific heat of 1200 J/kgK. The thickness of the insulation (for columns and beams) was determined in the FEM so the temperature of the steel members does not exceed the accepted critical temperature for steel of 538 °C (as per ASTM E119-16a (2016)).

The thermal analysis of beams and columns was performed in SAFIR[®], with the following characteristics of steel: heat transfer coefficient of 35 W/m²K, density of 7850 kg/m³ and radiative emissivity of 0.7 (Buchanan, (2001)). The temperature evolution of beams and columns



in different compartments is presented in Figure 3 for fire scenario 1 and in Figure 4 for fire scenario 2.

Structural analysis and the global response

The structural model was built with the following assumptions: the supports for the ground floor columns are assumed to be fixed while the beams and columns are assumed rigidly connected. The steel beams and columns are model using beam elements, and the beam element length was around 0.25 m. The Eurocode (CEN. EN 1991-1-2:2005) temperature dependent material properties are used for columns and beams.

Figure 5 shows the mid-span displacement, the axial force, and the mid-span bending moment versus time, developed within the beams on floor 1 and floor 2. The failure (defined as the

formation of plastic hinges at the beam ends, thus leading to global instability of the building) occurred at about 360 min for both cases, S1 and S2. At the time of failure, temperature in the compartment where the fire initiated was on the descending branch (cooling phase) and the fire had already spread to the compartment above and to the compartments adjacent to the compartment where the fire initiated. Therefore, the figures presented in this section focus on the members of floor 1 and floor 2. The fire initiated in the corner compartment (Floor 1) in the case of fire scenario 1 (S1) and in the central compartment (Floor 1) in the case of the fire scenario 2 (S2). The mid-span displacements were higher in the case of S2 than in the case of S1. The axial restraint for the beam in the central compartment was higher than in the case of the beam in the corner compartment. This is due to the restrain to thermal expansion which is provided by multiple columns, whereas, for the fire scenario S1, the restrain to thermal expansion is provided by only one column on one of the sides. The axial force developed in the beams close to the fire compartments oscillated between compression and tension in the case of S1. In the case of S2, the Floor 1 beams exposed to fire were under compression during the entire duration of the analysis while the Floor 2 beams were starting to be compressed at about 240 min (before the beam temperature of compartment C7 reached the maximum temperature). Larger bending moments developed in the central beam (in the case of S2) than in the corner beam (in the case of S2). The mid-span bending moment developed in the corner beam (in the case of S1) reached the peak value of about 100 kNm, while the central beam (in the case of S2) reached the peak value of about 140 kNm.



10

Sauca, Ana; Seif, Mina; Zhang, Chao. "Stability of Steel Structures at Elevated Temperatures: A Hybrid Fire Testing Approach." Paper presented at Annual Stability Conference / NASCC: The Steel Conference, Baltimore, MD, United States. April 10, 2018 - April 13, 2018.



Figure 5. Developments of mid-span displacements, axial force, mid-span bending moment within beam of the floor 1 and 2 for fire scenario 1 (S1) and 2 (S2)

Figure 6 presents the axial displacement developed within columns of floors 1 and 2 for the fire scenarios S1 and S2. Larger axial displacements developed in the exterior columns, where the forces were lower than in the case of the central columns.



Figure 6. Developments of axial displacements within columns of the floor 1 and 2 for fire scenario 1 (S1) and 2 (S2)

From the global thermal analysis of the 10 story building, it is determined that the beam where the fire initiated (first floor corner beam (S1) and first floor central beam (S2)) is most critical, and therefore will be the focus of the structural analysis described in this section.

Numerical analysis was performed to compare and observe the behavior of the beam when exposed to fire scenarios S1, and S2, in different possible testing configurations. The global behavior of the building was presented in the previous sections and a full scale test would reproduce similar results. Since, as mentioned above, full scale testing is not a cost-effective approach, standard tests are often performed on individual structural members (or sub-structural assemblies). In standard testing, the beam is tested as an individual structural member under constant mechanical boundary conditions. Here, various boundary conditions were considered for the individual beam testing, i.e., f-f (fixed-fixed ends), f-th (fixed-fixed ends with free thermal expansions), f-h (fixed-hinged ends), f-r (fixed-rolling supports), s-s (simply supported beam), h-h (hinged-hinged ends).

Figure 7 presents the mid-span displacements of the beam situated in the compartment where the fire initiated, in different testing configurations. The fire initiates in the first floor corner compartment in the case of fire scenario S1 and in the first floor central compartment for the fire scenario S2. The beams in these two cases are exposed to the same temperature.

Fig. 7 shows that when considering the standard testing approach, regardless of the end conditions, no failure is observed. However, if a full-scale testing were to be performed instead, the failure would occur in the both cases, when exposed to S1 or S2.



Figure 7. The mid-span displacement of the beams (compartment where the fire initiated) in different testing configurations and web temperature when exposed to design fire scenarios S1 and S2

This is a clear example that standard fire testing (with constant mechanical boundary conditions) is not always a safe approach when the design fire scenarios are to be considered. Moreover, the standard fire testing cannot capture the effects of the fire spreading from one compartment to the others. Even if the temperature of the steel elements did not exceed the critical temperature of 550 °C, the failure still occurred by the time the fire spread to the adjacent compartments. The unheated surrounding plays an important role in the behavior of the structure since it restrains the heated parts of the structure to thermally expand, thus inducing critical stresses in some structural members. The real behavior can only be reproduced if a full-scale test is performed, which, as discussed, is not often feasible. Nevertheless, the same global behavior can be reproduced using HFT. This is possible since the surrounding structure is modeled aside (on a computer) and the interaction with the tested member (which is usually an individual structural member or sub-assembly) is considered during the entire fire test. Thus, HFT represents the global behavior of the building (which proved to be critical compared with the individual testing) and at the same time can consider the fire spread in the adjacent compartments.

4.2. Virtual Hybrid Fire Testing

The goal in this section is to investigate the global structural response of the 10 story building when exposed to the fire scenarios 1 through a Virtual Hybrid Fire Test (VHFT). VHFT refers to the case when the tested PS is also modeled in a computer (tested virtually) instead of being physically tested in a lab. Therefore, in a VHFT, the physical substructure PS (tested substructure) is represented by using a finite element model and aside, the numerical substructure

NS (surrounding structure) is also modeled in a finite element software. The VHFT will help validate the available methodologies prior to the real tests. Also, some parameters need to be selected prior conducting real HFT, and when theoretical formulations are not available, VHFT helps determine the proper values of these parameters. No attempt was made to capture the connection behavior or localized instability of the beams and columns in this study.

Framework of the VHFT

In the VHFT, the PS and NS were separately represented in SAFIR. During the analysis, the two substructures were communicating every time step, thus simulating a real hybrid fire test. The exchange of information (i.e., interface displacements and reaction forces) between the PS and NS is done every time step and in order to increase the efficiency, some modifications were done in the nonlinear finite element SAFIR code to allow this exchange. The framework for communication for this VHFT was developed in Matlab (2016) and National Instrument's LabView (Elliott et al. (2007)) was used to establish data exchange between SAFIR and Matlab (Figure 8).



Results of the VHFT

Fire scenario S1 was analyzed by means of VHFT. The analyzed structure was divided to the PS and NS presented in Figure 9. The beam of the initial fire compartment represents the PS while the surrounding structure represents the NS. Since a 2D analysis was performed, 6 DoF were controlled at the interface of the PS and NS. The displacements and forces of the PS and NS for these DoFs need to be compatible and in equilibrium for the entire duration of the VHFT. The VHFT must reproduce; the global behavior of the analyzed structure, in addition to the interface compatibility and equilibrium.



Figure 9. The PS and NS in the case of the VHFT

presents the evolution of forces and displacements for the interface DoFs at nodes a and b. Their notation is shown on Fig. 9 as follows: 1 - axial DoF of node a, 2 - vertical DoF of node a, 3 - vertical DoF of node a, a - vertical DoF of node a.

rotational DoF of node a, 4 – axial DoF of node b, 5 – vertical DoF of node b, and 6 – rotational DoF of node b. Each sub-plot illustrates the interface solutions of the PS and NS resulting from the VHFT versus the global solution from the global analysis of the structure. The compatibility and equilibrium are ensured if the solution of PS equals the solution of the NS in the case of the VHFT. Their results should in turn match the global behavior of the analyzed structure.

To perform the VHFT, parameters such as the stiffnesses of the PS and NS, as well as the time step must be defined prior the analysis. The exact values of the stiffnesses of the PS and NS are continually changing during the VHFT, and hard to specifically predict at each time step. Their values at ambient temperature were used and kept constant in this initial phase of the study. Since the stiffnesses were kept constant, iterations were performed to converge to the global solution. Moreover, since the fire effect is continuous (continually changing the boundary (interface) conditions), this iteration process needs to be done continuously. In this initial analysis, it was assumed that 10 *s* are required to perform each time step's calculations and to apply the updated boundary conditions to the PS and the NS. Comparing the interface forces and displacements between the PS and the NS, initial results show that although equilibrium is not fully satisfied, axial and rotational (DoF 1, 3, 4, and 6) results show a good match. However vertical (Dof 2 and 5) had significant differences (up to 30 %). Also, the axial and rotational (DoF 1, 3, 4, and 6) results followed the global behavior very closely.

It is noted that these are preliminary results presented as a demonstration of the framework and as a proof of concept. An optimization process is needed (currently underway) to optimize the parameters used in the solution (such as the time step and the stiffnesses). Initial results from the optimization analysis already show that using a smaller time step dramatically improves the solution's convergence. However, note that the smaller the time step is, the faster of a reaction time of the HFT actuators needs to be. It also requires using higher computation power. Thus, allowing the new boundary conditions to be applied faster on the PS, and the computation of each solution step to be done faster. Since, the actuators and computers limit how small of a time step can be used, it is beneficial to also optimize the updating process of the PS and NS stiffnesses values used in the calculations for each time step. Finally, after optimizing the parameters, an uncertainty analysis is required to quantify the precision and accuracy of the solution.

SP-496



Figure 10. Interface forces versus displacements for the six interface DoF

16

Sauca, Ana; Seif, Mina; Zhang, Chao. "Stability of Steel Structures at Elevated Temperatures: A Hybrid Fire Testing Approach." Paper presented at Annual Stability Conference / NASCC: The Steel Conference, Baltimore, MD, United States. April 10, 2018 - April 13, 2018.

5. Summary

This paper presented the results of a heat analysis performed on a 10 story steel building when exposed to spreading designed fire scenarios. First, the global behavior of the structure was presented for two fire scenarios. In one fire scenario (S1), the fire initiated in the first-floor corner compartment and spread to the adjacent compartments. In the second fire scenario (S2), the fire initiated in the first-floor central compartment and then spreads in the adjacent compartments. The results showed the impact of the unheated compartments on the global behavior of the structure, thus when fire initiates in the central compartment (S2), the thermal expansion is restrained by the unheated surrounding, leading to the increase of mid-span displacements, axial forces and mid-span bending moments. When the fire initiates in the corner compartment (S1), the thermal expansion is restrained in one direction but less restrained in the other direction, therefore the mid-span displacements, axial forces, and bending moments are not as significant as in the S2 case. The temperature of the structural members does not reach the critical temperature of 550 °C, nevertheless, the failure occurs due to the fire spread in the compartments. This analysis highlighted the importance of considering the effect of surrounding members (the structure as a whole system) when performing tests on individual structural elements. For the analyzed case, when the effect of surrounding members is neglected, and individual "standard" member tests are to be performed, no failure is observed, whereas failure of those members was to occur in the global behavior. Afterwards, a virtual hybrid fire test was performed on the 10-story building for the case S1. The PS and NS were modeled in SAFIR and the framework for communication was developed in Matlab. The VHFT showed that the global solution can be reproduced if proper parameters are selected, such as the stiffnesses of the PS, NS and the time step of the analysis. The VHFT is useful in preparing the real HFT to properly select parameters (e.g., the time step of the analysis) when analytical formulations are not available so the interface equilibrium and compatibility are satisfied and the global behavior of the structure is satisfied.

Disclaimer

Certain commercial entities, equipment, products, software, or materials are identified in this document in order to describe a procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, products, software, materials, or equipment are necessarily the best available for the purpose.

References

ABAQUS (2011) `ABAQUS Documentation', Dassault Systèmes, Providence, RI, USA.

Armer, G. S. T., & Moore, D. B. (1994). "Full-Scale Testing on Complete Multistorey Structures." *The Structural Engineer*, Volume 72, No 2/18 January 1994, pp 30-31.

ASTM E119-16a (2016). Standard test methods for fire tests of building construction and materials, 2016.

Bisby, L., Gales, J., and Maluk, C. (2013). "A contemporary review of large-scale non-standard structural fire testing", *Fire Science Review* 2(1): 1-27.

Buchanan, A. H. (2001). "Structural Design for Fire Safety." John Wiley & Sons, Ltd.

CEN. EN 1991-1-2:2005 – Eurocode 3. Design of steel structures. General rules. Structural fire design; 2005. doi: ISBN 978 0 580 66390 1.

Elliott, C., Vijayakumar, V., Zink, W., Hansen, R. (2007), "National Instruments LabVIEW: A Programming Environment for Laboratory Automation and Measurement" Cytokinetics, Inc., San Francisco, CA.

Franssen, J.-M. (2005), "SAFIR, A Thermal/Structural Program Modelling Structures under Fire", A.I.S.C. Engineering Journal, 42 (3) 143–158.

- Franssen, J.-M. and Gernay, T. (2017), "Modeling structures in fire with SAFIR®:, Theoretical background and capabilities", Journal of Structural Fire Engineering, 8 (3), 300-323.
- Kiel, M. (1989). "Entwicklung einer intelligenten Prüfmaschine für brandbeanspruchte Gesamttragwerke." Proceedings of the Braunschweiger Brandschutztage 1989, Braunschweig, Germany.
- Korzen, M., Magonette, G., Buchet, Ph. (1999). "Mechanical Loading of Columns in Fire Tests by Means of the Substructuring Method." Zeitschrift für Angewandte Mathematik und Mechanik, Vol. 79, pp. S617-S618.
- Lennon, T. (2003). "Whole building behavior Results from a series of large scale tests." CIB-CTBUH International Conference on Tall Buildings, 8-10 May 2003, Malaysia.
- Mostafei, H. (2013a). "Hybrid Fire Testing for Assessing Performance of Structures in Fire Application." Fire Safety Journal, Vol. 56, pp. 30-38.
- Mostafaei, H. (2013b). "Hybrid fire testing for assessing performance of structures in fire -Methodology." Fire Safety Journal, 2013, 58, 170-179.

MATLAB 7.10.0, Natick, Massachusetts, The MathWorks Inc., 2016.

- Newman, G. M., Robinson, J. T., Bailey, C. G. (2000). "Fire Safe Design, A New Approach to Multi-storey Steel Framed Buildings." SCI P-288, The Steel Construction Institute, Ascot (United Kingdom).
- OpenSees 2016. "Open System for Earthquake Engineering Simulation." UC Berkley
- OpenFresco 2016. "Open Framework for Experimental Setup and Control." UC Berkley
- Peacock, R.D., Reneke, P.A., Forney, G.P. "CAFST-consolidated model of fire growth and smoke transport (version 7), volume 2: user's guide." NIST Technical Note 1889v2, 2017.
- Rackauskaite, E., Kotsovinos, P., Rein, G. (2017a). "Structural Response of a steel-frame building to horizontal and vertical travelling fires in multiple floors." Fire Safety Journal, 91, 542-552.
- Rackauskaite, E., Kotsovinos, P., Jeffers, A., Rein, G. (2017b). "Structural analysis of multi-storey steel frame exposed to travelling fires and traditional design fires." Engineering Structures, 150, 271-287.
- Robert, F., Rimlinger, S., Collignon, C., (2009). "Promethee, Fire Resistance Facility Taking Into Account the Surrounding Structure." 1st international Workshop on Concrete Spalling due to Fire Exposure, 2009, 3-5 sept.
- Robert, F., Rimlinger, S., Collignon. C., (2010). "Structure fire resistance: a joint approach between modelling and full scale testing (substructuring system)." Proceedings of 3rd International fib Congress incorporating the PCI Annual Convention and Bridge Conference 2010, Washington, DC, USA
- Sadek, F., Main, J. A., Lew, H. S., Robert, S. D., Chiarito, V. P., and El-Tawil, S. (2010). "NIST Technical Note 1669. An Experimental and Computational Study of Steel Moment Connection under a Column Removal Scenario.'
- Saouma, V.E. & Sivaselvan, M.V. (2008). "Hybrid simulation: Theory, implementation and application." Taylor and Francis, London, UK. 2008.
- Sauca, A., Gernay, T., Robert, F. & Franssen, J.M. (2015a). "Characteristics and implementation of Hybrid Fire Testing (HFT)." presented at the Engineering Mechanics Institute Conference, 2015, June 16-19.
- Sauca, A., Gernay, T., Robert, F. and Franssen, J.M. (2015b). "Analysis of a concrete building exposed to natural fire." International Conference ASFE, 2015, October 15-16.
- Sauca, A., Gernay, T., Robert, F., Tondini, N. and Franssen, J.M. (2016a). "Stability in Hybrid Fire Testing." 9th International Conference on Structures in Fire, 2016, June 8-10.
- Sauca, A., Gernay, T., Robert, F., Tondini, N. and Franssen, J.M. (2016b). "A novel method for Hybrid Fire Testing." 6th European Conference on Structural Control, 2016, July 11-13.
- Sauca, A., (2017). "Development and implementation of a methodology for hybrid fire testing applied to concrete structures with elastic boundary conditions." Doctoral Thesis, University of Liege, Liege, Belgium, 2017
- Sauca, A., Mergny, E., Gernay, T., and Franssen, J.M. (2017a). "A method for Hybrid Fire Testing: Development, implementation and numerical application." Proceedings of Applications of Structural Fire Engineering (ASFI'17), 2017, September 7-8.
- Sauca, A., Gernay, T., Robert, F., Tondini, N., & Franssen, J.-M. (2017b). "Hybrid Fire Testing: Discussion on Stability and Implementation of a New Method in a Virtual Environment." Journal of Structural Fire Engineering (in press).
- Schellenberg, A.H., Mahin, S.A., Fenves G.L., (2009). "Advanced implementation of Hybrid Simulation." PEER Report 2009/104, University of California, Berkeley.
- Schulthess, P., Neuenschwander, M., Knoblock, M. and Fontana, M. (2016). "Consolidated Fire Analysis Coupled Thermo-Mechanical Modelling for Global Structural Fire Analysis", 9th International conference on Structures in Fire, 8-10 June, pg. 819-826.
- Takanashi, K., et al. (1975). "Non-linear earthquake response analysis of structures by a computer-actuator online system - Part 1: detail of the system." Transactions of the Architectural Institute of Japan (229): 77-83.

- Tondini, N., Hoang, V. L., Demonceau, J.-F., & Franssen, J.-M. (2013). "Experimental and numerical investigation of high-strength steel circular columns subjected to fire." *Journal of Constructional Steel Research*, 80, 57-81.
- Tondini, N., Abbiati, G., Posidente, L. and Stojadinovici, B. (2016). "A Static Partitioned Solver for Hybrid Fire Testing", *9th International conference on Structures in Fire*, 8-10 June, pg. 827-835.

- Whyte, C.A., Mackie, K.R. and Stojadinovici, B. (2016). "Hybrid Simulation of Thermomechanical Structural Response", *Journal of Structural Engineering*, 142(2): 04015107-1 04015107-11.
- Zhang, C., Gross, J.L., McAllister, T.P., Li G.Q. (2015). "Behavior of unrestrained and restrained bare steel columns subjected to localized fire." *Journal of Structural Engineering* ASCE., Vol. 141.
- Zhang, C., Gross, J.L., McAllister, T.P. (2013a). "Lateral torsional buckling of steel W-beams subjected to localized fires." *Journal of Constructional Steel Research*, Vol. 88, pp.330-8.
- Zhang, C., Li, G.Q., Usmani, A. (2013b). "Simulating the behavior of restrained steel beams to flame impingement from localized-fires." *Journal of Constructional Steel Research*, Vol. 83, pp.156-65

UI-SimCor, University of Illinoi

Calibration of solid-phase microextraction for measuring indoor gas phase **SVOC concentration**

Mengyan Gong^{1,*}, Dustin Poppendieck¹

¹Indoor Air Quality and Ventilation Group, Engineering Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA *Corresponding email: mengyan.gong@nist.gov

SUMMARY

The most common method to calibrate solid-phase microextraction (SPME) for measuring gas phase semi-volatile organic compound (SVOC) concentrations, i.e., splitless liquid injection, may be associated with large measurement uncertainty. This study aims to evaluate the direct syringe-fiber loading method for calibrating SPME by examining repeatability, influence of solvent evaporation, and linearity of standard curves for several typical indoor SVOCs. The direct loading method has been successfully validated for DEHP. Evaluation for TDCPP, B[a]P and PCB-52, and application to DEHP concentration in air of a sealed chamber with PVC flooring will be conducted.

KEYWORDS

SPME, SVOC emissions, DEHP, PVC flooring, Indoor air quality

1 INTRODUCTION

Measurement of gas phase SVOC concentrations is required for indoor exposure assessment. Compared with other SVOC sampling techniques, e.g., sorbent tube and polyurethane foam, SPME has advantages of small sampling volume, ease of sample preparation and sampling, and simple injection (Ouyang and Pawliszyn, 2008). In addition, SPME is the only suitable method for sampling in sealed environments, e.g., the sealed stainless steel chamber used in Cao et al. (2016). Therefore, SPME has been increasingly used to sample gas phase SVOCs in both sealed and ventilated environments. However, to use SPME to measure SVOC concentration in air, quantification of the absolute amount of analyte on the SPME fiber is required. To calibrate the amount of analyte on the fiber, a separate liquid splitless injection to a gas chromatography-mass spectrometer (GC-MS) is often used, which assumes that SPME injection has the same transfer efficiency as liquid splitless injection. However, the transfer efficiencies can be significantly different for the two injection techniques (Ouyang et al., 2005). Thus, calibrating with liquid splitless injection may cause large uncertainties in measured SVOC concentrations. The purpose of this study is to improve and evaluate the current direct syringe-fiber loading method by determining the reproducibility, influence of solvent evaporation, and linearity for several typical indoor SVOCs, and demonstrating the method's applicability by measuring bis(2-ethylhexyl) phthalate (DEHP) concentration in air of a sealed chamber with PVC flooring.

2 METHODS

For calibration, as shown in Figure 1, 1 µL of standard liquid at different concentrations and 1 µL of a corresponding internal standard were spiked onto the SPME fiber. After the solvent evaporated, the SPME fiber was injected into the GC-MS. We used several typical SVOCs, e.g. DEHP, Tris(1,3-dichloro-2-propyl)phosphate (TDCPP),



Figure 1. Illustration of direct loading of calibration liquid onto SPME fiber.

Benzo[a]pyrene (B[a]P), and 2,2',5,5'-Tetrachlorobiphenyl (PCB-52) (and their corresponding internal standards DEHP-d₄, TDCPP-d₁₅, BaP-d₁₂ and ¹³C PCB-52) to evaluate this method.

First, the repeatability of standard loading was examined by running the same standard multiple times. Second, to examine the standard loss during solvent evaporation, 1 μ L of standard liquid

was spiked onto the fiber and evaporated for different times, while 1 μ L internal standard was spiked onto the fiber but evaporated for a fixed time for each sample. Third, the linearity of the standard curve was evaluated. After evaluation of the calibration method, SPME will be used to sample in a specially designed sealed chamber containing PVC flooring for different times at 25 °C, as show in Figure 2. The DEHP concentration in the air of the sealed chamber was then calculated based on the method described in Cao et al. (2016).



Figure 2. SPME sampling from a sealed chamber.

3 RESULTS AND DISCUSSION

The results show that the response ratio of DEHP and DEHP-d₄ for multiple injections are consistent with relative standard deviations less than 10 %. Figure 3 shows that the evaporation time of DEHP doesn't significantly influence the response. Hence, if the evaporation time of DEHP is within the examined time range (less than 32 min), the evaporation loss will be insignificant. The linearity of the standard curve for direct loading is high with R² larger than 0.99, as shown in Figure 4. Based on these results, it appears that the direct loading method can be used to calibrate the amount of DEHP sorbed to SPME. However, since different chemicals have varying evaporation rates, the method may not work for chemicals with higher volatility, and application for other chemicals needs to be further examined.



Figure 3. Influence of DEHP evaporation loss. Figure 4. SPME calibration curve for DEHP.

4 CONCLUSIONS

The method of calibrating SPME by direct syringe-fiber loading has been validated for DEHP. Evaluation for other selected SVOCs and application of the method for measurement of DEHP concentration in air of sealed chamber with PVC flooring will be conducted. If successful, the direct loading method could be used as a calibration method for using SPME to measure gas phase SVOC concentrations. However, the applicability of this method for chemicals not evaluated in this study needs to be examined using the procedure in this study.

5. REFERENCES

Cao J, Zhang X, Little JC, Zhang Y. 2016. *Indoor Air*. 27(2), 417-426. Ouyang G, Chen Y, Setkova L, Pawliszyn J. 2005. *Journal of Chromatography A*, 1097(2): 9-16. Ouyang G, Pawliszyn J. 2008. *Analytica Chimica Acta*, 627(2): 184-197.

Towards Augmented Reality Interfaces for Human-Robot Interaction in Manufacturing Environments

Shelly Bagchi National Institute of Standards and Technology Gaithersburg, MD shelly.bagchi@nist.gov

ABSTRACT

As part of the U.S. Department of Commerce, the National Institute of Standards and Technology (NIST) is examining new approaches to increase the competitiveness of smaller U.S.-based manufacturers. Collaborative robotics is one way to increase automation, leveraging the specialized skills of human workers along with the increased repeatability and throughput of robots. However, in order for humans to efficiently work alongside robots, new types of interfaces are needed to streamline interactions. Augmented reality, especially wearable devices, are a new approach that can enable task training, hands-free interactions, and increased safety through situational awareness. This paper will explore some preliminary approaches to integrating augmented reality for a collaborative manufacturing environment.

CCS CONCEPTS

• Human-centered computing \rightarrow HCI design and evaluation methods; Interactive systems and tools; Ubiquitous and mobile devices; \cdot Computer systems organization \rightarrow Embedded systems; Robotics;

KEYWORDS

augmented reality, human-robot interaction, manufacturing, interfaces, metrics, industrial robots

ACM Reference Format:

Shelly Bagchi and Jeremy A. Marvel. 2018. Towards Augmented Reality Interfaces for Human-Robot Interaction in Manufacturing Environments. In Proceedings of ACM Human-Robot Interaction Conference, Virtual, Augmented, and Mixed Reality Workshop (VAM-HRI). ACM, New York, NY, USA, Article 8,

1 INTRODUCTION

Automation, particularly the use of robotics, is becoming more prevalent among the larger U.S. manufacturers [1, 10], and this trend has been predicted to continue growing exponentially in the near future [12]. Recent trends show robots are increasingly being used in small- and medium-sized enterprises (SMEs) as well, principally due to decreasing costs and advances in safety [8]. In particular, collaborative robots offer SMEs the advantage of decreased training overhead and increased safety without requiring changes in the

VAM-HRI, March 2018, Chicago, IL USA

2018. ACM ISBN 978-x-xxxx-xx/YY/MM...\$15.00

Jeremy A. Marvel National Institute of Standards and Technology Gaithersburg, MD jeremy.marvel@nist.gov

work environment to accommodate cages, thus leading to greater flexibility and multi-purpose use [7].

Current off-the-shelf collaborative robots, however, are generally only collaborative in the sense that they are expected to be safety-rated to operate near humans, incorporating force-limited motions or collision avoidance. To enable truly collaborative tasks that can leverage the abilities of existing human workers, not only do robots need to be aware of their environment and human partner(s), but human-facing interfaces need to evolve as well to enable real-time, two-way communication. Traditional teach pendant interfaces do not typically allow for easy or convenient use while also performing a task. Likewise, teach pendants make it difficult to obtain feedback or task status from the robot in an easy and efficient manner. New types of interfaces need to be developed to simplify interactions and incorporate system diagnostics, prognostics, and situation awareness (SA) before truly collaborative robotics can become a reality

Augmented Reality (AR) technology has advanced rapidly in recent years. Now that such devices are widely available commercially, their acceptance is becoming increasingly common as their reliability improves. This increased acceptance and reliability enables integrators to leverage their use as human-robot interaction (HRI) interfaces for industrial applications. In particular, wearable AR devices have been shown to increase situational awareness (SA) and lead to better task performance [3], as well as grounding in the robot state [4].

Most AR devices can be split into two categories: "see-through" or monitor-based displays [9]. As the name implies, see-through displays project information onto a transparent medium that permits users to retrieve information without having to divert attention from the principal area of focus. Monitor-based displays, in contrast, require the user to focus attention on a separate display or interface through which the world is observed. This paper will explore multiple types of AR interfaces and their differing uses as industrial robot interfaces. As prototype interfaces are developed, future work will allow the collection of user data and enable further conclusions about the utility of AR devices for HRI.

2 WEARABLE DEVICES

In the category of see-through, wearable AR technology, there are generally two types of devices available: headsets and smart glasses. Smart glasses generally have one or two small screens projected into the user's direct or peripheral view, and look fairly similar to regular eyeglasses. Notably, the smaller form factor of these devices makes it difficult to incorporate advanced sensing beyond a single camera and microphone. Natural language control is possible, but may not be as useful in a loud manufacturing environment. This

Bagchi, Shelly; Marvel, Jeremy. "Towards Augmented Reality Interfaces for Human-Robot Interaction in Manufacturing Environments." Paper presented at Workshop on Virtual, Augmented and Mixed Reality for Human-Robot Interaction, Chicago, IL, United States. March 5,

2018 - March 5, 2018.

This paper is authored by an employee(s) of the United States Government and is in the public domain. Non-exclusive copying or redistribution is allowed, provided that the article citation is given and the authors and agency are clearly identified as its source

limits the user's interaction with the real world through virtual objects ("mixed reality") as well as the user's communication with the robot, and thus limits the application of these devices in HRI.

For the purpose of industrial collaborative robotics, AR headsets are more frequently being considered over smart glasses. In addition, there are manufacturing-specific devices being developed that are directly integrated into personal protective equipment (PPE) so as to ensure safety even while using immersive interfaces. Untethered headsets with wireless communication capabilities are ideal for performing manufacturing tasks while communicating with a robot partner. Additionally, these devices contain multiple cameras and inertial measurement units (IMUs) to perform head tracking and hand gesture recognition. The ability to project larger, threedimensional holograms, as well as perform object recognition and other perception techniques, makes these AR devices promising for HRI applications.

2.1 Task Augmentation

Smaller manufacturers are often retasking their industrial robots for new products or applications [5, 11]. When retasking, human workers must be trained and industrial robots must be programmed before new production can begin. AR headsets can be useful on both sides of this equation by performing task-specific environment augmentations.

While wearing an AR headset, task training for the operator could be performed in an abbreviated training session or potentially even on-the-job. For example, relevant parts for an assembly can be identified by the headset and highlighted in the kit tray using a holographic overlay. Instructions can appear to the user textually, graphically, or verbally, depending on the situation or the user's preferences. The robot's part of the collaborative task can be identified by the headset as well, allowing the human to train it using the AR control interface, or other options such as kinesthetic learning or teaching by demonstration. In this way, both task partners can benefit from the AR interface.

Task augmentation can also be performed as a method of error detection. In the case that the wrong part is chosen for an assembly, for instance, the headset's perception system could detect a mismatch and alert the operator. This should save time down the line in the quality assurance process.

2.2 Robot Feedback

Similar to error detection, robot fault detection can be integrated into an AR interface. The robot constantly reports its status to the headset, which can monitor several relevant parameters such as joint torque limits or Cartesian speed. Should a given parameter begin to approach its limit, the headset can alert the operator to the issue and display potential causes or methods of handling the issue

The robot can also use communication with the headset to query its human teammate. There may be issues where the robot's perception system is occluded or unsure which object is the goal, or an object may be out of the robot's reach. The robot can transmit these types of questions to the user - in text, visually, or verbally through natural language processing - to get assistance and make use of human knowledge or flexibility.

2.3 Situation Awareness

In addition to robot fault reporting as discussed above, there are multiple ways AR headsets and interfaces can improve operator situation awareness and grounding. One such option is for the robot to inform the operator of its next action, either through a virtual model (hologram) or by highlighting objects in the environment. This would allow the human teammate to adjust their actions accordingly, helping accomplish preconditions for the robot's next task, and potentially improving safety as the human knows the robot's intent. The operator can also intervene if the robot intends to perform the wrong task or has the potential of colliding with other objects. In this way, the collaborative task can take advantage of the robot's efficiency while still incorporating human oversight. Additionally, the human knows more about the robot's current state, and potentially feels more comfortable working directly with the robot.

3 **PORTABLE DEVICES**

On the opposite side of the spectrum from see-through or wearable devices are simpler devices with display screens and rear cameras (e.g., tablets and smart phones). These consumer devices are widely available for a relatively low price point, and can execute AR software comparable to headsets without the low screen real estate of smart glasses. External camera feeds can also be used with any computer, but are not significantly different from existing interfaces, so they will not be discussed here. Tablets are particularly interesting due to the potential of touchscreen interfaces. The recent prevalence of touchscreen devices means that interfaces on similar devices will require less training for novice users. The portable nature of these devices also gives them further utility for use in technical support and diagnostics applications.

3.1 Diagnostics & Prognostics

An example scenario we have envisioned for walk-up diagnostics is as follows: A technician is making the rounds of the industrial robots on a factory floor. At each robot, they use their tablet camera to scan an AR target (an image or other symbol, defined ahead of time). The scan connects them to the particular robot they are examining, via the factory's local wireless network, and launches the AR application. The technician can now point the camera at the robot, which is overlaid with a virtual model displaying the current robot status - for example, the torque at every joint. Problem areas reported by the robot will be highlighted in the AR view. Additionally, the technician can easily control the robot via the touchscreen interface to test operations or access areas in need of service.

Figure 1 shows a prototype tablet interface that incorporates several of the features mentioned above. A virtual model of the robot appears after scanning the image target shown in the photo (here a printed photo of the robot in question). Joint-specific controls show on the right side of the screen, along with options for communicating with the non-virtual robot. The robot's current configuration is updated live through both the virtual avatar and the joint sliders. Through this type of interface, a technician can observe and monitor the robot, as well as control it and receive relevant status information of the robot's current task.

Bagchi, Shelly; Marvel, Jeremy. "Towards Augmented Reality Interfaces for Human-Robot Interaction in Manufacturing Environments." Paper presented at Workshop on Virtual, Augmented and Mixed Reality for Human-Robot Interaction, Chicago, IL, United States. March 5,

2018 - March 5, 2018.

Towards Augmented Reality Interfaces for Human-Robot Interaction in Manufacturing Environmented Reality Interfaces for Human-Robot Interaction Interac



Figure 1: Prototype tablet interface using Augmented Reality technology.

4 EVALUATION & FUTURE WORK

Before new technology will be accepted by manufacturers, it must be thoroughly tested for performance, reliability, and safety. SMEs, in particular, have lower risk-tolerance, which may prevent them from investing in technologies that have not been exhaustively evaluated, or vetted through broad industry use [13]. Therefore, we intend to conduct a comprehensive study into the benefits of AR interfaces as compared to current methods.

Currently, several versions of AR interfaces for HRI are in development. These include headset, tablet, and touchscreen desktop applications. Once development is completed, user studies are planned to compare these interfaces to more traditional interfaces such as a teach pendant. This will allow us to collect user data in the form of both objective and subjective measures. For the former, these might include performance metrics such as task completion time or measures of success. For the latter, participants will be asked to evaluate each interface according to their preferences, answering post-task survey questions on a Likert scale. In addition, measures of mental workload will be collected from the Task-Load Index (TLX) developed by the National Aeronautics and Space Administration (NASA) [6], and task situation awareness will be measured from the situation awareness rating technique (SAGAT) or the situation awareness global assessment technique (SAGAT) tests [2].

Collecting a wide variety of metrics and surveying a large pool of participants will allow us to draw conclusions about the strengths and weaknesses of each type of interface. Ultimately, our goal is also to evaluate which metrics are most important for HRI interfaces. In this way, we can help guide the creation of commercial HRI interfaces that have proven performance, are thoroughly safety tested, and held to accepted standards – thus more likely to be adopted in industrial settings.

5 CONCLUSIONS

As manufacturers move towards greater automation, new types of technology are needed to create a truly collaborative working environment. Augmented Reality can be a very useful tool for making human workers more effective, particularly for humanrobot collaboration. As robot interfaces, AR devices can provide not only effective robot control, but also task training, feedback, and better SA. The process of robot diagnostics, root cause analysis, health monitoring, and prognostics can also be improved through the use of AR. In the future, we expect to quantitatively evaluate the potential performance benefits of AR used for HRI.

DISCLAIMER

Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

REFERENCES

- B. Doyle. 2015. North American Robotics Market Sets New Records in 2015. (2015). "http://www.robotics.org/content-detail.cfm/Industrial-Robotics-News/ North-American-Robotics-Market-Sets-New-Records-in-2015/content_id/ 5951"
- [2] Mica R Endsley and Debra G. Jones. 2016. Designing for situation awareness: An approach to user-centered design. CRC press.
 [3] Scott A. Green, Mark Billinghurst, XiaoQi Chen, and J. Geoffrey Chase. 2008.
- [3] Scott A. Green, Mark Billinghurst, XiaoQi Chen, and J. Geoffrey Chase. 2008. Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design. International Journal of Advanced Robotic Systems 5, 1 (2008), 1. https://doi.org/10.5772/5664 arXiv:https://doi.org/10.5772/5664
- [4] S. A. Green, J. G. Chase, X. Chen, and M. Billinghurst. 2008. Evaluating the Augmented Reality Human-Robot Collaboration System. In 2008 15th International Conference on Mechatronics and Machine Vision in Practice. 521–526. https://doi.org/10.1109/MMVIP.2008.4749586
- [5] V. C. Gungor and G. P. Hancke. 2009. Industrial Wireless Sensor Networks: Challenges, Design Principles, and Technical Approaches. *IEEE Transactions on Industrial Electronics* 56, 10 (Oct 2009), 4258–4265. https://doi.org/10.1109/TIE. 2009.2015754
- [6] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Adv. Psychol. 52 (1988), 139–183.
- [7] J. Marvel. 2014. Collaborative Robotics: A Gateway into Factory Automation. (2014). http://news.thomasnet.com/imt/2014/09/03/ collaborative-robots-a-gateway-into-factory-automation
 [8] Jeremy Marvel, Elena Messina, Brian Antonishek, Lisa Fronczek, and Karl Van
- [8] Jeremy Marvel, Elena Messina, Brian Antonishek, Lisa Fronczek, and Karl Van Wyk. 2015. NISTIR 8093: Tools for collaborative robots within SME workcells. Technical Report. National Institute of Standards and Technology.
- [9] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: a class of displays on the reality-virtuality continuum. (1995), 2351 - 2351 - 11 pages. https://doi.org/10.1117/12.197321
- [10] M. Orcutt. 2014. Robots Rising. (2014). "https://www.technologyreview.com/s/ 529971/robots-rising/"
- [11] Craig Schlenoff, Tsai Hong, Connie Liu, Roger Eastman, and Sebti Foufou. 2013. A literature review of sensor ontologies for manufacturing applications. In *Robotic* and Sensors Environments (ROSE), 2013 IEEE International Symposium on. IEEE, 96-101.
- [12] L. Wood. 2014. Research and Markets: Global Articulated Robots Market Growth of 16.272016-2019. (2014). "http://www.businesswire.com/news/home/20160108005449/en/ Research-Markets-Global-Articulated-Robots-Market-Growth"
- [13] M. Zimmerman and J. Marvel. 2017. Smart Manufacturing and The Promotion of Artificially-Intelligent Human-Robot Collaborations in Small- and Medium-sized Enterprises. In Association for the Advancement of Artificial Intelligence 2017 Fall Symposium Series.

Bagchi, Shelly; Marvel, Jeremy.

"Towards Augmented Reality Interfaces for Human-Robot Interaction in Manufacturing Environments." Paper presented at Workshop on Virtual, Augmented and Mixed Reality for Human-Robot Interaction, Chicago, IL, United States. March 5,

2018 - March 5, 2018.

Application of Digital Image Correlation to Structures in Fire

Christopher M. Smith¹, Matthew S. Hoehler²

¹ Berkshire Hathaway Specialty Insurance, 100 Federal Street, 20th Floor, Boston, MA 02110, USA

² National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA

KEYWORDS: digital image correlation; DIC; fire; narrow-spectrum illumination; blue light

1 Introduction

The behavior of engineering structures in fire is commonly studied through large-scale experiments. However, temperaturevarying material properties, spatially-varying thermal loading and complicated structure geometries result in these structures deforming in complex and often unpredictable ways. This makes it difficult to fully characterize the response of the structure with traditional (point) sensors [1]. Full-field, noncontact measurement techniques such as Digital Image Correlation (DIC) are potentially ideal for such experiments; however, the presence of light emitted by the flames, thermal radiation from the heated structure, and convective thermal gradients in the air make this a challenging application for DIC.

2 Imaging Through Fire Using Narrow-Spectrum Illumination

A simple and robust method has been developed to enable the use of DIC in large, low-soot, fires to measure targets at temperatures up to 800 °C. This method, which builds on work by Grant [2] and Pan [3], uses narrow-spectrum blue light (450 nm ± 10 nm at one standard deviation) and spectrally-matched band-pass optical filters to increase signal-to-noise ratio and filter undesired radiant energy before it reaches the camera [4]. Fig. 1a shows a benchtop-scale demonstration of this method. The distance from the camera to the target is about 1 m. Fig. 1b contains three images of the same natural gas flame with different illumination and filtering methods applied. In the image on the left, no filters are used, and the target is illuminated only by the flame and ambient lighting in the room. In the middle image, a 10 W 450 nm light emitting diode (LED) source is added to illuminate the target; improving the visibility of the target. In the image on the right, a band-pass optical filter is placed in front of the camera, blocking the light from the flame and allowing the target behind the flame to be seen clearly. This technique can reduce the observed intensity of the flame by a factor of 10⁴, which is sufficient to image full-scale structural experiments using less than 50 W of applied illumination [4].



Fig. 1 Demonstration of imaging method: (a) schematic of setup and (b) target viewed through 5 kW natural gas flame when illuminated with i) ambient light, ii) 10 W of 450 nm light, and iii) 10 W of 450 nm light and imaged through a band-pass filter.

3 Applications to Optical Metrology

The proposed filtering method allows DIC measurements to be made through flames that have low soot content. Converting the images to quantitative measurements poses additional challenges because the gas temperature gradients optically distort the images causing fictitious (apparent) strains. While there is a steady-state aspect of this distortion caused by the overall shape of the fire, in a large diffusion flame, the turbulent flow (flicker) of the flame is responsible for most of the distortion. The characteristic frequency of the flame flicker is on the order of ~ 10 Hz at the small scale [5, 6], and lower frequency transient distortions can occur due to variations fire ventilation at the large (1 m) scale. Meanwhile the characteristic dimension of individual flame eddies is on the order of 5 cm. As the strain rates in structural testing are slow relative to the flame motion, temporal and spatial averaging can be used to improve the accuracy of measurements.

4 Application to Full-Scale Experiments

The method described above was applied to several repeated full-scale experiments in which a 6 m long W16×26 steel beam was supported over a 700 kW fire from a natural gas diffusion burner (Fig. 2a). Measured gas temperatures in the flame exceeded 1400 °C and the steel beam reached temperatures over 700 °C [7]. A commercial DIC system, with optical filters installed, was placed at 3 m stand-off and the specimen was illuminated using two 100 W blue LED arrays. The target area was approximately 0.4 m high by 1 m long. Prior to painting a pattern to enable DIC, the surface oxide layer ('mill scale') on the beam was removed using sandblasting, as it naturally flakes off when heated above 400 °C or when the steel yields. A base layer of white high-temperature spray paint was applied followed by a random pattern stenciled in black spray paint to provide contrast (Fig. 2a). The combination of a white base coat with a black pattern was found to be more temperature-resistant than the inverse. The optical distortion from the flame, and necessary time averaging, have the effect of blurring the image to be processed with DIC, so the pattern was designed to be coarser (10-20 pixels/feature) than necessary in an unobstructed test (3-5 pixels/feature).



Fig. 2 Full-scale validation tests: (a) test setup and (b) example of measured longitudinal strain field (units in mm/mm).

During the experiment, the beam was heated by the flame, and the thermal strains were measured using DIC. Images were recorded at 1 Hz and time-averaged over 30 seconds to remove transient flame effects. The time-averaging was performed on the raw images, which were subsequently reimported into the DIC system for processing. After performing the DIC, the resulting strains were spatially averaged over 4 cm regions (100 pixels). As illustrated in Fig 2b, this method was sufficient to differentiate thermal strains in the steel of 3 microstrain from the flame-induced distortion. The tests also suggest that time averaging is preferable to spatial averaging, at least for structural experiments for which spatial resolution is preferable to temporal resolution. Further work is needed to better understand the post-processing options and methods for this application and to quantify the uncertainty associated with these processes.

5 Conclusions

Noncontact measurement techniques, such as Digital Image Correlation (DIC), are desirable for use in structural fire experiments because deformation and failure in large structures subject to fire is often unpredictable. The presence of light emitted by flames, thermal radiation from the heated structure, and convective thermal gradients in the air have challenged the use of DIC. A simple and inexpensive method to improve the signal-to-noise ratio (10⁴ compared to ambient lighting) in images of objects taken through low soot-yield fires is described and has been successfully deployed on large scale fire tests. The resulting images were temporally and spatially averaged during post-processing to smooth out false distortions of the images caused by the thermal gradients in and around the flames before DIC techniques were applied to resolve strain. Additional work to quantify the uncertainty associated with these processes is needed.

References

- 1. McAllister T, Luecke W, Iadicola M, Bundy M (2012) Measurement of temperature, displacement, and strain in structural components subject to fire effects: concepts and candidate approaches. doi: 10.6028/NIST.TN.1768
- Grant BMB, Stone HJ, Withers PJ, Preuss M (2009) High-temperature strain field measurement using digital image correlation. J Strain Anal Eng Des 44:263–271. doi: 10.1243/03093247JSA478
- 3. Pan B, Wu D, Wang Z, Xia Y (2011) High-temperature digital image correlation method for full-field deformation measurement at 1200 °C. Meas Sci Technol 22:15701. doi: 10.1088/0957-0233/22/1/015701
- 4. Smith CM, Hoehler MS Imaging Through Fire Using Narrow-Spectrum Illumination. Fire Technol.
- 5. Ballantyne A, Bray KNC (1977) Investigations into the structure of jet diffusion flames using time-resolved optical measuring techniques. Symp Combust 16:777–787. doi: 10.1016/S0082-0784(77)80371-8
- 6. Buckmaster J, Peters N (1988) The infinite candle and its stability—A paradigm for flickering diffusion flames. Symp Combust 21:1829–1836. doi: 10.1016/S0082-0784(88)80417-X
- 7. Choe L, Ramesh S, Hoehler M, et al (2018) National fire research laboratory commissioning project: testing steel beams under localized fire exposure. doi: 10.6028/NIST.TN.1983

STUDY ON FIREBRAND GENERATION FROM ROOFING ASSEMBLY COMBUSTION

Samuel L. Manzello¹, Sayaka Suzuki², and Tomohiro Naruse³ ¹National Institute of Standards and Technology (NIST), USA ²National Research Institute of Fire and Disaster (NRIFD), Japan ³Building Research Institute (BRI), Japan

1. INTRODUCTION

Wildland fires that spread into communities, termed wildland-urban interface (WUI) fires, are frequently seen in media reports and have resulted in loss of life, injuries, and property damage. There have been a significant number of WUI fires all over the world. During October 2017, multiple WUI fires in California destroyed more than five thousand structures and resulted in more than 40 deaths. In 2017, more than 60 people perished in WUI fires in Portugal.

In Japan, many cities are densely populated and there exists the potential for large-scale urban fires, even without the presence of earthquakes. The most recent of these urban fires occurred in Itoigawa City, Japan on December 2016. In this fire, 147 structures were damaged by fire, with 120 of 147 destroyed. The Great Hanshin earthquake in Kobe, Japan, in 1995 had urban fires that occurred after a strong earthquake. More than 285 fires were reported that burned more than 7,000 structures.

A governing mechanism of structure ignition in both WUI and urban fires is the production of firebrands. Hardly any data exist for firebrand size distributions from actual structures or large outdoor fires, such as WUI and urban fires [1]. In WUI fires, the structures themselves may be a large source of firebrands, in addition to the vegetation. It is believed that production from structures is a significant source of firebrands in WUI fires [2]. In the case of urban fires in Japan, structures are responsible for firebrand production, but in the case of WUI fires, little is known about firebrand size/mass distributions produced.

An experimental protocol has been developed to ignite full-scale roofing assemblies and quantify the degree of firebrand production during the combustion process. As wind is an important factor in firebrand generation, the experiments were conducted in a full-scale wind tunnel. These results are being incorporated into a comprehensive database the authors are developing related to firebrand production from structures.

2. EXPERIMENTAL DESCRIPTION

Experiments were performed in the Building Research Institute's (BRI) Fire Research

Wind Tunnel Facility (FRWTF) by varying the wind speed. As part of the database development, wall assemblies and re-entrant corner assemblies were used in experiments; here roofing assemblies have been considered. The roofing assemblies used in the experiments were 1.2 m by 1.2 m in dimension. Wood joists (2 x 6) were used as framing, and oriented strand board (OSB thickness - 11 mm; plywood thickness 11.9 mm) was used as the base sheathing material (see Fig. 1). As a firststep, specific roofing treatments, such as roof tiles, were not applied. OSB and plywood were selected as both are common construction materials used in USA and Japan. The assemblies were installed inside the test section of the FRWTF at BRI. A custom frame was used to support the roofing assemblies and roof angle was fixed at 25°. This roofing angle was selected as it is within the range of those found in practice.

The FRWTF was fitted with a 4.0 m diameter axial fan to produce a wind field up to a 10 m/s (\pm 10 %). The wind velocity distribution was verified using a hot wire anemometer array. To track the evolution of the size and mass distribution of firebrands, a series of water pans was placed downstream of the assemblies.

The ignition method used in this study was developed based on lessons learned from igniting wall assemblies for firebrand collection [3]. This ignition method provides repeatable conditions to collect firebrands during combustion of the assemblies. The effect of wind speed on firebrand size and mass distribution generated from a given roofing assembly configuration is an important parameter to study, as large-scale disastrous WUI and urban fires are reported under windy conditions. For this study, each roofing assembly was ignited using propane gas (commercial grade) T-shaped burner with a heat release rate of 26 kW positioned adjacent to the assemblies for 10 min under conditions of no wind. T-shaped burners were superior to point ignition methods such as a single propane torch. The T-shaped burner was placed on the outside of the roofing assemblies, at the leading edge, since the purpose of this study was to simulate ignition from an outside fire. If the burner was applied with wind (e.g., 6 m/s), flaming

combustion of the assembly was difficult to achieve due to convective heat loss from the applied wind flow. Another advantage of ignition without wind (with the burner applied for the same duration for each assembly) was that it provided more repeatable initial conditions for the experiments. Specifically, the area exposed to direct flame contact (1.44 m^2) was similar for a given assembly. If igniting with wind, in addition to large convective heat loss, the contact area of the flame onto the OSB or plywood surface of the assembly became unsteady.



Figure 1 Drawing Schematic of roofing assemblies used for the experiments. Top view is shown. Either OSB or Plywood was used as base sheathing.

If an ignition time less than 10 min was applied, once the wind field was added, the roofing assembly would self-extinguish. If an ignition time longer than 10 min was applied, the roofing assembly was observed to be consumed a great deal before the application of wind, thus rendering it unstable for firebrand collection (i.e. large holes were formed due to long ignition combustion process compromising the structural integrity).

With the application of the T-shaped burner to the assembly under no wind for 10 min duration, flaming combustion was observed on the exterior of roofing assembly. Once the burner was switched off, the wind tunnel was switched on. Firebrands were collected until the assemblies were consumed to such a degree that they could no longer support themselves (loss of structural integrity). Firebrands were collected using multiple water pans placed behind the roofing assembly. Water was necessary to stop the combustion of the generated firebrands. After deposition into the water pans, firebrands were filtered from the water using a series of fine mesh filters. Firebrands were dried in an oven at 104 °C for 24 h (to remove moisture).

3. RESULTS AND DISCUSSIONS

Experiments were conducted for a wind speed of 6 m/s. Image analysis software (commercially available packages exist; these are not identified since no endorsement is desired) was used to determine the projected area of a firebrand by converting the pixel area using an appropriate scale factor. The projected areas with the maximum dimension and the second maximum dimension of three dimensions were measured. Images of specific shapes that have known areas (e.g. circles) were used to determine the ability of the image analysis method to calculate the projected area. The standard uncertainty in determining the projected area was ± 10 %. Repeat measurements of known calibration masses were measured by the balance which was used for the firebrand mass analysis. The standard uncertainty in the firebrand mass was approximately ± 1 %.

Fig. 2 displays a comparison of the collected firebrand size and mass distribution collected from these experiments. To more clearly show the results, the graph is plotted using a loglog scale and direct comparison of sheathing materials (OSB vs plywood) is shown. Both OSB, plywood, and wood stud combustion resulted in firebrand production.



Figure 2 Firebrands collected from roofing assembly combustion (6 m/s). The base sheathing material was varied.

4. SUMMARY

An experimental protocol was developed to ignite full-scale roofing assemblies and quantify the degree of firebrand production during the combustion process. While more work remains, the experimental method demonstrated the capability to discern differences in firebrand production from different base sheathing materials.

5. ACKNOWELDGEMENTS

Mr. Marco Fernandez of NIST is acknowledged for shipping materials.

6. REFERENCES

- [1] Manzello, S.L., et al., (2017), NIST SP 1213.
- [2] Maranghides, A., et al., (2015), NIST TN1910.
- [3] Suzuki, S., et al., (2013) Proc. Comb. Inst. 34:2479-2485.

EXPERIMENTAL STUDY ON IGNITIONS OF THATCHED ROOFING ASSEMBLIES EXPOSED FROM FIREBRANDS

Sayaka Suzuki¹ and Samuel L. Manzello² ¹National Research Institute of Fire and Disaster, Japan ²National Institute of Standards and Technology, USA

1. INTRODUCTION

It is well known that structures with thatched roofing assemblies are prone to ignition [1]. Counter measures to prevent structures with thatched roofing assemblies igniting have been studied and external sprinklers were implemented, especially for historical buildings, such as UNESCO world heritage sites in Japan [1]. It is believed that firebrands penetrate the thatched roofing assembly and smolder, which eventually leads to ignition and damage to buildings [1]. Yet, fundamental ignition studies of thatched roofing assemblies by firebrands were never carried out. Past experiments with thatched roofing assemblies were performed with a burning wood crib placed on the roof top or using a gas burner, rather than simulating the actual phenomena of firebrand showers attacking the thatched roofing [2, 3].

In this study, the reduced-scale continuousfeed firebrand generator (the continuous-feed baby dragon) was used to simulate firebrand showers attacking a mockup thatched roofing assembly under a 6 m/s wind in NRIFD's wind facility.

2. EXPERIMENTS

Experiments were performed by using the continuous-feed baby Dragon, shown in **Fig. 1**. Details of the experimental facility and experiments using other mockup roofing assembly types are elsewhere [4]. The reduced-scale continuous-feed firebrand generator consisted of two parts; the main body and continuous feeding component. The capability of a smaller-sized firebrand generator to develop continuous firebrand showers has been described [4].

For all experiments here, Japanese Cypress wood chips were used to produce firebrands. These were provided from a supplier and upon arrival, these chips were filtered using a 1 cm mesh to remove very fine wood pieces. The chips were also oven dried. These size wood pieces were selected to produce firebrands with larger projected area at a specific mass than that used in our prior studies using continuous firebrand generation focused on vegetative firebrands. The wood feed rate used here was 80 g/min, which is near the upper limit for this reduced-scale firebrand generator [4].



Figure 1 Experimental Settings.

As the base of the fan used to generate the wind in the NRIFD facility is located 1.6 m from the floor, the conveyer was placed under a custom stage designed for experiments when using NRIFD's wind facility. The wind field exits from a 4.0 m diameter fan, and it is possible to generate wind speeds up to 10 m/s. The flow field was measured to be within \pm 10 % over a cross-section of 2.0 m by 2.0 m.

When the blower was set to provide an average velocity below 4.0 m/s measured at the exit of the firebrand generator with no wood pieces loaded, insufficient air was supplied for combustion and this resulted in smoke being generated in addition to firebrands. Above 4.0 m/s, smoke production was mitigated but then many firebrands produced were in a state of flaming combustion as opposed to glowing combustion. In these experiments, glowing firebrands were desired [4].

A mock-up thatched roofing assembly was constructed for this experimental series and the size of the mock-up was 0.9 m (W) x 1.1 m (H) x 0.4 m (maximum thickness) with an angle of 45 degrees. The thatched roofing assembly was made from water reed (main materials), bamboo and wood frame, shown in **Fig. 2**. The thatched roofing assembly was placed at 0.5 m downwind from the

baby Dragon. This location was picked for the roof to receive adequate amount of firebrand showers.



Figure 2 Thatched roofing assembly used for the experiment.

3. RESULTS & DISUSSIONS

Not all firebrands from the Dragon arrive at the roof and the number of firebrands arriving on thatched roof was measured to be 7.3 /sec. Most of firebrands landing on the roof landed on the lower half of roof, but again not all firebrands landing on roof stayed on the roof. Some were blown away, some penetrated between the water reeds, and some stayed on the surface for a while. Images of an experiment were shown in **Fig. 3**. No smoke was observed until ignition and once ignited, fire was rapidly spread within the thatched roofing assembly.

4. SUMMARY

Experiments with thatched roofing assemblies were performed by using the continuous-feed reduced-scale firebrand generator in NRIFD's wind facility to investigate fundamental ignition mechanism. The wind speed of 6 m/s was selected to observe the ignition and flame spread of thatched roofing assembly. It is revealed that firebrands penetrated into the thatched roofing assembly, sometimes unseen from the outside, resulting in ignition and ultimately rapid flame spread. Basic information obtained in this study would be useful to compare with the results using burning wood cribs done in past research and subsequently to evaluate and develop effective counter measures to protect historical structures with thatched roofing assemblies.

5. REFERENCES

[1] FDMA, Final Report on methods to protect Cultural Property from fire, 2011.03. http://www.fdma.go.jp/html/data/tuchi2304/pdf/23 0422-index.pdf (in Japanese) [2] Murata, S. et al, Proceedings of annual JAFSE conference, 250-251, 2016. (in Japanese) [3] Murata, S. et al, Proceedings of annual JAFSE conference, 256-257, 2017. (in Japanese) [4] Suzuki, S., and Manzello, S.L., Fire Safety Journal, 91, 784-790, 2017.







Figure 3 Images of thatched roofing assembly during an experiment under a 6 m/s wind.

Proceedings of the ASME 2018 13th International Manufacturing Science and Engineering Conference **MSEC2018** June 18-22, 2018, College Station, Texas, USA

MSEC2018-6707

TOWARDS A STANDARDS-BASED METHODOLOGY FOR EXTENDING MANUFACTURING PROCESS MODELS FOR SUSTAINABILITY ASSESSMENT

Arvind Shankar Raman

School of Mechanical, Industrial and Manufacturing Engineering Oregon State University Corvallis, Oregon 97331, USA shankara@oregonstate.edu

Karl R. Haapala School of Mechanical, Industrial and Manufacturing Engineering Oregon State University Corvallis, Oregon 97331, USA Karl.Haapala@oregonstate.edu

KC Morris

Systems Integration Division National Institute of Standards and Technology, Gaithersburg, MD 20899, USA kcm@nist.gov

ABSTRACT

Over the past decade, several efforts have characterized manufacturing processes from a sustainability perspective. In addition, frameworks, methodologies, and standards development for characterizing and linking unit manufacturing process (UMP) models to construct manufacturing system models for supporting sustainability assessment have been pursued. In this paper these research efforts are first briefly reviewed, and then, ASTM standards derived from this work are described and built upon. The contribution of this research is to demonstrate how more formalization of these prior efforts will facilitate systematic reuse of developed models by encapsulating different aspects of complex processes into reusable building blocks. The research proposes a methodology to define template UMP information models, which can further be abstracted and customized to represent an application-specific, upgraded manufacturing process. The methodology developed is based on the ASTM standards of characterizing manufacturing process for sustainability characterization. The approach is demonstrated for analyzing manual and computer numerically controlled (CNC) machining processes.

1. INTRODUCTION

While manufacturing has primarily focused on gaining profits and securing market share, manufacturing with a sustainability focus has gained traction [1]. Globally, a number of sustainable manufacturing efforts have been pursued in response to growing societal concerns over the non-monetized impacts of manufacturing [2,3]. In particular, green, or environmentally-responsible design and manufacturing philosophies have paved the way for assessment tools that promote sustainable manufacturing during the conceptual and early design stages of the product life cycle [4].

A variety of software tools are available to perform product life cycle assessment (LCA), such as SimaPro and GaBi [5,6]. LCA tools are able to guide manufacturers in making more informed decisions about the environmental impacts of their production processes and supply-chain activities [7,8]. These tools can sometimes also offer insight into product- and processrelated economic and social impacts during design and, thus, aid manufacturers in developing and implementing sustainable product design and manufacturing modifications. One major drawback of LCA tools, however, has been the generic representation of manufacturing processes for analysis [8,9].

Manufacturing phase, or gate-to-gate, LCAs often do not address process-specific impacts, but rather utilize generalized process models, which are often not representative of the machine tool setup in the setting evaluated [9,10]. For example, a comparative LCA for machining a one-kilogram sphere and a one-kilogram cube would yield identical results, since the machining process model in the database reports impact based on the mass of the part processed. However, since the machine setups, cutting paths, and volume of material removed would all be different, the impacts also would vary significantly between the two parts. To overcome this limitation of manufacturing process models, efforts have been undertaken to improve manufacturing process characterization. One aim of these efforts is to enhance the ability of LCA tools to more accurately assess the environmental impacts of unit manufacturing processes

Haapala, Karl; Morris, Katherine; Shankar Raman, Arvind.

(UMPs) [8]. Further, more accurate process models will enhance manufacturing system evaluations for other metrics and indicators (e.g., cost and productivity).

UMPs have been defined as "the individual steps required to produce finished goods by transforming raw material and adding value to the workpiece as it becomes a finished product" [11]. A UMP has also been defined as "the smallest elementary manufacturing activity required for a specific taxonomological [referring to a taxonomy of manufacturing process types] transformation and composed of machines, devices, or equipment" [12]. As noted above, characterizing a UMP enables a deeper understanding of the process and improves processlevel decision making. Further, being able to link characterized UMPs to form a manufacturing system model will enable system-level characterization and enhance sustainability assessment.

To demonstrate the operational application of these concepts, the research presented herein focuses on development of a methodology for constructing reusable abstractions of UMPs based on two ASTM standards (ASTM E2986-15 and ASTM E3012-16). First, related prior work is briefly introduced. Next, the methodology is presented and demonstrated for several abstractions of a machining process (manual and computer numerically controlled milling). Finally, several advantages of the methodology are presented.

2. BACKGROUND

Several efforts have addressed the development of methods for UMP characterization. One of the initial efforts was under the Unit Process Life Cycle Inventory (UPLCI) project [13]. The goal of the UPLCI project was to formalize a systematic framework for inventory analysis of the manufacturing phase of LCA. By dividing a manufacturing process into sub-processes, representative models are more reliable and precise. Thus, the UPLCI framework proposed the creation of a toolset that would help compile life cycle inventories (LCIs) for UMPs to support LCA. The framework could enable manufacturing system analyses by aggregating LCI data for individual manufacturing processes involved in the production of a part [5,6,14].

UPLCI framework development work was later undertaken in conjunction with the Cooperative Effort on Process Emissions (CO2PE!) in Manufacturing, an initiative undertaken by the International Academy for Production Engineering (CIRP) [8]. CO2PE! was launched to address the lack of precise and specific environmental impact data in LCI databases. The effort aimed to compile a repository of data from research labs and other organizations from various geographic locations. The focus was to emphasize the coordination of the various global efforts in consolidating and analyzing environmental impacts of UMPs toward sustainability characterization of manufacturing [10,14]. In merging these two initiatives, the UPLCI effort formed a screening method for building LCI databases, while the CO2PE! effort presented an in-depth approach for quantifying LCI data. No recent developments toward LCI data collection and sustainability characterization have been reported under this initiative

In addition to these two relatively ad hoc efforts, the International Organization on Standardization (ISO) published the ISO 20140:2013 standard, titled "Automation systems and integration - Evaluating energy efficiency and other factors of manufacturing systems that influence the environment" [15]. This standard instituted a method for environmental performance evaluation (EPE) of individual manufacturing processes by assessing the energy efficiency and other factors of manufacturing systems. The standard helps in conducting EPEs of manufacturing systems by aggregating UMP EPE data.

Despite the fact that these methods and standards have helped in characterizing discrete manufacturing processes for sustainability evaluation, there has not been much recent development. Also, these prior methods have focused on developing distinct and specific information models of UMPs. Developing these information models from scratch requires a high level of expertise and knowledge in characterizing specialized manufacturing processes and, thereby, also requires significant time and effort. Having robust information models that can be reused and expanded upon to specify configurations of manufacturing processes would greatly benefit manufacturers and researchers alike. Prior methods of UMP model development have not focused on creating reusable abstractions for information models that can be instantiated for sustainability characterization in a variety of settings.

To overcome this inherent gap in existing methods, research collaborations with the ASTM International working group have engaged in standards development to support sustainable manufacturing. The collaboration contributed to the ASTM E2986-15 standard, titled "Standard Guide for Evaluation of Environmental Aspects of Sustainability of Manufacturing Processes," which provides a method for the evaluation of manufacturing process-related environmental impacts [16]. A second standard, ASTM E3012-16, titled "Standard Guide for Characterizing Environmental Aspects of Manufacturing Processes," further develops ASTM E2986-15 [17]. This standard helps analysts and decision makers in the systematic characterization of the environmental impacts of a UMP. It provides a defined structure for representing a UMP, enabling industry practitioners and researchers to more easily share UMP models [18,19]. The structure is formalized in XML (eXtensible Markup Language) using XSD (XML Schema Definition). The standard also provides for the specification of variables for linking, or composing, multiple UMPs for sustainability characterization of manufacturing systems. However, the implementation of this concept is not fully developed in the standard. Composability is enabled by the use of linking variables defined for a UMP that are appropriate for a subsequent UMP

The methodology presented below extends prior framework development efforts based on ASTM standards. For example, we previously reported a complementary framework which solely focused on composing UMPs to enable sustainability assessment of manufacturing systems [20,21]. However, the previous

Haapala, Karl; Morris, Katherine; Shankar Raman, Arvind,

framework lacked aspects of model reusability and extensibility, which is being addressed in this research. Other efforts at NIST have explored the needs for creating a repository of UMP models [22–24]. The reuse of models in such a repository is an ongoing research challenge that this work addresses. Here, we posit that information models can be created for a specific manufacturing process and then abstracted to characterize variations of that manufacturing process. Using these abstractions of the UMPs, process model composability can be performed to evaluate systemic sustainability assessment.

3. METHODOLOGY

As noted above, the aim of the research presented here is to build on the existing ASTM E3012-16 standard to improve the reusability, extensibility, and composability of UMP models. The standard provides a graphical model structure to represent UMPs (Fig. 1). This standard structure defines five aspects: inputs, outputs, resources, product and process information, and transformation equations. Inputs indicate the types of energy, materials, and consumables flowing into the process. Outputs indicate the product and, when relevant, co-products and byproducts, types of wastes/emissions, and process feedbacks (e.g., status of consumables and tools). Resources define information related to resources used by the process, such as tooling/fixtures, equipment, software, and people. Product and process information is the information needed to enable transformation functions (equations), which includes information related to the material, part, process plans, and control programs.





The model structure includes transformation equations. These equations describe the physical transformations that occur based on pertinent engineering information (for example, part dimensions and material properties) into information describing the physical outputs of the process, including information transferred to the subsequent UMP model. These equations maybe also used to calculate sustainability metrics and key performance indicators (KPIs), typically represented as product and process information for the manufacturing process. Thus, the transformation equations provide a physical basis for characterizing the sustainability performance of the UMP. The standard provides a formal representation of all five UMP aspects using an XML schema. Since these aspects are represented as element blocks in the standard model structure, they are easy to read, edit, and expand upon from a software programming perspective.

The research presented here contributes a methodology for abstracting an existing model and molding it into a specific model for a particular application (instantiation). We propose the development of template models that can be reused, extended, and composed. Figure 2 shows the activities comprising the methodology.



This methodology defines what constitutes a template model, devises a method to develop and represent a template model, and presents an approach for abstracting models for extensibility. The remainder of this section discusses each of these activities in greater detail. The next section then demonstrates the activities using a case study for milling operations.

3.1 Define the template model for the process

The ASTM standard guides researchers and industry practitioners in developing process-specific UMP models. Here, we propose the concept of a template model to represent discrete manufacturing processes of a particular classification. We then evaluate the feasibility of extending these template models to use-specific models. First, template models will need to be established for each class of manufacturing process. A template model (abstraction) can be defined as a model that completely characterizes the most simplistic instantiation of a manufacturing process that has varying levels of machine configurations. Thus, a template model would be developed for the most basic machine form for a manufacturing process that utilizes multiple alternatives of machine types. The template model would be able to be expanded to accommodate models for similar machine configurations or higher complexity machine configurations. This can be explained using the example of a milling operation.

A UMP model for a process using a manual milling machine is considered as a template model for milling (multi-point material removal). A manual knee and column mill is understood to be the basic physical representation of most vertical milling machines [25]. While the spindle is electrically powered, all other capabilities of the machine are manually controlled (e.g., spindle speed, feed rate, and depth of cut). By adopting the manual milling machine to create the template model, any milling machine configurations with enhanced or upgraded

Haapala, Karl: Morris, Katherine; Shankar Raman, Arvind,

capabilities would be considered as extensions of the template (manual milling) model.

While it is expected that these template models can be applied to the majority of upgraded machine configurations, there will be complex machine/process models that require further processing to facilitate abstraction from template models. Such cases would derive from machine configurations that are combinations of multiple manufacturing processes within a single machine configuration, e.g., five-axis milling or hybrid manufacturing. The methodology is generally applicable to these complex configurations, but requires a complete and thorough understanding of the machine and process to model accurately.

3.2 Represent template model in a structured format

Next, the identified template model must be represented in a structured manner to enable software tool implementation. The schema proposed in the standard is currently being applied in this research. Changes in the schema do not affect the proposed methodology, which is independent of the structure and can be applied to any XML schema definition. Software tools will facilitate adoption and use of manufacturing process and system modeling and analysis. We investigated how UMP models could be represented for software implementation using XML, since it is capable of handling functional modeling of manufacturing systems [26,27]. We found that XML schema can handle complex relationships, has a defined structure, which is beneficial for model development, and is amenable to extension for software programming [28].

In addition, XML models are capable of handling the research-specific needs for model reusability, extensibility, and composability due to their structured and compartmentalized way of representing data [29]. Also, by representing models as XML documents, parsing, analyzing, and processing data is not software platform dependent and can be handled by any language that can work with XML. In the current format, the transformation equations are in string, written as free form. The National Institute of Standards and Technology (NIST) has an ongoing effort to create a web tool, called the UMP Builder [30] to model and represent UMPs as a standardized XML document. The UMP Builder handles these transformation equations as MathML functions and can be processed through MathML interpreters to be read as executable computations. The language is relatively easy to learn for non-expert practitioners, supporting adoption of the standard. For industry practitioners and researchers to perform sustainability assessment, models must be represented as real-time operational standardized XML documents. By conforming to the standards, these models can be used by other researchers and practitioners by expanding them into application-specific process models.

3.3 Extend template model to represent an instantiation

The next step focuses on extending the template models to represent more complex, use-specific manufacturing systems. Complex variations could include extensions of physical inputs/outputs or instantiations of product and process information to reflect a specific physical piece of equipment and would require modification of the transformation equations. Instantiations of these template models to depict more complex variations of the processes studied are called extensions. Since the template models are created based on ASTM E3012-16, the model structure inherently allows for extensions of unit process models. For example, let model UMP A in Fig. 3 be the template model for all similar manufacturing processes, A. Let UMP A1 be a complex variation of UMP A. To develop a model for UMP A1, an instantiation of the template model of UMP A is generated. This model is then extended using information related to UMP A1 as a layer (Layer A1) of template model UMP A. We posit that extensibility can not only be done by building upon template models using such layers, but also by building upon already extended models using additional layers, which will develop higher order UMP models. This concept is illustrated for milling in the next section.



FIGURE 3. EXTENSIBILITY OF TEMPLATE MODEL UMP A TO FORM MODEL UMP A1 USING LAYER A1

Another important aspect of extensibility is that the layers that can be added to the template model are not just the higher order variants of the primary process. For example, auxiliary systems (e.g., exhaust gas pressure control systems, monitoring equipment, and electric boosting systems) that are essential to support the manufacturing process, but might not directly modify the workpiece, can be added as layers to create specific models of the equipment in use. To be amenable for reuse, the template models require certain information and characteristics to be instantiated and expanded. To establish these characteristics, template models have been developed under this research for a few manufacturing processes (i.e., milling, inertial welding, and heat treatment using a natural-gas fired furnace). These efforts

Haapala, Karl; Morris, Katherine; Shankar Raman, Arvind

(to be reported later) have aided in characterizing the processes and in developing a generic model structure for these processes.

4. DEMONSTRATION CASE: MILLING PROCESS

To demonstrate the application of the methodology, a template model for a manual milling machine was developed using the steps above. The template model was then extended to model a milling machine with computer numerical control (CNC) of its x-axis and y-axis movement (referred to here as a "two and a half axis milling machine"). It was further extended to consider use of a lubrication system. These extensions were created by instantiating the template model and adding information layers, representing a two and a half axis milling machine (Layer 1) and the lubrication system (Layer 2), as shown in Fig. 4.



FIGURE 4. EXTENSIBILITY OF MANUAL MILLING FOR TWO AND HALF AXIS MILLING WITH LUBRICATION SYSTEM

A representative template model of manual milling machine is illustrated in Appendix A, as specified by the standard. The representative manual milling machine has just the spindle powered in this demonstration case. The model represented here primarily performs an energy characterization of the milling process. Energy characterization provides a detailed view to improve energy performance of the process. The development of the template model begins with capturing the physical inputs and outputs of the process. Inputs and outputs are a description of the physical inputs and outputs (e.g., bar stock, work in progress, electrical input, and type of waste) to the manufacturing process. Product and process information holds information related to the product data (e.g., length, width, thickness, and density), process data (e.g., speed, feed, and depth of cut), and sustainability metrics (e.g., energy consumption, total cost of energy, and mass

of greenhouse gas emissions) and KPIs (e.g., energy consumption per kilogram of part). The transformation equations contain the mathematical functions required to quantify the desired metrics and KPIs. UMP model resources capture all the information pertaining to process resources (e.g., software, tools, fixtures, and workers). Resources do not have a direct effect on either the product or process, but are needed to aid in the functioning of the machine.

An equivalent information model representation is also developed based on the ASTM standard, and documents the five aspects of a UMP model (i.e., inputs, outputs, product/process information, resources, and transformation equations) as elements in the XML documents. Appendix B reports the representative information model for the manual milling case in XML format [31–34].

With the template model in place, it can be extended to create models of different milling processes based on the specific machine infrastructure. As noted above, for this case study, two and a half axis milling with lubrication is considered. The two and a half axis milling model can be created by extending the template model by adding a CNC table (Layer 1). Similarly, the lubrication system (Layer 2) is an auxiliary system to the milling machine. Thus, the extended model with Layers 1 and 2 added captures the information related to both the two and a half axis milling machine and the lubrication system, as shown in Appendix C. Information such as coolant flow rate, volume of coolant used during the milling operation, and the energy required to run the lubricant system represent some of the additional information pertaining to the lubricant system layer. Similarly, table motor power, basic power, and basic time represent some of the added information that is related to the two and half axis layer. Added information is indicated by bold blue text in the UMP representation of two and a half axis milling with lubrication. In addition to information being added to the product and process information, extension of the template models requires editing of the transformation functions to accommodate the addition of layers. For example, on-site energy in the manual milling model is the energy supplied to the motor, whereas, in the two and half axis milling model, it is the sum of basic energy (energy required for part setup and idle) and energy supplied to the table motors and lubricant system. Development of the models herein relied on prior research [35-38].

Similar to the template model, the extended UMP model is an energy-based model. XML representation of the template model can be updated by editing individual elements of the XML document reported in Appendix B to accommodate the two and a half axis milling machine layer and the lubricant system layer. This additional XML information is presented in Appendix D. Input and output elements remain unchanged as the information is the same in both the models. The product and process information for Layer 1 and Layer 2 is appended to the product and process information for the manual milling model. Similarly, new transformation equations for the extended layers are added as new elements. This XML representation captures the different aspects (inputs, outputs, product and process information,

resources, and transformations) and adheres to the standard representation.

5. CONCLUSION

The methodology presented establishes a mechanism to create reusable abstractions (models) of unit manufacturing processes (UMPs) for characterizing the sustainability performance of a variety of manufacturing processes and systems. The methodology facilitates the creation of extensible and composable UMP models, and enables manufacturers and researchers to develop more accurate system models for sustainability characterization by tailoring existing validated models for their specific needs. This approach offers several advantages over current and prior practices:

- Straightforward development of basic and extended UMP models supported by a standardized model structure;
- · Simplified tracking for evaluating a UMP model and validating modifications made to extend a model;
- · Improved model reusability and extensibility through multilayer buildup of an existing validated UMP model; and
- · Maintained reusability, extensibility, and composability characteristics of the UMP model after extension.

The methodology proposed in this research is portable (UMP models can be incorporated into computer-aided engineering tools) and scalable (models can be developed for processes and systems of varying complexity from a variety of domains). To realize the vision of facilitated model creation, extensibility, and application to sustainable manufacturing characterization, future work will be needed to build a repository of template models and extension layers for a broad set of manufacturing processes [22]. This effort can be accelerated by creation of software capable of validating the models, as well as tools that can aid decision makers from various domains in composing the models for system analysis.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the National Institute of Standards and Technology (NIST Award No. 70NANB16H223) for their support of this research. We would especially like to thank William Bernstein, David Lechavalier, Swee Leong, and Kevin Lyons, for their generous inputs to the research.

REFERENCES

- Jovane, F., Westkämper, E., and Williams, D., 2009, The [1] ManuFuture road: towards competitive and sustainable high-adding-value manufacturing, Springer, Berlin.
- [2] Tilton, J. E., 1996, "Exhaustible resources and sustainable development: Two different paradigms," Resour. Policy, 22(1-2), pp. 91-97.
- Allouche, J., 2011, "The sustainability and resilience of [3] global water and food systems: Political analysis of the interplay between security, resource scarcity, political systems and global trade," Food Policy, 36, pp. S3-S8.

- Dornfeld, D., and Wright, P., 2007, "Technology Wedges' [4] for Implementing Green Manufacturing," NAMRI/SME, 35(1), pp. 193-200.
- Overcash, M., and Twomey, J., 2012, "Unit Process Life [5] Cycle Inventory (UPLCI) - A Structured Framework to Complete Product Life Cycle Studies," Leveraging Technology for a Sustainable World, D.A. Dornfeld, and B.S. Linke, eds., Springer Berlin Heidelberg, pp. 1-4.
- [6] Overcash, M., Twomey, J., and Kalla, D., 2009, "Unit Process Life Cycle Inventory for Product Manufacturing Operations," ASME International Manufacturing Science and Engineering Conference, ASME, West Lafayette, IN, pp. 49-55.
- Reap, J., Roman, F., Duncan, S., and Bras, B., 2008, "A [7] Survey of Unresolved Problems in Life Cycle Assessment: Part 2: Impact Assessment and Interpretation," Int. J. Life Cycle Assess., 13(5), pp. 374–388.
- Kellens, K., Dewulf, W., Duflou, J. R., and others, 2010, [8] "The CO2PE!-initiative (cooperative effort on process emissions in manufacturing)," International framework for sustainable production., Netherlands, p. 13.
- [9] Overcash, M., Duflou, J., Kellens, K., Looman, K., Dewulf, W., Sutherland, J., Twomey, J., and Isaacs, J., 2012, "Unit Process Life Cycle Inventories CO2PE! -UPLCI Workshop," CIRP Annals - Manufacturing Technology, Berkeley, CA.
- [10] Duflou, J. R., Kellens, K., Renaldi, Guo, Y., and Dewulf, W., 2012, "Critical Comparison of Methods to Determine the Energy Input for Discrete Manufacturing Processes," CIRP Ann. - Manuf. Technol., 61(1), pp. 63-66.
- [11] Mani, M., Madan, J., Lee, J. H., Lyons, K. W., and Gupta, S. K., 2014, "Sustainability Characterization for Manufacturing Processes," Int. J. Prod. Res., 52(20), pp. 1 - 18
- [12] Garretson, I. C., 2015, "A Unit Manufacturing Process Characterization Methodology and Supporting Terminology for Sustainable Manufacturing Assessment," Master of Science, Oregon State University.
- [13] Luttikhuis, E. O., Toxopeus, M. E., and Overcash, M., 2013, "Applying Unit Process Life Cycle Inventory Methodology (UPLCI) in Product/Packaging Combinations," Re-engineering Manufacturing for Sustainability, A.Y.C. Nee, B. Song, and S.-K. Ong, eds., Springer Singapore, pp. 15-20.
- [14] Duflou, J. R., Sutherland, J. W., Dornfeld, D., Herrmann, C., Jeswiet, J., Kara, S., Hauschild, M., and Kellens, K., 2012, "Towards Energy and Resource Efficient Manufacturing: A Processes and Systems Approach," CIRP Ann. - Manuf. Technol., 61(2), pp. 587-609.
- [15] International Organization for Standardization [ISO], 2013, "Automation systems and integration -- Evaluating energy efficiency and other factors of manufacturing systems that influence the environment (ISO 20140)."
- [16] ASTM, 2015, "Standard Guide for Evaluation of Environmental Aspects of Sustainability of Manufacturing Processes (ASTM E2986)."

Haapala, Karl; Morris, Katherine; Shankar Raman, Arvind,

- [17] ASTM, 2016, "Standard Guide for Characterizing Environmental Aspects of Manufacturing Processes (ASTM E3012)."
- [18] Mani, M., Larborn, J., Johansson, B., Lyons, K. W., and Morris, K. C., 2016, "Standard Representations for Sustainability Characterization of Industrial Processes," J. Manuf. Sci. Eng., 138(10), p. 101008.
- [19] Rebouillat, L., Barletta, I., Johansson, B., Mani, M., Bernstein, W. Z., Morris, K. ., and Lyons, K. W., 2016, "Understanding Sustainability Data through Unit Manufacturing Process Representations: A Case Study on Stone Production," Procedia CIRP, 57, pp. 686-691.
- [20] Smullin, M. M., 2016, "An Information Framework for Composing Manufacturing Processes," Master of Science, Oregon State University.
- [21] Smullin, M. M., Iman, Z., and Haapala, K. R., 2017, "A Desktop Application for Sustainability Performance Assessment of Composed Unit-Based Manufacturing Systems," Proceedings of the 12th International Manufacturing Science and Engineering Conference, ASME, Los Angeles, CA, p. V004T05A022; 11 pages.
- [22] Bernstein, W. Z., Mani, M., Lyons, K. W., Morris, K. C., and Johansson, B., 2016, "An Open web-based repository for capturing manufacturing process information,' Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, ASME, p. V004T05A028; 8 pages.
- [23] Bernstein, W. Z., Bala Subramaniyan, A., Brodsky, A., Garretson, I. C., Haapala, K. R., Libes, D., Morris, K. C., Pan, R., Prabhu, V., Sarkar, A., Shankar Raman, A., and Wu, Z., 2017, "Research directions for an open unit manufacturing process repository: A collaborative vision," Manuf. Lett.
- [24] Brodsky, A., Krishnamoorthy, M., Nachawati, M. O., Bernstein, W. Z., and Menasce, D. A., 2017, "Manufacturing and contract service networks: Composition, optimization and tradeoff analysis based on a reusable repository of performance models," IEEE, pp. 1716-1725.
- [25] Groover, M. P., 2015, Fundamentals of Modern Manufacturing, Wiley, New York.
- [26] Sung, R. C., Ritchie, J. M., Lim, T., and Kosmadoudi, Z., 2012, "Automated generation of engineering rationale, knowledge and intent representations during the product life cycle," Virtual Real., 16(1), pp. 69-85.
- [27] Meihua, Z., Aiping, L., and Liyun, X., 2008, "Integrated Modeling and Information Interaction Mechanism of

Collaborative Production Information," Proceeding of the International Conference on Information Management, Innovation Management and Industrial Engineering, IEEE, Taipei, Taiwan, pp. 67–70.

- [28] Yadong, F., Ping, K., and Lei, Z., 2009, "Research and application of machine tools life-cycle management system based on web," Proceedings of the ISECS International Colloquium on Computing, Communication, Control, and Management, IEEE, Sanya, China, pp. 117-120.
- [29] Guha, R. V., Brickley, D., and Macbeth, S., 2016, "Schema.org: evolution of structured data on the web," Commun. Assoc. Comput. Mach., 59(2), pp. 44–51.
- [30] Bernstein, W. Z., Libes, D. E., and Lechevalier, D., 2018, "UMP Builder: capturing and exchanging manufacturing models for sustainability," Proceedings of the 46th SME North American Manufacturing Research Conference (NAMRC) and the ASME Manufacturing Science and Engineering Conference (MSEC 2018), ASME, College Station, Texas.
- Kara, S., and Li, W., 2011, "Unit process energy [31] consumption models for material removal processes, CIRP Ann. - Manuf. Technol., 60(1), pp. 37-40.
- Avram, O. I., and Xirouchakis, P., 2011, "Evaluating the [32] use phase energy requirements of a machine tool system," J. Clean. Prod., 19(6-7), pp. 699-711.
- [33] Kalla, D., Twomey, J., and Overcash, M., 2009, MR 3 Milling Process Unit Process Life Cycle Inventory.
- [34] Branker, K., 2011, A study of energy, carbon dioxide emissions and economics in machining: milling and single point incremental forming, Queen's University (Canada).
- [35] He, Y., Liu, B., Zhang, X., Gao, H., and Liu, X., 2012, "A modeling method of task-oriented energy consumption for machining manufacturing system," J. Clean. Prod., 23(1), pp. 167-174.
- [36] Overcash, M., Twomey, J., Isaacs, J., Duflou, J., and Sutherland, J., 2010, "Energy efficiency improvements in manufacturing life cycle technology for analysis of machines and products," Soc. Manuf. Eng. Tech. Pap. TP10PUB8 Dearborn MI, p. 5.
- [37] Li, W., Zein, A., Kara, S., and Herrmann, C., 2011, "An investigation into fixed energy consumption of machine tools," Glocalized Solut. Sustain. Manuf., pp. 268-273.
- [38] Dahmus, J. B., and Gutowski, T. G., 2004, "An environmental analysis of machining," Proceedings of the International Mechanical Engineering Congress and Exposition, ASME, Anaheim, CA, pp. 643-652.

Haapala, Karl: Morris, Katherine; Shankar Raman, Arvind,



APPENDIX A: MANUAL MILLING MACHINE MODEL REPRESENTED BASED ON ASTM STANDARD E3012-16

Haapala, Karl; Morris, Katherine; Shankar Raman, Arvind. "Towards a standards-based methodology for extending manufacturing process models for sustainability assessment." Paper presented at 46th SME North American Manufacturing Research Conference (NAMRC) and the ASME Manufacturing Science and Engineering Conference (MSEC 2018), College Station, TX, United States. June 18, 2018 - June 22, 2018.

APPENDIX B: XML REPRESENTATION OF MANUAL MILLING INFORMATION MODEL

<UMP name="Manual Milling" type="Material Removal" description="Manual milling model"> //INPUT SECTION <Input name="Bar stock" description= "Type of workpiece input to the process" category=""type="Workpiece" unit="" /> <Input name="Electrical Energy" description="Input electrical energy to the process" category=""type="Energy" unit="kWh" / > //PRODUCT AND PROCESS INFORMATION SECTION ProductProcessInformation name="Cutting force" description="Force on the cutting tool" category="Process"value="" unit="N" / >ProductProcessInformation name="Cutting speed" description="Speed of cut" category="Process"value="" unit="mm/sec" / > <ProductProcessInformation name="Feed" description="input feed of tool" category="Process"value="" unit="mm/s" / > <ProductProcessInformation name="Depth of cut" description="Axial depth of cut per pass" category="Process" value="" unit="mm" / > Conduct Process information name="Cutting time" description="Total cutting time" category="Process "value="" unit="s" /></product Process Information name="Cutting time" description="Volume of workpiece before operation" category="Product"value="" unit="mm^3" /> <ProductProcessInformation name="Efficiency of motor" description="Efficiency of motor" category="Process"value="" unit="" /> <ProductProcessInformation name="Volume removed" description="Total volume of material removed" category="Product"value="" unit="mm^3" / > <ProductProcessInformation name="Material removal rate" description="Rate of material removal" category="Product"value="" unit="mm^3/s" / > <ProductProcessInformation name="Tool wear rate" description="Rate of tool wear" category="Process"value="" unit=" mm^3/s" / > ProductProcessInformation name="Cutting power" description="Power required to cut material" category="Process"value="" unit="kW" / > <ProductProcessInformation name="Motor/Spindle power" description="Power at the motor/spindle" category="Process"value="" unit="kW" / > <ProductProcessInformation name="Generation rate of CO2" description="Mass of CO2 produced for unit energy use" category="Process"value="" unit="kg CO2/kWh" / > ProductProcessInformation name="Generation rate of CH4" description="CH4 produced in equivalent mass of CO2 for unit energy use" category="Process"value="" unit="kg CO2e/kWh"/>
<ProductProcessInformation name="Generation rate of NO2" description="NO2 produced in equivalent mass of CO2 for unit energy use"</pre> category="Processinformation name="Cutting energy" description="Roz produced in equivalent mass of 602 for this energy doe <ProductProcessInformation name="Cutting energy" description="Energy required to cut the part" category="Process"value="" unit="kJ" /> <ProductProcessInformation name="Energy onsite" description="Onsite energy consumption" category="Process"value="" unit="kJ" /> <ProductProcessInformation name="Energy offsite" description="Offsite energy consumption" category="Process"value="" unit="kJ" /> <ProductProcessInformation name="Energy offsite" description="Offsite energy consumption" category="Process"value="" unit="kJ" /> <ProductProcessInformation name="Energy offsite" description="Offsite energy consumption" category="Process"value="" unit="kJ" /> <ProductProcessInformation name="Total energy consumption" description="Total energy consumption" category="Process"value="" unit="kJ" /> Conduct ProcessInformation name="Unit cost of energy" description="Cost of 1kWh of energy" category="Process"value="" unit="\$/kWh" / > ProductProcessInformation name="Mass of GHG emissions" description="Greenhouse gas emissions in equivalent mass of CO2" category="Process"value=""unit="kg CO2e" / > //TRANSFORMATION SECTION <Transformation> <Equation description="Volume removed" set="">V_r = V_i - V_f</Equation>
<Equation description="Material removal rate" set="">MRR = v_c * f * d</Equation> <Equation description="Specific power" set="">P_cut= F_c * V_c</Equation> <Equation description="Specific power" set="">P_cut= F_c * V_c</Equation> <Equation description="Motor power" set="">P_m= P_cut / Eff</Equation> <Equation description="Cutting energy" set="">E_c = P_cut / Eff</Equation> <Equation description="Onsite energy" set="">E_o = P_cut / C_{Equation> <Equation description="Onsite energy" set="">E_o = P_m * T_c</Equation> <Equation description="Total energy consumption" set="">E_T = E_on + E_off</Equation> <Equation description="Total cost of energy" set="">C = E_T * C_e</Equation> <Equation description="Total cost of energy" set="">C = E_T * C_e</Equation> <Equation description="GHG emission" set="">GHG = E_T * (rCO2 + rCH4 + rNO2)</Equation> </Transformation> //RESOURCE SECTION <Resource name="Software" description="Software used for computer control" value="Linux CNC" / > <Resource name="Machine ID" description="ID of the machine that is being used" value="MM01" / > <Resource name="Operator" description="Name of operator" value="John Doe" / > //OUTPUT SECTION

<Output name="Finished Part" description="Number of workpieces produced in an hour" category=""type="workpiece" unit="" / > <Output name="Waste" description="Total waste of stock material" category="Waste"type="workpiece" unit="kg"/>

</UMP>

Haapala, Karl; Morris, Katherine; Shankar Raman, Arvind,

APPENDIX C: MANUAL MILLING EXTENDED TO TWO AND A HALF AXIS MILLING WITH LUBRICATION SYSTEM



APPENDIX D: ADDITIONAL INFORMATION MODEL FOR EXTENDED LAYERS (TWO AND A HALF AXIS MILLING AND LUBRICATION SYSTEM)

<ump description="Two and a half axis milling with lubrication system" name="Two-Axis Milling with Lubrication System" type="Material Removal"></ump>
//INPUT SECTION
<input/> // This section is the same as the input section in Appendix B
//OUTPUT SECTION from Appendix B
<output> // This section is the same as the output section in Appendix B</output>
//Additional product and process information specific to two and a half axis milling with lubrication system, appended to product and process information in Appendix B
<productprocessinformation category="Process" description="Volumetric flow rate of coolant" name="Coolant flow rate" unit="L/s" value=""></productprocessinformation>
<productprocessinformation category="Process" coolant="" description="volume" l="" name="volume" of="" used="" value="unit="></productprocessinformation>
<productprocessinformation category="Process" description="Time to setup and idle" name="Basic time" unit="s" value=""></productprocessinformation>
Product rocessified matter rate motor power description = "Power to table motor" category = "Process" value = "unit= kW" / >
<productprocessinformation category="Process" description="Energy to setup and idle" name="Basic energy" unit="kJ" value=""> <productprocessinformation category="Process" description="Energy for cutting" name="Ready energy" unit="kJ" value=""></productprocessinformation></productprocessinformation>
//Additional and updated transformation functions appended to transformation section in Appendix B
<transformation></transformation>
<equation description="Ready Energy" set="">E_Ready = P_m * T_cutting </equation>
<equation description="Onsite Energy Consumption" set="">E_on = E_Basic + E_Ready</equation>
<equation description="Onsite energy" set="">E_on = E_Basic + E_Ready</equation>
//RESOURCE SECTION from Appendix B
<resource></resource>

Model-based Cosimulation for Industrial Wireless Networks

Jing Geng*, Honglei Li*, Yanzhou Liu*, Yongkang Liu[†],

Mohamed Kashef[†], Richard Candell[†], Shuvra S. Bhattacharyya^{*‡},

* Department of Electrical and Computer Engineering, University of Maryland, College Park, USA

[†] Intelligent Systems Division, National Institute of Standards and Technology, USA

[‡] Department of Pervasive Computing, Tampere University of Technology, Finland

Email: {jgeng, honglei, yzliu}@umd.edu, {yongkang.liu, mohamed.hany, richard.candell}@nist.gov, ssb@umd.edu

Abstract—Wireless communications technology has the potential to provide major benefits in lowering the cost and increasing the efficiency of factory automation (FA) systems. However, design of FA systems that employ wireless networks involves stringent constraints on real-time performance and reliability, and requires the assessment of and experimentation with complex interactions among process control, factory topology construction (layout and connectivity of subsystems, such as machines, rails, etc.), and wireless communication. In this paper, we introduce a novel simulation framework to support such assessment and experimentation in the design of next-generation FA systems. Our simulation framework employs model-based design principles to enhance design reliability, and enable systematic and efficient integration of control, topology, and network modeling aspects. We demonstrate the utility of our framework through a case study that involves topology design and scalability analysis for a large class of FA systems. Our results demonstrate the ability of the proposed framework to provide insights on complex design trade-offs, while the underlying model-based features enhance efficient and reliable system-level integration.

I. INTRODUCTION

Wireless communications technology has the potential to provide major benefits in lowering the cost and increasing the efficiency of factory automation (FA) systems. In particular, wireless communication significantly reduces costs for network deployment, and allows for integration of sensing devices and associated monitoring capabilities in parts of the system that are difficult or impossible to connect through wired sensing installations (e.g., see [1]).

However, to realize the full potential of wireless communications integration with FA, significant challenges must be overcome in ensuring the reliability, real-time operational speed, and energy efficiency of industrial wireless networks. This is a major challenge due to the conflicting relationships among these metrics, and the complexity of the system-level design space, which includes key parameters at the layers of communication protocols, FA system topologies (placement and connectivity of units), and the embedded processing platforms (hardware and software) within the network nodes.

To enable efficient experimentation, validation, and optimization processes in the context of this complex design space, effective simulation methodologies are of critical importance.

In this paper, we introduce a novel simulation framework to address the challenge of simulating complex FA systems that are integrated with wireless communication techniques, networks, and connections. Our framework is novel in its model-based approach, where different models are employed to efficiently and flexibly integrate the design perspectives of FA topologies (system-level structure in the form of networks of machinery and control units), FA nodes (processing within individual nodes of a networked FA system), and wireless communications (protocol functionality and channel characteristics). Model-based design is an important paradigm for cyber-physical systems in which systems are specified through components that interact through formal models of computation [2].

In particular, we apply the *dataflow* model of computation in new ways to formally capture key aspects of FA system behavior and integrate these with discrete event models of computation that are used in state-of-the-art simulators for communication networks. Our application of dataflow principles enables dataflow-based analysis of FA process flows (system-level analysis) and signal and information processing that takes places within individual FA nodes, while providing efficient mechanisms for leveraging and interfacing to arbitrary discrete-event-based network simulation tools.

Our simulation framework is built on top of the *lightweight* dataflow environment (LIDE), which is a design tool for dataflow-based design and implementation of signal and information processing systems [3], [4]. Distinguishing features of LIDE are that it is minimally intrusive on existing design processes, and it is easily adaptable to different platformor simulation-oriented programming languages, such as C, C++, Verilog, VHDL, and MATLAB. LIDE achieves these features through its formulation in terms of a compact set of application programming interfaces (APIs) for constructing software or hardware components that communicate through dataflow interfaces, and for building up graphs as connections of such dataflow-based components. In our simulator prototype, we specifically employ LIDE-C, which provides Clanguage implementations of the LIDE APIs. However, the simulator can be retargeted readily to other languages by exploiting the retargetability of LIDE described above.

In our simulation framework for FA systems, we incorporate

extensions to LIDE for managing time because the dataflow model of computation, which LIDE is based on, is an untimed model with no built-in concept of time or time stamps. In dataflow, functional components (actors) synchronize with one another based only on the availability of data. The "timeextended" version of LIDE that we employ is referred to as τ LIDE, where τ represents the incorporated notion of time.

Based on its foundations in τ LIDE, we refer to our new factory simulation framework as τ LIDE-factorysim, which we abbreviate as TLFS (Tau Lide Factory Sim). The TLFS framework can be viewed as a design tool that applies timed dataflow concepts in novel ways to enable model-based cosimulation of factory process flows together with discrete-event simulation of communication networks that link physicallyseparated subsystems within the factories.

Rather than re-invent functionality for simulating communication networks, TLFS defines a structured process for integrating third-party network simulation tools with the dataflowbased factory simulation capabilities within TLFS. Thus, TLFS is agnostic to any specific network simulation tool, and can be adapted in a modular fashion to work with different network simulators depending on the designer's preference. In our first-version prototype of TLFS, we apply NS-3 [5] as the network simulation tool that is incorporated into the framework. However, the underlying architecture of TLFS is not dependent on NS-3, and can be adapted to work with other network simulation tools, as described above.

The remainder of this paper is organized as follows. Section II provides background on dataflow modeling that the later sections in the paper depend on. Section III discusses related work and positions the contribution of this paper in the context of other simulation techniques and tools. Section IV presents the architecture of the TLFS simulation framework. Section V presents an example to concretely demonstrate the process of modeling factory process flows in TLFS using its' underlying dataflow capabilities. Section VI presents experiments that demonstrate the utility of TLFS in assessing design tradeoffs across different combinations of factory process flow and networking configurations. Section VII summarizes the contributions of the paper and points to useful directions for further investigation.

II. BACKGROUND

Dataflow is widely used as a programming model in many areas of embedded and cyber-physical systems, especially in areas that involve significant emphasis on digital communication, signal and information processing, or both. Model-based design processes in terms of dataflow graphs provide useful features such as efficient retargetability to different implementation languages and platforms; support for deterministic, concurrent programming; and support for a wide variety of analysis and optimization techniques (e.g., see [6], [7]).

A variety of specific forms of dataflow has been developed and continues to evolve. These different forms of dataflow provide different trade-offs between generality (the class of

that is, the tractability of analysis and optimization problems or the availability of algorithms that address such problems. For background on dataflow models and various specialized forms of dataflow in embedded and cyber-physical systems, we refer the reader to [7].

The specific form of dataflow used in TLFS, inherited from the modeling approach in LIDE, is referred to as core functional dataflow (CFDF) [8]. CFDF is a general (Turing complete) form of dataflow that can be used to model a wide variety of applications. It is possible to apply more restricted models in TLFS that trade-off some of the generality of CFDF to enable more powerful analysis and optimization techniques. Such exploration is a useful direction for future work that builds naturally on the developments of this paper.

In the remainder of this section, we provide a brief introduction to CFDF semantics. This background is useful in understanding the methodology employed in TLFS for modeling FA systems.

In CFDF, as in other dataflow programming variants, the application system is represented as a directed graph in which vertices (actors) correspond to functional modules, and each edge e corresponds to a first-in, first-out (FIFO) buffer. This buffer represents the storage of data as it passes from the source actor of e to the sink actor of e. Each unit of data that passes along an edge is referred to as a token. Tokens can be of arbitrary data types, such as integers, floating point numbers, or objects of any complexity.

In a simulation model of a factory system, each token that passes through a given edge can represent some physical entity (such as a part) that moves from one factory subsystem (e.g., a machine) to another subsystem. Alternatively, a token can represent some form of information, such as a command that is sent from a factory controller module to a machine or rail that is controlled by the module.

In dataflow, the execution of an actor is decomposed into discrete units of execution called *firings* of the actor. On each firing, an actor consumes zero or more tokens from each of its input edges and produces zero or more tokens onto each output edge. The numbers of tokens that are produced and consumed in this way are referred to as the production and consumption rates that are associated with the firings.

An actor A in CFDF consists of a set $\mu(A)$ of one or more *modes*, where each $m \in \mu(A)$ corresponds to a distinct state in which the actor can fire. In other words, each firing f of the actor is a associated with a specific mode in $\mu(A)$. As a side effect, each firing determines the *next mode* $n \in \mu(A)$ of the actor, which is the mode that will be associated with the next firing of the actor. An initial next mode is assigned to each actor as part of the initialization of a CFDF graph.

In a given CFDF mode, the production or consumption rate on each incident edge must be constant — that is, all firings in the same mode must have identical production (consumption) rates on a given output (input) edge of the actor. This is a key "design rule" of CFDF. However, the rates can vary between distinct modes. A CFDF actor can fire whenever the number

Bhattacharyya, Shuvra; Candell, Richard; Hany, Mohamed; Jeng, Ging; Li, Honglei; Liu, Yanzhou; Liu, Yongkang. "Model-based Cosimulation for Industrial Wireless Networks." Paper presented at 14th IEEE International Workshop on Factory Communication Systems, Imperia, Italy. June 12, 2018 - June 15, 2018.
of tokens buffered on each input edge is at least equal to the consumption rate of the edge for the actor's next mode.

As a simple example of a CFDF actor, consider the integertyped switch actor that is illustrated in Figure 1. This actor has two input ports x_d (data) and x_c (control), and two output ports y_f and y_t . The ports x_d , y_f , and y_t have integer token type, while the token type of x_c is Boolean. The actor operates by copying each *i*th token consumed from x_d to y_f or y_t depending on whether the *i*th value consumed from x_c is false or true, respectively.



Fig. 1. A switch actor used to illustrate CFDF modeling.

This functionality can be implemented as a CFDF actor with three modes $\mu = \{m_c, m_t, m_f\}$. The subscripts here stand for "control", "true", and "false", respectively. In m_c , the actor consumes the next token ("control token") from x_c , and sets the actor's next mode to m_f or m_t depending on whether the value consumed is false or true, respectively. In m_f (m_t) , the actor outputs a copy of the most-recently consumed control token onto $y_f(y_t)$ and unconditionally sets the actor's next mode (back) to m_c .

One can easily verify, based on this description, that the number of tokens consumed and produced on each actor port is constant for each given mode. This is the basic "dataflow requirement" when decomposing an actor into modes for CFDF modeling. For example, the production rates associated with ports x_c, x_d, y_t, y_f in mode m_c are 1, 0, 0, 0, respectively.

For more detailed background on dataflow modeling that is oriented to embedded and cyber-physical systems, we refer the reader to [6], [7]. For further details on CFDF, we refer the reader to [8]. For foundational concepts related to the dataflow paradigm, we refer the reader to [9], [10].

III. RELATED WORK

A variety of approaches have been developed over the years that are relevant to the effective simulation of networked factory automation systems.

For example, Neema et al. have developed a model-based platform that integrates heterogeneous simulation tools for cyber-physical systems [11]. Düngen et al. demonstrate a simulation approach for parallel sequence spread spectrum signals in the context of industrial wireless communication environments [12]. Liu et al. present a simulation framework that integrates process control system modeling and wireless network modeling into a unified discrete-event simulator [13]. Bause et al. present a tool that integrates the ProC/B toolset for process chain modeling with the OMNeT++ tool for simulating communication networks [14]. Won et al. present a

tool that integrates the targeted dataflow interchange format, a dataflow tool for embedded signal processing, with the NS-2 environment for communication network simulation [15].

The work presented in this paper differs from the related work in its emphasis on dataflow-based modeling of factory process flows, and its general approach for integrating dataflow-based process flow models with arbitrary discreteevent-based tools for communication network simulation.

For concreteness, we demonstrate our models and methods by integrating a specific dataflow-based design tool, LIDE, with a specific network simulation tool, NS-3. However, the architecture of TLFS is formulated carefully in terms of abstract dataflow principles and model-based interfacing principles between dataflow and network modeling subsystems. This use of abstraction promotes the adaptability of TLFS to incorporate other network simulation tools as well as other tools for dataflow-based design. This flexibility is important given that a variety of different network simulation tools exist with different combinations of features (e.g., see [16], [17]).

In this way, our contribution is complementary to related work such as the works discussed previously in this section. For example, the network simulation techniques applied in the works by Liu [13] and Düngen [12] can be interfaced with dataflow-based factory models using TLFS, and similarly, the techniques for embedded software synthesis used in the work of Won [15] can be applied to networked factory automation systems. Such extensions of TLFS represent a useful direction for future work.

IV. SIMULATION APPROACH

In this section, we present our simulation framework, called TLFS, for FA systems that are integrated with wireless communication networks. We describe the architecture of TLFS, which is designed to be integrated with arbitrary simulators that are specialized for communication network simulation (CNS). In Section VI, we present experiments using this architecture with the integration of a specific CNS tool.

Figure 2 illustrates the architecture of TLFS. The blue-, green-, and red-colored parts of the diagram (the top 2, middle 2, and bottom 2 blocks, respectively) correspond to parts that pertain to factory process flow simulation, interfacing (cosimulation) between the factory and network simulation subsystems, and communication network simulation.

In TLFS, a factory system is modeled as a dataflow graph in which vertices (actors) represent distinct physical or computational components within the system, and edges represent the flow of information or physical entities (e.g., parts for products that are being manufactured). Examples of TLFS actors include actors for representing individual machines within a factory, individual rails that connect different machines, and machine controllers that send signals to machines to instruct the machines to perform specific functions (such as loading parts on rails when specific conditions are met). More detailed actor examples will be presented in Section V.

As illustrated in Figure 2, cosimulation is enabled by the TLFS coordination module (or just coordination module in

Bhattacharyya, Shuvra; Candell, Richard; Hany, Mohamed; Jeng, Ging; Li, Honglei; Liu, Yanzhou; Liu, Yongkang. "Model-based Cosimulation for Industrial Wireless Networks." Paper presented at 14th IEEE International Workshop on Factory Communication Systems, Imperia, Italy. June 12, 2018 - June 15, 2018.



Fig. 2. Cosimulation architecture based on TLFS.

abbreviated form), which transfers relevant simulation events between the factory and network simulation subsystems. There are two types of events that trigger event transfer through the coordination module: dataflow send events, which are generated in the factory simulation (dataflow simulator), and network receive events, which are generated in the CNS. These events are communicated between simulation subsystems using a dedicated event list, which is represented by the block in Figure 2 labeled "communication event list", together with the coordination module, which acts as a gateway to achieve proper synchronization of communication events. More details on the process of managing dataflow send events and network receive events are described in Section IV-C.

A. Coordination Module

Figure 3 shows a pseudocode sketch of the coordination module. Here, ntime represents the value of simulated time in the CNS (network time), and nresult is a simple data structure that is used to communicate selected status information from the CNS tool to the coordination module. In particular, nresult.result time gives the current value of network time, and nresult.receive_events provides the list of network receive events that have been generated in the CNS tool during the most recent call to the CNS tool. The function save_receive_events removes events from nresult.receive_events, and generates corresponding events in $\tau LIDE$ that trigger processing of the received data (see Figure 3). The dataflow simulator has access to the communication event list directly through special communication interface actors within the factory process flow model. Communication interface actors are discussed further in Section IV-C.

The next_lide_event_time function is a method of τ LIDE that returns the time of the earliest pending event within the event list of the dataflow (factory) simulation. The simulate_network function serves as a "wrapper" for

```
while ((ntime < time limit) and (!deadlock)) {
    simulate_network(next_lide_event_time(),
            nresult);
    if (!is_empty(nresult.receive_events)) {
       save_receive_events (nresult.receive_events) ;
    ntime = nresult.time;
    deadlock = nresult.deadlock;
    if (!deadlock) {
        simulate lide (ntime, lide result);
        transfer_comm_events(comm_event_list);
} /* while */
```

Fig. 3. A pseudocode sketch of the block in Figure 2 that is labeled TLFS coordination module.

whatever CNS tool is being used in the given application of TLFS. This function uses application programming interfaces of the CNS tool to simulate the FA system's communication network until a new receive event is generated or the next_lide_event_time is reached, whichever happens first. If (a) the return value from next_lide_event_time is infinite, which means that there are no pending dataflow events, and (b) there are no more pending communication events in NS-3, then simulate_network stops with network_result.deadlock = true. The simulation (while loop in Figure 3) stops when this deadlock condition is reached or when the predefined simulation time limit time_limit is reached.

B. Latency Modeling

The passage of time in the factory process flow model incorporates latencies that are associated with the actors in the model. These latencies are specified as part of the modeling process. Additionally, the physical placement of factory components (machines, controllers, etc.) with respect to the enclosing factory environment is annotated as attributes of actors that represent the components in the dataflow model. This placement information is passed to the CNS tool when the cosimulation is initialized.

Similarly, the CNS tool is responsible for simulating latencies associated with sending and receiving data across the communication network. In general, CNS tools take into account channel characteristics, network traffic conditions, and transmitter-receiver separation (distance) in determining these latencies. The interoperability of TLFS with arbitrary CNS tools allows designers to leverage the various communication latency modeling approaches of available tools.

Each dataflow actor A in a TLFS factory process flow model has an associated execution time estimation function, which is denoted by θ_A or by θ if A is understood from context. The arguments to the function include any parameters and state variables of A.

When a new firing of A becomes enabled at some simulated time t, TLFS calls the θ_A function to determine the amount of simulated time T_f that will be expended by the firing. Here, by an enabled firing, we mean a firing for which there is sufficient data on the input edges of the associated actor A, as defined by the consumption rate specification for the next mode of A.

Bhattacharyya, Shuvra; Candell, Richard; Hany, Mohamed; Jeng, Ging; Li, Honglei; Liu, Yanzhou; Liu, Yongkang. "Model-based Cosimulation for Industrial Wireless Networks." Paper presented at 14th IEEE International Workshop on Factory Communication Systems, Imperia, Italy. June 12, 2018 - June 15, 2018.

Upon determining the value T_f , TLFS schedules a *firing* completion event to be processed by the simulator at time $(t+T_f)$. The firing completion event triggers the execution of a τ LIDE function that carries out a single firing of A, which in turn updates the input and output edges (FIFOs) of A based on the token consumption and token production that occurs as part of the firing.

Just as $\tau LIDE$ is used to manage simulated time delays associated with actor firings (execution of factory subsystems), the CNS tool is responsible for managing the simulated delays associated with sending and receiving communication packets through the wireless communication network. The TLFS coordination module (see Figure 2) is responsible for synchronizing the advancement of time across the two cooperating simulators for factory dataflow and network communication, respectively.

C. Communication Interface Actors

For dataflow modeling in TLFS, we introduce a special type of actor called a communication interface actor, as mentioned in Section IV-A. Communication interface actors are used in factory process flow models to represent functionality for sending or receiving data across wireless communication channels. These actors are used to provide modular interfaces between the dataflow subsystem within a TLFS simulation model and the TLFS coordination module (see Figure 2), which is used to provide time synchronization and information transfer between the dataflow simulator and the CNS tool. Communication across these interfaces goes through the dataflow simulator and communication event list.

In TLFS, we use different types of communication interface actors for sending and receiving data across the wireless network. These are referred to, respectively, as send interface actors (SIAs) and receive interface actors (RIAs). When an SIA fires (executes) in the dataflow simulation, it injects one or more events into the communication event list. This list is used, as described in Section IV-A, by the TLFS coordination module to transfer events between the dataflow simulator and the CNS tool.

Similarly, events associated with the reception of data in the CNS tool (network receive events) trigger the firing of RIAs in the dataflow simulation. This triggering is enabled again by the coordination module, which injects an event into the τ LIDE event list corresponding to each network receive event that it detects. This process of injecting reception-related events into τLIDE is represented by the function call labeled save_receive_events in Figure 3. More specifically, an RIA firing completion event is scheduled in τ LIDE for each packet reception. The time of the event is the receive time as determined by the CNS tool.

Note that there need not be a one-to-one correspondence between the SIAs and RIAs in an TLFS model. The routing of packets between SIAs and RIAs is achieved through information in the packets (as they are assembled in the dataflow simulation), and the communication protocols that are being used (as they are simulated in the CNS tool).

V. FACTORY SYSTEM MODELING

In this section, we concretely demonstrate the process of modeling factory process flows in TLFS using the underlying dataflow capabilities. Once modeled within the dataflow framework of τ LIDE, the factory models can be integrated systematically with arbitrary CNS tools using the TLFS framework defined in Section IV.

A. Factory Process Flow Model

Figure 4 illustrates a simple factory system that we use to demonstrate the dataflow-based modeling process in TLFS. The factory processes parts that are generated by a Parts Generator subsystem. Each part is processed by a pipeline that consists of three distinct machines. A given machine could, for example, be responsible for adding some specific feature to the part. The third machine outputs the fully-processed parts, which are subsequently collected for storage or further processing by a subsystem called the Parts Sink. In this and related examples of this paper, we are not concerned with the details of what the Parts Sink does; we consider storage within the Parts Sink to represent the last stage of processing within the FA system (pipeline) that is being investigated.



Fig. 4. A simple factory system that we use to illustrate the dataflow-based modeling process in TLFS.

Adjacent machines and rails in Figure 4 are controlled by controller subsystems, which we refer to as machine controllers. Each of the lower three blocks in the figure represents a Dual-Rail, Single Machine (DRSM) controller, which is a machine controller that is designed to interface with two rails and a single machine. For example, the DRSM controller for Machine 1, Rail 1 and Rail 2 sends a command to Machine 1 to output the next part onto Rail 2, and similarly, it sends a command to Rail 1 when it is time for Machine 1 to take in the next part to process.

When a machine or rail completes a command that is sent from an associated machine controller, it sends an acknowledgment message back to that controller indicating the completion of the command operation. Additionally, state changes within the machines and rails may trigger notification messages to inform the corresponding machine controllers. The communication between machines and rails and their associated machine controllers is assumed to be carried out using wireless connections, which are shown as the dashed, bidirectional edges in Figure 4.

Figure 5 illustrates a TLFS-based dataflow model that can be used to model the factory system of Figure 4. This model is an actual test case for TLFS, and can be used by TLFS to simulate the functionality and performance of the factory system together with specific communication protocols that are used to implement the wireless networking, as well as

Bhattacharyya, Shuvra; Candell, Richard; Hany, Mohamed; Jeng, Ging; Li, Honglei; Liu, Yanzhou; Liu, Yongkang. "Model-based Cosimulation for Industrial Wireless Networks." Paper presented at 14th IEEE International Workshop on Factory Communication Systems, Imperia, Italy. June 12, 2018 - June 15, 2018.

specific models of the wireless channels within the factory environment.



Fig. 5. A dataflow model for the factory system of Figure 4.

In the dataflow graph of Figure 5, the actors M_1, M_2, M_3 and r_1, r_2, r_3 model the machines and rails, respectively, of the factory pipeline that is depicted in Figure 4. Similarly, the actors C_1, C_2, C_3 represent the three machine controllers that provide commands to different subsets of machines and rails. The actors G and K are used to model the Parts Generator and Parts Sink.

The rail r_1 is a special rail that incorporates storage for multiple parts that are generated by the Parts Generator. If this storage fills up (i.e., if parts are being generated faster than they are processed in the pipeline), then the Parts Generator stops generating new parts until there is space in r_1 to accommodate them. This functionality ensures that parts do not keep accumulating at the output of the Parts Generator if they cannot be processed by the machines fast enough.

The numbers next to or inside the actors and next to the edges in Figure 5 are used as unique labels (indices) to help identify the different graph components.

The actors in Figure 5 that are labeled $S_1, S_2, \ldots S_{12}$ are SIAs. As described in Section IV-C, SIAs and their counterparts called RIAs are special actors that model wireless communication interfaces for transmitting and receiving data across the FA network. Actors R_1, R_2, \ldots, R_{14} are RIAs. The RIAs and SIAs are actors that encapsulate the functionality for interfacing the dataflow-based FA process flow model to the event-based model that is used in the CNS tool. More details about SIAs and RIAs are discussed in Section IV-C.

B. Actor Modeling Example

Actors that model individual factory subsystems, such as machines, rails, and the Parts Generator, are specified in TLFS using CFDF semantics, which were introduced in Section II. In CFDF, a given actor can be viewed as a state machine where the current mode of the actor defines the present state, and state transitions occur between firings based on the determination of the actor's next mode as part of each firing. A graphical, finite state machine representation of a CFDF actor is therefore referred to as a mode transition graph. In a mode transition

graph, vertices represents actor modes for a given actor, and each edge (m_a, m_b) indicates that it is possible for the actor to determine its next mode as m_b when it is executing in mode m_a .

To illustrate actor modeling in TLFS, we discuss the modeling of machine controllers in our example factory process flow model of Figure 5. Figure 6 illustrates the interface (input and output ports) for each of the machine controller actors C_1, C_2 and C_3 in Figure 5. These actors are among those that have the most complex mode transition graphs in our factory system example, and help to concretely illustrate the modeling of factory subsystems in TLFS. Actors C_1 , C_2 and C_3 represent DRSM controllers.



Fig. 6. Interface of a DRSM controller actor.

A DRSM controller is associated with a specific machine M in the factory system, an input rail R_{in} of the machine, and an output rail R_{out} of the machine (or to some subsystem that has a similar input interface as a rail). The DRSM controller communicates through a wireless channel to monitor updates to the state of M, R_{in} , and R_{out} that are detected by sensors on these three subsystem. The DRSM controller also sends actuation commands to M, and R_{in} through the wireless channel.

In our factory model, the DRSM C_3 interfaces to the Parts Sink actor in place of an output rail. However, the input interface of the Parts Sink is modeled in a similar fashion as that of a rail. Thus, a DRSM is used to model the controller for Machine M_3 even though the output of this machine is connected to the Parts Sink rather than to a rail.

The input ports M_S , R_{inS} , R_{outS} , shown in Figure 6 represent ports for receiving monitoring messages from the relevant sensors of M, R_{in} , and R_{out} , respectively. Similarly, the output ports M_A and R_{inA} represent ports for sending actuation messages to M and R_{in} . In the notation associated with these ports, R, M, s, and a stand respectively for rail, machine, sensing, and actuation. Note that in this factory system example, the actuation associated with R_{out} (the transfer of parts from the rail to the next factory pipeline unit) is assumed to be performed by a different controller that is also associated with R_{out} .

The approach described here for DRSM controllers is just one possible way of modeling interfaces in TLFS between machine controllers and their associated rails and machines; many other interfacing models are possible given the flexibility of the underlying dataflow model of computation.

Bhattacharyya, Shuvra; Candell, Richard; Hany, Mohamed; Jeng, Ging; Li, Honglei; Liu, Yanzhou; Liu, Yongkang. "Model-based Cosimulation for Industrial Wireless Networks." Paper presented at 14th IEEE International Workshop on Factory Communication Systems, Imperia, Italy. June 12, 2018 - June 15, 2018.

C. Mode Transition Graph

In Section V-B, we discussed the mode transition graph as a useful representation of actor functionality. Figure 7 illustrates the mode transition graph for a DRSM controller actor.



Fig. 7. Mode transition graph for a DRSM controller actor.

As a DRSM controller actor D operates, it keeps track of (as part of its internal state) the status of the associated factory subsystems M, R_{in} and R_{out} through messages it receives through the M_S , R_{inS} , and R_{outS} ports of D, respectively. Note that dataflow actors for embedded applications are typically allowed to maintain state. Such state can be represented formally within a dataflow model as a self loop edge — that is, an edge whose source and sink actors are identical. To avoid clutter in diagrams of dataflow graphs, we often omit the self-loop edges.

The status information that a DRSM keeps track of includes the state of the machine M and the number of parts that reside at any given time on the rails R_{in} and R_{out} . This control information is used to help ensure the constraints that at any given time there should be at most one part in the machine, and that the number of parts on a rail at any given time should not exceed its capacity. The capacity of a rail (the maximum number of parts that can reside on it at a given time) is an actor parameter in the dataflow model. In our experiments, we use a capacity of 10 for rail r_1 in Figure 5, which connects to the output of the parts generator, and we use a capacity of 1 for each of the other rails.

The DRSM operates through sequences of message processing cycles, where each cycle involves a trajectory through multiple actor modes. Each cycle begins in the selectMessage mode, where the next input message to process is selected. This selection is made based on the unprocessed messages that are buffered at the actor inputs, and some priority scheme

for selecting messages if messages are available on multiple inputs. If no messages are available at the inputs, then the actor waits in the selectMessage mode until a message becomes available. The TLFS framework can be used to prototype and experiment with different priority schemes. In our experiments, we assumed the simple priority scheme in which messages are selected in decreasing priority order from the inputs R_{outS} , R_{inS} , and M_S , respectively.

The processRout, processMachine, and processRin modes shown in Figure 7 are used to perform initial processing of messages received from actor inputs R_{inS} , R_{outS} , and M_S , respectively. These actors use the newly received sensor data and the current state information saved within the DRSM to determine whether a new command message needs to be sent to R_{in} , R_{out} or M. If a new command is not required to be sent at this time, the DRSM completes the current message processing cycle by returning to the selectMessage mode, where the next cycle will subsequently begin.

Otherwise — if the DRSM determines in the processRout, processMachine, or processRin mode that a command message needs to be sent — then it transitions to Mode machineReady, machineNewPart, or machinePartDone depending on the type of message that needs to be sent. Mode machineReady represents the situation where a part has entered the machine, and the machine is waiting for a command to start performing its processing on the part. Similarly, Mode machineNewPart is entered when the machine is ready to receive a new part for processing from its input rail, and a part is available to be ingested from the rail. Mode machinePartDone represents the situation where the machine has finished its processing of a part that is still inside the machine, and there is a vacancy on the output rail so that the part can now be transferred to the rail.

All of the modes in Figure 7 are based on standard CFDF semantics (see Section II), except for the selectMessage mode, which is called a transition-only mode. This type of mode can be used to model nondeterminate behaviors in LIDE. In our case, nondeterminism arises because of the unpredictable order in which messages in general arrive at wireless communications interfaces of network nodes. Support for transition-only modes is a new feature that we have added to LIDE as part of developing TLFS.

In its transition-only mode, the machine controller actor does not consume any input tokens. Instead, it checks the populations of the input edges, and determines the next mode as a function of these populations. The machine controller actors in our factory process flow model are the only actors that have transition-only modes, and each of these actors has exactly one such mode. Thus, the overall factory system model is technically based mostly on CFDF - all actors are CFDF, except the three machine controllers, which are equipped with transition-only modes, as described above.

The use of transition-only modes allows factory system models in TLFS to incorporate non-determinism at the communication network level together with deterministic, CFDFbased modeling of signal and information processing that is

Bhattacharyya, Shuvra; Candell, Richard; Hany, Mohamed; Jeng, Ging; Li, Honglei; Liu, Yanzhou; Liu, Yongkang. "Model-based Cosimulation for Industrial Wireless Networks." Paper presented at 14th IEEE International Workshop on Factory Communication Systems, Imperia, Italy. June 12, 2018 - June 15, 2018.

TABLE I DATAFLOW TABLE FOR A DRSM CONTROLLER ACTOR.

Mode	R_{inS}	M_S	R_{outS}	M_A	R_{inA}
selectMessage	0	0	0	0	0
processRin	-1	0	0	0	0
processMachine	0	-1	0	0	0
processRout	0	0	-1	0	0
machineReady	0	0	0	0	1
machinePartDone	0	0	0	1	0
machineNewPart	0	0	0	1	0

embedded within individual factory subsystems. This multilayer modeling is achieved all within a unified dataflow framework along with model-based interfacing to CNS tools. Experimentation with complex signal and information processing functionality in TLFS — for example, by integrating machine learning subsystems into factory workflow models is a useful direction for future work.

D. Dataflow Table

In addition to the mode transition graph, another data structure that is useful for representing the behavior of CFDF actors is the dataflow table [4]. The dataflow table for a CFDF actor A is a matrix D whose rows are in one-to-one correspondence with the mode set $\mu(A)$ and whose columns are in one-to-one correspondence with the ports of A. for each port p of A and each mode $m \in \mu(A)$, the matrix element D[m, p] gives the net change in the token population on the edge connected to p that results from firing A in Mode m. If p is connected to a self-loop edge or if m does not produce or consume data on p during m, then D[m, p] = 0. Otherwise, D[m, p] gives the number of tokens produced on p during m if p is an output port of A, and D[m, p] gives the negative of the number of tokens consumed from p during m if p is an input port of A.

Table I shows the dataflow table for a DRSM controller actor. By definition, all of the entries for the selectMessage mode are zero in the table because this is a transition-only mode. As shown in the table, all of the other modes perform dataflow (consume or produce tokens) on at least one actor port. The dataflow table is useful, for example, as a data structure for model-based testing. Using the dataflow table, one can generate tests for validating that the implementation of an actor is consistent with the interface behavior defined by the actor model.

E. Summary

In this section, we have presented concretely, through an example, the modeling of a factory process workflow in TLFS. We have focused in the example on modeling the interfacing between different subsystems in a factory model. Of course, this is just one example of a factory model in TLFS, and many other kinds of process flow models can be developed and experimented with using the tool.

While models of the type demonstrated in this section are relatively abstract, they can be useful for fast design space exploration across large design spaces involving communication protocols, factory topologies, and distributed control policies. Such early stage design space exploration can be used to inform later stage simulation experiments that investigate specific combinations of protocols, topologies, and policies in more depth. Experimenting with TLFS for such detailed simulation experiments, which incorporate lower levels of process flow design abstraction, is a useful direction for future work.

VI. EXPERIMENTS

In this section, we present experiments to demonstrate the utility of TLFS. First, we use TLFS to study the effect of different wireless communication protocols on communication system performance for a given factory model. Then we demonstrate how communication performance changes as the complexity of the factory system increases, and also as the distances between network nodes in the factory increase. These experiments are conducted with NS-3 as the CNS tool.

The experiments presented in this section are representative of the kinds of studies that can be performed with TLFS. They are by no means intended to be comprehensive. Indeed, the flexible architecture of TLFS, which facilitates integration with arbitrary CNS tools, enables a wide variety of different kinds of trade-off studies, and investigations into the influence of parameters and other design decisions on factory system performance. Development of more extensive case studies in simulation and design using TLFS is an interesting direction for future work.

A. Simulation Parameters

Each factory model simulated in our experiments consists of one or more pipelines of the form illustrated in Figure 4 and Figure 5. Each factory model has two size-related parameters - the number of pipelines N_p , and the number of machines per pipeline N_m . Thus, the example shown in Figure 4 and Figure 5 corresponds to $N_p = 1$ and $N_m = 3$.

Table II summarizes the other key simulation parameters used in our experiments. A given data point in the experiments is derived by executing a simulation with the same settings N_s times, and averaging the results over the N_s executions. Each such simulation involves N_i generated parts for each Parts Generator actor in the factory dataflow graph. The simulation completes when all of the generated parts are fully processed in their respective pipelines. Since there is one Parts Generator actor per pipeline, this means that each simulation involves processing a total of $(N_p \times N_j)$ parts.

The values of t_m and t_r give, respectively, the estimated execution time values used in the simulation models for a machine to process a part, and for a rail r to move a part from one end of r to the other end. Similarly, t_i is the estimated time required to generate a new part after the previous part has been generated. The values of t_m , t_r , and t_i are used in the execution time estimation functions (θ s) for the relevant actors (see Section IV-B).

Bhattacharyya, Shuvra; Candell, Richard; Hany, Mohamed; Jeng, Ging; Li, Honglei; Liu, Yanzhou; Liu, Yongkang. "Model-based Cosimulation for Industrial Wireless Networks." Paper presented at 14th IEEE International Workshop on Factory Communication Systems, Imperia, Italy. June 12, 2018 - June 15, 2018.

TABLE II SIMULATION PARAMETERS.

Parts Generated Per Pipeline N _i	100		
Number of Simulation Iterations N_s	10		
Machine Processing Time t_m	10 sec		
Rail Transfer Time t_r	4 sec		
Part Generation Interval t _i	10 sec		
Channel Frequency	2.4 GHz		
Large Scale Path Loss Model	Log-distance		
Decay Exponent α	3		
Distance Reference d_0	1 m		
Loss at Reference L_0	46.6777 dB		

The parameters α , d_0 , and L_0 in Table II are related to the simulation of propagation path loss. In our simulations, we apply features in NS-3 for using the log-distance path loss model to estimate signal loss in communication channels. The log-distance model is often used to estimate path loss within buildings. In this model, the power loss at the receiver side when transmitting over a distance d is calculated by

$$L = L_0 + 10\alpha \log_{10}(\frac{d}{d_0}) + Z,$$
 (1)

where L_0 is the path loss at the reference distance, d_0 is the reference distance, α is the decay exponent, and Z is the lognormal shadowing.

B. Experiments with Different Protocols

We first use TLFS to study the average communication delay T_c for a fixed factory size, and the variation in T_c for different communication protocols - in particular, for different variants of IEEE 802.11. Using TLFS, we measure T_c as the average time difference between the time when a communication packet P is successfully received (through an RIA), and the time when P was transmitted (through an SIA). This average is taken over all packet communications within a given simulation.

In this experiment, we use a factory model with a single pipeline that contains 3 machines — that is, $N_p = 1, N_m = 3$.

Figure 8 shows a box plot representation of how T_c was found to vary across four different protocols - IEEE 802.11xx, for $xx \in \{ac, b, g, n\}$. Significant performance variation is shown between the best-performing protocol in this context (IEEE 802.11g) and the worst-performing one (IEEE 802.11b). One reason for the relatively low performance of IEEE 802.11b may be its low maximum data rate. Compared with other protocols, which can achieve 54 Mbps speed, the maximum speed of IEEE 802.11b is only 11 Mbps.

Overall, the results in Figure 8 show a clear advantage of IEEE 802.11g in terms of T_c for the factory model studied in this experiment.

C. Scalability Experiments

Next, we study how communication performance changes as we increase the factory size (N_p, N_m) and the distance between factory subsystems. Here we study five different factory dataflow graphs, denoted (a) through (e),



Fig. 8. A box plot representation of the variation in average communication delay across different communication protocols.

with sizes that are defined, respectively, as $(N_p, N_m) =$ (1,3), (1,4), (1,5), (4,4), (5,5). The total number of network nodes in the CNS tool modeling subsystem are, respectively, 11, 14, 17, 56, and 85. Each rail, machine, and machine controller corresponds to a distinct network node. Additionally, each Parts Generator and Parts Sink is modeled as a separate network node.

We define a *distance parameter* d associated with the simulations in this experiment. The units of this parameter are meters. For each factory pipeline, the spacing between adjacent network nodes is set to d. Additionally, for factory models that consist of multiple pipelines, the successive pipelines are spaced apart by distance d. Note that for the experiments reported in Section VI-B, we used a constant distance value of d = 10.

Figure 9 and Figure 10 show the variation in average communication delay and packet retransmission rate, respectively, for the five different factory sizes defined above, and for different settings of d for each factory size. In each of these two figures, each of the five plots corresponds to a distinct (N_p, N_m) pair, and each curve within a given plot corresponds to a distinct value of $d \in \{5, 10, 15, 20, 25\}$. The data is plotted for each of the four IEEE 802.11 variants discussed in Section VI-B. In Figure 10, the vertical axis represents the fraction of packet transmissions that have to be repeated due to errors in the original transmission. For example, a value of 0.5 means that 50% of the packets have to be retransmitted. In addition to increasing communication delays, packet retransmissions result in energy consumption overhead due to the increased operational load placed on the communication transceivers in the system.

The results in Figure 9 and Figure 10 show the general trends that one would expect of increasing communication delay and packet retransmission rate with increases in the distance parameter value d, and increases in the factory size. The results also provide insight into how different IEEE 802.11 protocol variants perform for the different factory size/distance combinations that are evaluated. The simulations carried out

Bhattacharyya, Shuvra; Candell, Richard; Hany, Mohamed; Jeng, Ging; Li, Honglei; Liu, Yanzhou; Liu, Yongkang. "Model-based Cosimulation for Industrial Wireless Networks." Paper presented at 14th IEEE International Workshop on Factory Communication Systems, Imperia, Italy. June 12, 2018 - June 15, 2018.



Fig. 9. Variation in average communication delay for different factory sizes, distance parameter settings, and IEEE 802.11 variants.



Fig. 10. Variation in packet retransmission rate for different factory sizes. distance parameter settings, and IEEE 802.11 variants.

using TLFS provide a quantitative assessment of all of these trends, and the associated factory system design trade-offs. The results also help to validate the capabilities of TLFS and demonstrate these capabilities with further concreteness.

VII. CONCLUSIONS

In this paper, we have introduced a simulation framework called TLFS (Tau Lide Factory Sim) to address the challenge of simulating complex factory automation systems that are integrated with wireless communication. TLFS incorporates novel techniques, based on dataflow concepts, for representing factory process flows, and for systematically integrating dataflow-based process flow simulations with discrete event simulations of communication network functionality. Useful directions for future work include automatically generating the corresponding communication network models from dataflow graph representations of networked factory process flows, and experimenting with more detailed and specialized factory process flows using TLFS.

DISCLAIMER

Certain commercial equipment, instruments, materials, software or systems are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

REFERENCES

- [1] A. A. K. S., K. Ovsthus, and L. M. Kristensen, "An industrial perspective on wireless sensor networks - a survey of requirements, protocols, and challenges," IEEE Communications Surveys & Tutorials, vol. 16, no. 3, pp. 1391–1412, 2014.
- [2] E. A. Lee and S. A. Seshia. Introduction to Embedded Systems. A Cyber-Physical Systems Approach, 2011, http://LeeSeshia.org, ISBN 978-0-557-70857-4.
- C. Shen, W. Plishker, H. Wu, and S. S. Bhattacharyya, "A lightweight [3] dataflow approach for design and implementation of SDR systems,' in Proceedings of the Wireless Innovation Conference and Product Exposition, Washington DC, USA, November 2010, pp. 640-645.
- [4] S. Lin, Y. Liu, K. Lee, L. Li, W. Plishker, and S. S. Bhattacharyya, "The DSPCAD framework for modeling and synthesis of signal processing systems," in Handbook of Hardware/Software Codesign, S. Ha and J. Teich, Eds. Springer, 2017, pp. 1-35.
- [5] ns-3 Tutorial, Release ns-3.25, ns-3 Project, 2016.
- [6] E. A. Lee and T. M. Parks, "Dataflow process networks," Proceedings of the IEEE, pp. 773-799, May 1995.
- [7] S. S. Bhattacharyya, E. Deprettere, R. Leupers, and J. Takala, Eds., Handbook of Signal Processing Systems, 2nd ed. Springer, 2013.
- [8] W. Plishker, N. Sane, M. Kiemb, and S. S. Bhattacharyya, "Heterogeneous design in functional DIF," in Transactions on High-Performance Embedded Architectures and Compilers IV, ser. Lecture Notes in Computer Science, P. Stenström, Ed. Springer Berlin / Heidelberg, 2011, vol. 6760, pp. 391-408. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-24568-8_20
- [9] G. Kahn, "The semantics of a simple language for parallel programming," in Proceedings of the IFIP Congress, 1974.
- [10] J. B. Dennis, "First version of a data flow procedure language," in Programming Symposium, ser. Lecture Notes in Computer Science, B. Robinet, Ed. Springer Berlin Heidelberg, 1974, vol. 19, pp. 362-376.
- [11] H. Neema et al., "Model-based integration platform for FMI cosimulation and heterogeneous simulations of cyber-physical systems," in Proceedings of the International Modelica Conference, 2014, pp. 235-245
- [12] M. Düngen, F. Hofmann, and H. Schulze, "Bit error rate simulation studies for PSSS with multi-user detection for industrial multipathfading environments," in IEEE International Workshop on Factory Communication Systems, 2017, pp. 1-6.
- [13] Y. Liu, R. Candell, K. Lee, and N. Moayeri, "A simulation framework for industrial wireless networks and process control systems," in IEEE World Conference on Factory Communication Systems, 2016, pp. 1–11.
- F. Bause, P. Buchholz, J. Kriege, and S. Vastag, "A simulation environ-ment for hierarchical process chains based on OMNeT++," *Simulation*, [14] vol. 86, no. 5-6, pp. 291-309, 2010.
- [15] S. Won, C. Shen, and S. S. Bhattacharyya, "NT-SIM: A co-simulator for networked signal processing applications," in Proceedings of the European Signal Processing Conference, Bucharest, Romania, August 2012, pp. 1094-1098.
- [16] A. Khan, S. M. Bilal, and M. Othman, "A performance comparison of open source network simulators for wireless networks," in Proceedings of the IEEE International Conference on Control System, Computing and Engineering, 2012, pp. 34-38.
- [17] M. Korkalainen, M. Sallinen, N. Kärkkäinen, and P. Tukeva, "Survey of wireless sensor networks simulation tools for demanding applications, in Proceedings of the International Conference on Networking and Services, 2009, pp. 102-106.

Bhattacharyya, Shuvra; Candell, Richard; Hany, Mohamed; Jeng, Ging; Li, Honglei; Liu, Yanzhou; Liu, Yongkang. "Model-based Cosimulation for Industrial Wireless Networks." Paper presented at 14th IEEE International Workshop on Factory Communication Systems, Imperia, Italy. June 12, 2018 - June 15, 2018.

INFLUENCE OF FIRE ON THE SHEAR CAPACITY OF STEEL-SHEATHED COLD-FORMED STEEL FRAMED SHEAR WALLS

Matthew S. Hoehler¹, Blanca Andres²

ABSTRACT

This paper discusses an experimental investigation of the influence of fire on the lateral load-carrying behaviour of steel-sheathed cold-formed steel framed shear walls. The completed Phase 1 tests showed an important example of structure-fire interaction and highlighted the relevance of the gypsum boards for both the thermal and mechanical behaviour of these systems. A second phase of the study is underway and is outlined in this paper with focus on the test program, specimen design and fire exposure development.

Keywords: cold-formed steel, fire, shear wall, steel-gypsum composite board

INTRODUCTION 1

Lightweight construction using cold-formed steel (CFS) studs represents roughly 20 % of the multistory nonresidential building market in the United States [1]. These buildings rely on their lateral force-resisting systems (LFRS) to resist horizontal loads; e.g. from wind or earthquakes. While information exists about the structural performance and fire resistance of cold-formed steel construction; see for example [2–5], there is limited knowledge about the performance of CFS-LFRS under combined hazards. This information is needed to inform: (i) fire compartmentation design when significant lateral deformation of a building is anticipated, (ii) post-fire assessment of a structure, and (iii) first responder decisions to enter a building when earthquake aftershocks are likely.

In 2016, a series of experiments (Phase 1) was performed at the National Fire Research Laboratory at the National Institute of Standards and Technology (NIST) to investigate the performance of earthquake-damaged steel-sheathed cold-formed steel shear walls under fire load [6]. A second phase of the project currently underway (Phase 2) extends the study to two additional levels of fire severity and two additional types of CFS-LFRS. In this paper, the authors summarize key findings from Phase 1 and present the test program and setup for Phase 2; with focus on the steel-gypsum composite board assemblies. The Phase 2 experiments are being conducted at the time of writing this paper.

PHASE 1 2

The Phase 1 experiments were performed in conjunction with the project Earthquake and Post-Earthquake Fire Performance of Mid-Rise Light-Gauge Cold-Formed Steel Framed Buildings conducted at the University of California, San Diego (UCSD), which investigated the earthquake and fire performance of a six-story, cold-formed steel framed test building [7]. The NIST tests were conducted immediately prior to the six-story building tests to experimentally determine the influence

Andres Valiente, Blanca; Hoehler, Matthew. "Influence of Fire on the Shear Capacity of Steel-Sheathed Cold-Formed Steel Framed Shear Walls." Paper presented at SiF18 – The 10th International Conference on Structures in Fire, Belfast, United Kingdom. June 6, 2018 - June 8, 2018.

¹Reserach Strucutral Engineer, National Institute of Standards and Technology, Gaithersburg, USA.

e-mail: matthew.hoehler@nist.gov

² PhD Student, Danish Institute of Fire and Security Technology, Hvidovre, Denmark.

e-mail: bav@dbi-net.dk

of the planned fires on the lateral load-resistance of the building to help inform the selection of the earthquake motion intensities used in the UCSD tests before and after the fires.

2.1 Test program and specimens

Table 1 shows the Phase 1 test matrix. Six identical 2.74 m by 3.66 m shear wall specimens were fabricated consisting of 150 mm wide CFS framing sheathed on one side with sheet steel adhered to 15.9 mm thick Type X (per [8]) gypsum board (Sure-Board 200^3) and on the opposite side with 15.9 mm thick Type X gypsum board (*Fig. 1*). Fabrication details are provided in [6]. The specimens were subjected sequentially to varied combinations of mechanical (shear) deformation and thermal (fire) loading. Specimen 1 (CFS01) was used to establish the monotonic 'pushover' load-displacement response of the wall system. Specimen 2 (CFS02) was loaded by symmetric-amplitude reverse-cyclic shear deformation to destruction to establish the cyclic load-displacement response. Specimen 3 and 4 (CFS03 and CFS04) were cycled to deformations just before and after the peak load was achieved, respectively, burned for 13 min and 20 s and then cycling was continued until destruction of the wall. For Specimen 5 (CFS05), an undamaged wall was exposed to fire for 13 min and 20 s and then cycled to destruction. Specimen 6 (CFS06) was tested similarly to Specimen 3, however, the burn duration was doubled.



Fig. 1. Phase 1 specimen geometry (units in meters unless noted)

Andres Valiente, Blanca; Hoehler, Matthew.

³ Certain commercial products are identified in this paper to specify the materials used and the procedures employed. In no case does such identification imply endorsement or recommendation by the National Institute of Standards and Technology, nor does it indicate that the products are necessarily the best available for the purpose.

2.2 Test setup and procedure

The test setup was informed by ASTM E2126-11 [9] but deviated slightly to accommodate a burn compartment on a rolling platform. Details are provided in [6].

The test specimens were loaded mechanically by holding the base of the wall fixed and applying a prescribed in-plane deformation to the top of the wall as shown in *Fig. 2a.* Out-of-plane movement of the wall was limited by four structural steel guide frames. Mechanical load was applied using a servo-hydraulically controlled actuator with a load capacity of 240 kN in tension and 365 kN in compression. Axial loading to the wall was limited to the self-weight of the specimen, actuator and top loading beam. ASTM E2126-11 Method C (CUREE Basic Loading Protocol) was used with a reference deformation based on the expected deformation at peak load under monotonic loading. This protocol is widely used in the United States to characterize the seismic horizontal load resistance of vertical elements intended to form the lateral force resisting system in buildings. The loading procedure involved displacement cycles grouped in phases at incrementally increasing displacement levels. The rate of displacement was selected to be 1.52 mm/s to minimize inertial influences.

The fire load was provided by a natural gas diffusion burner in a compartment (interior dimensions: 2.9 m high \times 3.5 m long \times 1.2 m deep; see *Fig. 2b*) designed to approximate a portion of the corridor in the six-story building tested at UCSD. The open side of the compartment mated with the test specimen on the steel-gypsum composite board side of the shear wall. The openings at the ends of the compartment were 1.7 m high by 1.2 m wide. The mass flow rate of the natural gas was controlled to approximate the predicted time-temperature curves for upper gas layer temperatures in the 2nd floor corridor of the six-story building at UCSD during the fire tests. This was achieved by rapidly increasing (rise time less than 60 s) the Heat Release Rate (HRR) of the burner to 1900 kW and holding it constant for 800 s (13 min 20 s); for specimen CFS06 the burn duration was doubled. While the compartment temperatures rose more rapidly than for the standard fire curve in American Society of Testing and Materials (ASTM) standard E119 [10] (or International Organization for Standardization (ISO) standard ISO 834 [11]), they were representative of conditions present in residential fires in modern buildings [12]. The burn duration for the UCSD tests that the NIST tests were intended to simulate, was limited to about 15 minutes by the fire authorities in San Diego.



Fig. 2. Photograph of test setups: a) Mechanical loading; b) Fire loading

2.3 Phase 1 results

The Phase 1 results are discussed in detail in [6,13] and data and videos can be downloaded at [14]. The uncertainties associated with all Phase 1 measurements are provided in [6]. The total expanded uncertainties (using a coverage factor of 2) associated with the forces and displacements reported in this paper are 1.6 kN and 2.3 mm, respectively.

Specimen 3 represents a typical loading case in Phase 1. The specimen was first cycled to a prescribed drift, subjected to fire load, and then, after the wall cooled, mechanical cycling was continued until the wall failed (defined as approximately 70 % reduction of post-peak load capacity). The resulting load-displacement curve is shown in *Fig. 3a*. The local maxima (load) for each step in the loading pattern are indicated for compression (circles) and tension (squares) excursions. The curves defined by these maxima show an abrupt reduction of force after the specimen was subjected to the fire.

The enveloping curves ('backbone') defined by the local maxima of applied force versus drift for the five cyclic tests are compared in Fig. 3b. Displacements (drifts) measured at the top of the specimens are converted to drift ratios by dividing by the specimen height (2.74 m). The portions of the curves indicated by dashed lines represent the mechanical response in the post-fire test. Test CFS02 represents the stiffness and capacity of the wall under ambient conditions. Test CFS05 represents the stiffness and capacity of the specimen after the steel-sheathed side has been subjected to the investigated fire load for 13 min 20 s. The reduction in peak load capacity was 35 % and the response was roughly symmetric for tension and compression cycles. The reduction in the peak load was accompanied by a shift in failure mode of the specimens from local buckling of the sheet steel (Fig. 4a) to global buckling of the sheet steel (Fig. 4b) for the unburned (CFS02) and burned (CFS05) walls, respectively. The fire severely damaged the gypsum on the burn side reducing the stiffness of the shear panels out-of-plane and creating a 15.9 mm standoff between the screw heads and the sheet steel; i.e. the thickness of the lost gypsum. This, in effect, transformed the specimen to a plain sheet steel shear wall with reduced constraint around the panel boundaries. Pre-damaging the specimen by reversed shear cycling to 1 % (CFS03) or 1.8 % (CFS04) drift ratio prior to the fire loading had no noticeable influence on the residual load bearing capacity of the wall. Doubling the burn time to 26 min 40 s (CFS06) caused additional reduction (approximately 15 %) of the post-fire lateral load bearing capacity due to the damage to the nonstructural gypsum board on the back side (cold side) of the wall during the longer burn; which was not present in the shorter tests.



Fig. 3. a) Lateral load versus drift for Specimen 3 (CFS03); b) Envelope force versus drift ratio curves for all cyclic tests (CFS02 to CFS06)



Fig. 4. Photograph of back of sheathed side of wall with the nonstructural gypsum removed after mechanical loading to failure: a) Unburned wall after cycling (local buckling); b) Cycling after burning (global buckling)

3 PHASE 2

The Phase 1 experiments showed a clear example of structure-fire interaction (the shift in failure modes for the relatively short-duration fire) and highlighted the relevance of the gypsum for both the thermal and mechanical behaviour of these systems. Moreover, it appeared that this behaviour was predictable. This motivated the Phase 2 tests which include two additional types of CFS-LFRS, two additional levels of fire exposure, and locates the fire on the non-egress side of the shear wall.

3.1 Test program

Table 2 provides the test matrix for Phase 2. In addition to retesting the sheet steel adhered to Type X gypsum board (Sure-Board 200), Oriented Strand Board (OSB), and strap braced walls will be investigated. As in Phase 1, the walls are subjected to sequences of mechanical and fire loading. The test setup and mechanical loading procedure are the same as in Phase 1, however, the fire loadings differ as discussed below.

Table 2 Test matrix for Phase 2

Wall	Specimen	Loading				
Туре	Name	Before Fire	Fire	After Fire		
Sure-Board	SB01	Cycle to failure	-	-		
	SB02	-	ASTM E119 (1-hour)	Cycle to failure		
	SB03	-	Severe Parametric	Cycle to failure		
	SB04	- Mild Parametric		Cycle to failure		
OSB	OSB01	Cycle to failure	-	-		
	OSB02	-	ASTM E119 (1-hour)	Cycle to failure		
	OSB03	-	Severe Parametric	Cycle to failure		
	OSB04	-	Mild Parametric	Cycle to failure		
	OSB05	Drift Level 3	Severe Parametric	Cycle to failure		
OSB06		Drift Level 2	Severe Parametric	Cycle to failure		
	OSB07	Drift Level 1	Severe Parametric	Cycle to failure		
Strap braced	S01	Cycle to failure	-	-		
	S02	-	ASTM E119 (1-hour)	Cycle to failure		
	S03	-	Severe Parametric	Cycle to failure		
	S04	-	Mild Parametric	Cycle to failure		
	S05	Drift Level 3	Severe Parametric	Cycle to failure		
	S06	Drift Level 2	Severe Parametric	Cycle to failure		
	S07	Drift Level 1	Severe Parametric	Cycle to failure		

Drift Level 1 = 0.5 %; Drift Level 2 = 1.0 %; Drift Level 3 = to be determined based on SB01, OSB01, and S01

3.2 Test specimen

The dimensioning of the walls is similar to Phase 1, however, the CFS framing design and tension holddowns are modified to be more representative of typical U.S. practice; the Phase 1 detailing focused on buildings with more than five storeys. The walls are designed using Allowable Stress Design (ASD) nominally following American Iron and Steel Institute (AISI) standards S400-15/S1-16 [15] and AISI S100-16 [16] assuming the walls are located on the interior of the building. The details for the gypsum and sheet-steel sheathed (Sure-Board 200) walls are shown in *Fig. 5*.

The walls are designed to achieve a 1-hour fire-resistance rating per ASTM E119 [10]. The design for fire-resistance of the steel-gypsum composite board walls is based on [17]. The walls use 15.9 mm thick Type X gypsum board with the joints taped and joints and fastener heads covered with one coat of joint compound on the fire side of the wall. The influence of insulation material in the wall cavity is not investigated.



Fig. 5. Phase 2 steel-gypsum composite board specimen geometry (units in meters unless noted)

3.3 Fire load

The target fire exposures are selected to represent various levels of fire severity. Three exposures are considered in Phase 2: (1) a 1-hour standard ASTM E119 fire curve (comparable to ISO 834), (2) a 'severe' fire exposure, and (3) a 'mild' fire exposure. The severe and mild fires represent realistic post-flashover compartment fire conditions with heating, fully-developed and decay phases. *Fig. 6* plots the target temperature-time curves along with a typical upper layer gas temperature measured during Phase 1.

The severity of the fire is defined in terms of exposure time and peak temperature. These values are informed by a statistical fit of data from compartment fire tests reported by Hunt [18]. A peak temperature of 1100 °C represents 95 % of the reported peak temperatures and 900 °C represents 50 % assuming a normal distribution. These values were selected as the maximum temperatures for the 'severe' and 'mild' fires, respectively. Likewise, assuming a normal distribution of the duration of the fire, 70 minutes and 50 minutes represent 70 % and 50 % of the reported data, respectively.



Fig. 6. Target temperature-time curves for Phase 2

The length of the plateau is calculated using the time-to-burnout for the enclosure fire (τ_b) per [18] and as given in Eq. (1).

$$\tau_b = \frac{\mathbf{E} \cdot \mathbf{A}_f}{90\mathbf{A}_0 \sqrt{H_0}} \text{ (minutes)} \tag{1}$$

where Eis the fuel (energy) load per unit floor area of the enclosure (MJ/m²),

is the floor area of the enclosure over which combustibles are present (m^2) , A_f

is the opening area (m²), and A_0

 H_0 is the opening height (m).

In multi-unit residential buildings, shear walls are commonly located along corridors adjacent to a kitchen. Assuming a kitchen compartment and taking the mean values of floor area and fuel load density reported by the National Research Council Canada [19] for multi-family dwellings (9.8 m² floor area with 805 MJ/m²), opening factors⁴ of 0.04 m^{1/2} and 0.09 m^{1/2} provide a time-to-burnout of 37 minutes and 16 minutes, respectively. These times defined the temperature plateaus for the 'severe' and 'mild' fires.

For comparison, the area under the target curve for the 'severe' fire represents a 20 % higher energy than ASTM E119 and the 'mild' fire corresponds to 40 % lower energy. The 'mild' fire is similar to the average upper gas layer time-temperature curves achieved in the Phase 1 tests.

SUMMARY 4

The results from Phase 1 indicate that a relatively short-duration (13 min 20 s) fire representative of modern furnished residential spaces acting directly on the shear panels caused a shift in the failure mode of the investigated steel-sheathed and gypsum shear walls under lateral loading from local to global bucking of the sheet steel with an accompanying reduction of the load capacity. The shear walls maintained a predictable, but reduced, lateral load capacity compared to ambient conditions. These observations prompted a second phase (Phase 2) of investigation described in this paper to extend the study to two additional shear wall systems and two additional fire loading exposures.

Paper presented at SiF18

⁴ Opening factor is defined as $\frac{A_0\sqrt{H_0}}{A_t}$, where A_0 is the opening area (m²), H_0 is the opening height (m), and A_t is the area total enclosure (m²).

Andres Valiente, Blanca; Hoehler, Matthew. "Influence of Fire on the Shear Capacity of Steel-Sheathed Cold-Formed Steel Framed Shear Walls." SiF18 – The 10th International Conference on Structures in Fire, Belfast, United Kingdom. June 6, 2018 - June 8, 2018

ACKNOWLEDGEMENT

We thank Sure-Board, CEMCO, Marino/WARE, Panel Systems, Inc., Larry Williams, Benjamin Schafer, Rob Madsen, and the National Fire Research Laboratory staff for their expertise and support.

REFERENCES

- [1] Steel Framing Industry Association, Market Share and Comparative Report: Cold-formed Steel Framing and Project Specification Trends, 2016.
- [2] M.A. Sultan, A Model for Predicting Heat Transfer through Noninsulated Unloaded Steel-Stud Gypsum Board Wall Assemblies Exposed to Fire, Fire Technol. 32 (1996) 239–257.
- [3] P. Keerthan, M. Mahendran, Numerical modelling of non-load-bearing light gauge cold-formed steel frame walls under fire conditions, J. Fire Sci. 30 (2012) 375–403. doi:10.1177/0734904112440688.
- [4] H. Takeda, A model to predict fire resistance of non-load bearing wood-stud walls, Fire Mater. 27 (2003) 19–39. doi:10.1002/fam.816.
- [5] X. Wang, E. Pantoli, T.C. Hutchinson, J.I. Restrepo, R.L. Wood, M.S. Hoehler, P. Grzesik, F.H. Sesma, Seismic Performance of Cold-Formed Steel Wall Systems in a Full-Scale Building, J. Struct. Eng. (United States). 141 (2015). doi:10.1061/(ASCE)ST.1943-541X.0001245.
- [6] M.S. Hoehler, C.M. Smith, Influence of fire on the lateral load capacity of steel-sheathed cold-formed steel shear walls - report of test, Gaithersburg, MD, 2016. doi:10.6028/NIST.IR.8160.
- [7] X. Wang, T.C. Hutchinson, G. Hegemier, S. Gunisetty, P. Kamath, B. Meacham, Earthquake and fire performance of a mid-rise cold-formed steel framed building – test program and test results: rapid release (preliminary) report (SSRP-2016/07), San Diego, CA, 2016.
- [8] ASTM C1396-17 Standard Specification for Gypsum Board, ASTM International, West Conshohocken, PA, 2015. doi:10.1520/C1396.
- [9] ASTM E2126-11 Standard Test Methods for Cyclic (Reversed) Load Test for Shear Resistance of Vertical Elements of the Lateral Force Resisting Systems for Buildings, ASTM International, West Conshohocken, PA, 2011. doi:10.1520/E2126-11.
- [10] ASTM E119-16a Standard Test Methods for Fire Tests of Building Construction and Materials, ASTM International, West Conshohocken, PA, 2016. doi:10.1520/E0119.
- [11] ISO 834-1 Fire Resistance Tests Elements of Building Construction Part 1: General Requirements, International Organization for Standardization, Genève, Switzerland, 1999.
- [12] S. Kerber, Analysis of Changing Residential Fire Dynamics and Its Implications on Firefighter Operational Timeframes, Fire Technol. 48 (2012) 865–891. doi:10.1007/s10694-011-0249-2.
- [13] M.S. Hoehler, C.M. Smith, T.C. Hutchinson, X. Wang, B.J. Meacham, P. Kamath, Behavior of steelsheathed shear walls subjected to seismic and fire loads, Fire Saf. J. 91 (2017) 524–531. doi:10.1016/j.firesaf.2017.03.021.
- [14] M.S. Hoehler, C.M. Smith, Dataset from Influence of Fire on the Lateral Resistance of Cold-Formed Steel Shear Walls - Phase I, (2018). doi:https://doi.org/10.18434/T4/1422515.
- [15] AISI S400-15 w/S1-16, North American Standard for Seismic Design of Cold-Formed Steel Structural Systems (with Supplement 1), American Iron and Steel Institute (AISI), Washington, DC, 2016.
- [16] AISI S100-16 North American Specification for the Design of Cold-Formed Steel Structural Members, American Iron and Steel Institute (AISI), Washington, DC, 2016.
- [17] IAPMO-ER-1261 Sure-Board Series 200, 200W, and 200B Strucutral Panels Installed on Cold-Formed Steel or Wood Framed Shear Walls, International Association of Plumbing and Mechanical Officials (IAPMO), 2018.
- [18] S.P. Hunt, J. Cutonilli, M. Hurley, Evaluation of Enclosure Temperature Emperical Models, Society of Fire Protection Engineers (SFPE), 2010.
- [19] A.C. Bwalya, G.D. Lougheed, A. Kashef, H.H. Saber, Survey Results of Combustible Contents and Floor Areas in Multi-Family Dwellings, National Research Council Canada, 2008.

FIRE SAFETY CHALLENGES OF TALL WOOD BUILDINGS: LARGE-SCALE CROSS LAMINATED TIMBER COMPARTMENT FIRE TESTS

Matthew S. Hoehler¹, Joseph Su², Pier-Simon Lafrance³, Matthew F. Bundy⁴, Amanda Kimball⁵, Daniel Brandon⁶, Birgit Östman⁷

ABSTRACT

This study investigates the contribution of cross laminated timber (CLT) building elements to compartment fires. Six compartments (9.1 m long \times 4.6 m wide \times 2.7 m high) were constructed using 175 mm thick 5-ply CLT structural panels and fire tested using residential contents and furnishings to provide a fuel load density of 550 MJ/m². The results show that gypsum board can delay or prevent the involvement of the CLT in the fire, and that the ventilation conditions and exposed surface area of the CLT play a decisive role in the outcome of the test. The results highlight the need to use heat-resistant adhesives in cross laminated timber to minimize delamination.

Keywords: cross laminated timber, compartment fires, fire tests, delamination, ventilation

1 INTRODUCTION

Recent architectural trends include the design and construction of increasingly tall buildings with structural components comprised of mass timber including cross laminated timber (CLT). These buildings are cited for their advantages for sustainability resulting from the use of wood as a renewable construction material.

This research aimed to quantify the contribution of CLT building elements to compartment fires, and to characterize the fire protection of the CLT using gypsum board for delaying or preventing its involvement in the fire. Six large-scale CLT compartment fire tests were planned in consultation with the Project Technical Panel [1–3] and in consideration of gaps identified by a literature review [4]. Modelling was also conducted to support the choice of test parameters [5].

¹Reserach Strucutral Engineer, National Institute of Standards and Technology, Gaithersburg, USA.

e-mail: matthew.hoehler@nist.gov

² Principal Research Officer, National Research Council of Canada, Ottawa, Canada.

e-mail: joseph.su@nrc-cnrc.gc.ca

³ Technical Officer, National Research Council of Canada, Ottawa, Canada.

e-mail: pier-simon.lafrance@nrc-cnrc.gc.ca

⁴ Director National Fire Research Laboratory, National Institute of Standards and Technology, Gaithersburg, USA.

e-mail: matthew.bundy@nist.gov

⁵ Research Director, Fire Protection Research Foundation, Quincy, MA, USA.

e-mail: <u>akimball@nfpa.org</u> ⁶Researcher, Research Institutes of Sweden, Borås, Sweden.

e-mail: daniel.brandon@ri.se

⁷ Senior Advisor, Linneaus University, Växjö, Sweden.

e-mail: birgit.ostman@lnu.se

North American codes require that all tall buildings be fully sprinklered in accordance with National Fire Protection Association (NFPA) standard NFPA 13 [6] and that fire services are required to respond to fire incidents in accordance with NFPA 1710 and 1720 [7,8]. However, to achieve the project objective to quantify the contribution of CLT building elements, the CLT compartment fire tests were conducted without sprinklers and without firefighting intervention.

2 TEST SETUP AND PROCEDURE

2.1 Test structure

Six compartments (9.1 m long \times 4.6 m wide \times 2.7 m high) were constructed using 175 mm thick 5-ply CLT structural panels manufactured using 2×4 spruce-pine-fir lumber glued with a polyurethane adhesive for all wall and ceiling assemblies. All CLT panels conformed to American National Standard ANSI/APA PRG-320 [9]. The ceiling assembly spanned in the 4.6 m direction parallel to walls W2 and W4 (*Fig. 1a*). The wall panels were connected using lap joints and the adjacent walls were connected using butt joints with screws. The ceiling panels were connected to the walls using butt joints with screws (*Fig. 1b*). The inside of the compartments was fully or partially lined using multiple layers of 15.9 mm thick Type X gypsum board.

The compartment had a rough opening in Wall W2 (front) of 1.8 m wide \times 2.0 m high in four tests and 3.6 m wide \times 2.0 m high in two tests. The ventilation factor was 0.03 m^{1/2} with the 1.8 m x 2.0 m opening, and 0.06 m^{1/2} with the 3.6 m \times 2.0 m opening, respectively. Additionally, two small openings of 150 mm diameter were created in wall W4 at 0.3 m and 1.8 m heights to provide an equivalent leakage of 0.035 m² without actually having a door located in the rear wall. Additional details and drawings for the compartments can be found in the final report [10].



Fig. 1. a) Plan view schematic of test compartments; b) Photograph of test compartments (bare structure)

2.2 Test program

Table 1 provides the test matrix. Test 1-1 served as a baseline for the tests with the 1.8 m \times 2.0 m opening. Three layers of the gypsum board were used to cover the CLT walls and ceiling in Test 1-1 so that the CLT would neither contribute to the fire nor develop char during the test. Test 1-2 served as a baseline for the tests with the 3.6 m \times 2.0 m opening. An exposed wall (W1) was tested in both ventilation configurations (Test 1-3 and Test 1-5). Test 1-4 and Test 1-6 involved an

exposed ceiling and a combination of exposed ceiling and wall, respectively, with the 1.8 m \times 2.0 m opening. In all tests, a structural load of 0.95 kN/m², which is one half of the design live load for a residential occupancy, was superimposed on top of the ceiling using eight cylindrical water tanks.

Rough	Compartment Surface						СІТ
Opening in Wall W2	W1 9.1m x 2.7m	W2 4.6m x 2.7m	W3 9.1m x 2.7m	W4 4.6m x 2.7m	Ceiling 9.1m x 4.6m	Test	Compartment
1.8m wide x 2.0m high	3GB	3GB	3GB	3GB	3GB	1-1	1
	3GB	3GB	3GB	3GB	exposed	1-4*	1*
	exposed	3GB	3GB	3GB	3GB	1-5	4
	exposed	3GB	3GB	3GB	exposed	1-6	3
3.6m wide x 2.0m high	2GB	2GB	2GB	2GB	2GB	1-2	2
	exposed	2GB	2GB	2GB	3GB	1-3*	2*

Table 1. Test matrix

GB: 15.9 mm (5/8 in.) thick Type X gypsum board; 2GB: 2 layers of GB; 3GB: 3 layers of GB * Reused CLT structure.

2.3 Ignition scenario and fuel load

Since the primary objective of the experiments was to evaluate the effect of the CLT structural elements on the fire growth and fire dynamics, the contents fire load and ignition scenario were designed to produce a medium fire growth rate. The movable fuel⁸ represented residential contents in a studio apartment with sleeping, living and kitchen areas (*Fig. 2a*) and had target a fire load density of 550 MJ/m². The mass, dimensions and position of all combustible material were measured prior to placement in the compartment. These measurements along with assumed calorific values for each material were used to create a 3D model of the fuel energy distribution in the compartment (*Fig. 2b*). The highest fuel density was in the kitchenette at the rear of the compartment, however, the rest of the fuel density was roughly evenly distributed throughout the compartment.

The ignition source was a natural gas burner (200 mm \times 400 mm) placed in the corner adjacent to walls W1 and W4 partially under a sideboard. A heat release rate of 50 kW (similar to a burning waste basket) from the burner was maintained until the total measured heat release rate from the compartment exceeded 1000 kW (due to ignited contents), at which time the burner was shut off.



Fig. 2. a) Photograph of movable fuel as installed; b) Model of fuel energy distribution in the compartment

Brandon, Daniel; Bundy, Matthew; Hoehler, Matthew; Kimball, Amanda; Lafrance, Pier-Simon; Ostman, Birgit; Su, Joseph. "Fire Safety Challenges of Tall Wood Buildings: Large-Scale Cross Laminated Timber Compartment Fire Tests." Paper presented at SiF18 – The 10th International Conference on Structures in Fire, Belfast, United Kingdom. June 6, 2018 - June 8, 2018.

3

⁸ Movable fuel refers to all combustible material inside of the compartment (including the laminated flooring) that is not a CLT structural element.

2.4 Measurement systems and instrumentation

Data was collected from multiple systems during the experiments. These systems can be grouped into four categories related to: (1) the calorimeter – dedicated to measuring fuel flow to the small ignition burner and performing oxygen depletion calorimetry, (2) gas sampling in the compartment, (3) the constituent measurements for the test specimens, and (4) imaging and video. 185 specimenrelated data channels were recorded. A detailed description of the instrumentation and the estimated uncertainty of the measurements are provided in the final report [10].

3 RESULTS

A large amount of technical data was produced, including measurements of the heat release rate, temperatures inside and outside the compartment, as well as through the encapsulation material and structural assemblies, simulated sprinkler response time, gas species concentrations (O_2 , CO_2 , CO_3 , pressure and flow conditions, interior and exterior heat fluxes, smoke density, and char depth. In this paper, we limit our discussion to the fire development, select temperature measurements, and comparisons of heat release rates. All of the measured data, as well as videos can be downloaded at [11].

Although the tests were conducted without sprinklers, simulated thermal elements (STEs) were placed at typical sprinkler heights to replicate the time temperature response of a quick-response sprinkler. The time from ignition for each STE to reach the temperature of 68 $^{\circ}$ C (a typical temperature rating for residential sprinklers) was always less than 5.8 minutes. At these times, the fires were still limited to the first item ignited. Had sprinklers been installed, flashover would not have occurred with effective sprinkler operations.

3.1 Fire development for the baseline ventilation conditions

Fig. 3 shows the typical pre-flashover condition of the compartment and a subsequent view of the compartment when most of the movable fuel has been consumed by the fire. In this example (Test 1-2), all CLT surfaces on the walls and ceiling were protected using two layers of gypsum board and the opening was 3.6 m \times 2.0 m (wide opening). Test 1-2 served as a baseline to define the contribution of the movable fire loads and to quantify CLT contribution to the fire in a subsequent test with the same ventilation conditions (Test 1-3).

The test started with the ignition of a sideboard in the room corner and progressed through the dining area. The smoke layer descended to the mid height in the compartment and stayed at this height until flashover. After flashover, an intense burning of room contents lasted for approximately 24 min. The fire started to decay at 37 min and the fire plume ceased to issue from the opening at 40 min. The test lasted for 104 min and was terminated after the heat release rate fell below 500 kW. A hose was used to lightly mist the debris on the floor with water to terminate the test.

The fire development described above was similar for both tests in which the CLT was fully protected by the gypsum and did not contribute to the heat release rate; i.e., Test 1-1 and Test 1-2. *Fig. 4* shows comparisons of the heat release rates during these tests. Test 1-2 with the larger ventilation opening had a higher heat release rate but earlier fire decay than Test 1-1. The two tests released virtually the same amount of total heat of 18 000 MJ, which is equivalent to the total energy of the movable fuel with 80 % combustion efficiency. However, the heat exposure conditions inside the compartment were more severe in Test 1-1 than in Test 1-2, because more intense burning occur outside the fire compartment in Test 1-2 [10]. The fire development for Test 1-3 to Test 1-6, which were affected to varying degrees by contributions from the burning CLT, is illustrated by the heat release rate comparisons in Section 3.3.



a) at 7 min 49 s after ignition

b) at 67 min

Fig. 3. Photographs of the CLT compartment during Test 1-2 (wide opening)



Fig. 4. Influence of opening size on the measured heat release rate for the same movable fuel load of 550 MJ/m² and no participation of CLT: a) Test 1-1 with narrow opening; b) Test 1-2 with wide opening

3.2 Encapsulation material and structural assembly temperatures

In this section we show temperature time-histories for a wall assembly that successfully prevented participation of the CLT in the fire for the compartment with a wide opening, as well as an unprotected wall assembly under the same compartment ventilation conditions.

Fig. 5 shows the temperatures measured at the interfaces between the adjacent gypsum board layers, and between the CLT interior surface and the gypsum board base layer, as well as inside the CLT panels at various depths for Wall W1 in Test 1-2 (wide opening) with two layers of gypsum. The heat transfer through the gypsum board followed the typical three-stage pattern as indicated by the temperature profiles at the interfaces: an initial temperature rise up to 100 °C, a period of gypsum calcinations at the constant temperature of 100 °C, then temperature increasing again after the calcination. Although the temperatures at the interface between the face and base layers of gypsum board stayed intact until the end of the test (*Fig. 3b*). The maximum temperature measured at the CLT interface with gypsum board was 263 °C on wall W1 at the 1.8 m height; at other measurement locations, the CLT wall interface temperatures were 100 °C to 170 °C. The CLT wall panels were at the ambient temperature on the exterior surface. For Test 1-1, which provided a baseline for the compartments with a narrow opening, three layers gypsum board were necessary to successfully protect the CLT structure, preventing the ignition and involvement of CLT structural elements in the fire.



Fig. 5. Temperatures in a wall assembly with two layers of 15.9 mm thick Type X gypsum protection in Test 1-2 (wide opening): a) Details of embedded thermocouples in Wall W1 with the drilled holes in a circular pattern and sealed from the outside; b) Temperatures in Wall W1 at various depths from the fire exposed surface (1 inch = 25.4 mm)

Fig. 6 shows temperatures measured on the surface of and inside the CLT panels for Wall W1 in Test 1-3 (wide opening) which had no gypsum protection. The temperature profiles show that before the flashover, the upper portion of the exposed W1 wall was already ignited. Based on the timing when the embedded thermocouples in Wall W1 (mid-length at 1.8 m height) measured 300 °C, the char front reached 20 mm, 35 mm, 50 mm and 65 mm deep at 32 min, 83 min, 172 min, and 213 min, respectively. The embedded thermocouple in 35 mm depth of the W1 wall shows a sharp temperature rise at around 80 min. The starting point of this temperature rise, 220 °C, is below the wood charring temperature. The end point of this temperature rise, 500 °C, is the prevailing compartment temperature. Given that 35 mm inside CLT corresponds with a glue line and that the char front should be at around 300 °C, this temperature rise indicates that the delamination of the first ply of the CLT occurred prior to the char front reaching the first glue line. The embedded thermocouple in 65 mm depth of the CLT prior to the char front reaching the delamination of the second ply of the CLT prior to the char front reaching the second glue line. At the end of the test (242 min), the temperatures at the 90 mm, 115 mm and 140 mm depths in the CLT panel were 132 °C, 80 °C, and 40 °C, respectively.



Fig. 6. Temperatures in a wall assembly without gypsum protection in Test 1-3 (wide opening): a) Details of embedded thermocouples in Wall W1 with the drilled holes in a circular pattern and sealed from the outside; b) Temperatures in Wall W1 at various depths from the fire exposed surface (1 inch = 25.4 mm)

3.3 Heat release rate

The influence of exposing CLT under the investigated fuel load density and ventilation conditions is illustrated by comparing the measured heat release rates.

Fig. 7a compares the baseline case for the wide opening and no CLT participation (Test 1-2) with the case where one wall of the compartment is exposed (Test 1-3). With the exposed W1 wall in Test 1-3, flashover occurred 3 min earlier and the heat release rate was approximately 2 MW higher than the baseline (Test 1-2) in the growth and fully developed fire stages. The CLT contribution is most prominent, however, during the decay phase of the fire when flaming in the compartment reoccurred as the first and second plys in Wall W1 delaminated. Based on the volume (depth × area) of the CLT charred, which was measured after the test, the CLT panels were estimated to contribute approximately 1100 kg of timber to the fire (mainly by the W1 wall panels), which translated to 540 MJ/m² fuel load density to the floor area in addition to the movable fuel load. This brought the effective fuel load density to 1090 MJ/m² in Test 1-3.

Fig. 7b compares the baseline case for the narrow opening and no CLT participation (Test 1-1) with various combinations of ceiling and wall CLT exposure (Test 1-4 to Test 1-6). In the three investigated cases (one wall exposed, ceiling exposed, and one wall and ceiling exposed), exposure of the CLT with the narrow opening resulted in a decreased time to flashover (3.4 min to 5.1 min sooner) and ultimately to runaway fire growth that required manual fire suppression. The exposed CLT panels translated to more fuel loads in addition to the movable fuel load (room contents) in the compartment. Due to the intensity and duration of the thermal exposure in the compartment, the CLT surfaces protected by three layers of 15.9-mm thick Type X gypsum also contributed to the fire in the later part of the tests. The estimated contributions of the CLT to the total fuel load for these tests depended on the time at which the fire was suppressed and can be found in [10].

Comparing the heat release rates for Tests 1-3 (*Fig. 7a*) and Test 1-5 (*Fig. 7b*) in which both tests had one exposed CLT wall, illustrates the importance of the ventilation conditions on the outcome of the test. In Test 1-3 with a wide opening (ventilation factor = $0.06 \text{ m}^{1/2}$), the delamination in the exposed wall does not prevent the fire from continuing to decay, whereas in Test 1-5 with a narrow opening (ventilation factor = $0.03 \text{ m}^{1/2}$), the delamination of the plys led to runaway growth of the fire.



Fig. 7. Comparison of CLT compartment heat release rates: a) compartments with wide openings. b) compartments with narrow openings

Brandon, Daniel; Bundy, Matthew; Hoehler, Matthew; Kimball, Amanda; Lafrance, Pier-Simon; Ostman, Birgit; Su, Joseph. "Fire Safety Challenges of Tall Wood Buildings: Large-Scale Cross Laminated Timber Compartment Fire Tests." Paper presented at SiF18 – The 10th International Conference on Structures in Fire, Belfast, United Kingdom. June 6, 2018 - June 8, 2018.

7

4 CONCLUSIONS

The following conclusions are made based on the results:

- Ventilation conditions had significant impacts on the fire development in the compartments;
- The larger opening (ventilation factor = $0.06 \text{ m}^{\frac{1}{2}}$) increased the peak heat release rate, but reduced the fire challenges to the CLT compartment structure inside;
- The CLT contribution to the fire increased with increasing exposed surface area of the CLT; and,
- There is a need to use better heat-resistant adhesives in CLT for exposed CLT applications to minimize delamination.

ACKNOWLEDGMENT

The authors gratefully acknowledge the project management provided by the Fire Protection Research Foundation and the technical oversight of its Project Technical Panel. Financial support from the Fire Protection Research Foundation, Property Insurance Research Group, U.S. Department of Agriculture, U.S. Forest Service, American Wood Council, and National Research Council Canada is gratefully acknowledged. Donations of STE devices from FM Global and gypsum board from USG Corporation are also gratefully acknowledged.

REFERENCES

- J. Su, G. Lougheed, Fire Safety Challenges of Tall Wood Buildings Phase 2, Test Options for Task 2 – A Discussion Paper, Ottawa, Ontario, 2016.
- [2] J. Su, G. Lougheed, Fire Safety Challenges of Tall Wood Buildings Phase 2, Test Options for Task 2 Second Discussion Paper, Ottawa, Ontario, 2016.
- [3] J. Su, G. Lougheed, Test Plan for Fire Safety Challenges of Tall Wood Buildings Phase 2, Ottawa, Ontario, 2016.
- [4] D. Brandon, B. Östman, FPRF Project Fire Safety Challenges of Tall Wood Buildings Phase 2, Task 1 – Literature review: The contribution of CLT to compartment fires, 2015.
- [5] D. Brandon, B. Östman, FPRF Project Fire Safety Challenges of Tall Wood Buildings Phase 2, Task 2 – Test plan, modeling: The contribution of CLT to compartment fires, 2016.
- [6] National Fire Protection Association (NFPA), NFPA 13. Standard for the Installation of Sprinkler Systems, Quincy, MA, 2016.
- [7] National Fire Protection Association (NFPA), NFPA 1710. Standard for the Organization and Deployment of Fire Suppression Operations, Emergency Medical Operations, and Special Operations to the Public by Career Fire Departments, Quincy, MA, 2016.
- [8] National Fire Protection Association (NFPA), NFPA 1720. Standard for the Organization and Deployment of Fire Suppression Operations, Emergency Medical Operations and Special Operations to the Public by Volunteer Fire Departments, Quincy, MA, 2014.
- [9] ANSI/APA PRG 320-2012. Standard for Performance-Rated Cross-Laminated Timber, 2012.
- [10] J. Su, P. Lafrance, M. Hoehler, M. Bundy, Fire Safety Challenges of Tall Wood Buildings Phase 2: Task 2 & 3 – Cross Laminated Timber Compartment Fire Tests, Fire Protection Research Foundation, Quincy, MA, 2018.
- [11] M.S. Hoehler, M.F. Bundy, J. Su, Dataset from Fire Safety Challenges of Tall Wood Buildings Phase 2: Task 3 - Cross Laminated Timber Compartment Fire Tests, (2018). doi:https://doi.org/10.18434/T4/1422512.

HYBRID FIRE TESTING OF A SINGLE DEGREE-OF-FREEDOM LINEAR SYSTEM

Ana Sauca¹, Chao Zhang², Artur Chernovsky³, Mina Seif⁴

ABSTRACT

As the structural engineering industry transitions towards performance based design methods, a better understanding of the performance of structures as full systems, especially under extreme loading conditions like fire, becomes a must. Full scale testing will provide such information, however, at an unrealistic cost. Hybrid Fire Testing (HFT) is inspired from sub-structuring method, which provides the insight through testing individual members exposed to elevated temperatures, while simultaneously accounting for the effect of the surrounding structure, which is numerically modelled aside. The tested member and the numerically represented surrounding must communicate during the entire test. This communication framework is a key component for a successful HFT. However, there is lack of availability of generic frameworks that enable the use of different software and hardware configurations. This paper presents the development of such generic communication framework. First, a validation of the communication framework was done in a virtual environment, where both the tested member and the surrounding are numerically modelled separately. Next, a hybrid fire test was performed on a single degree-of-freedom linear system.

Keywords: Hybrid fire testing, fire tests, substructures, communication framework

INTRODUCTION 1

There is a lack of understanding of how structures, as whole systems, perform under realistic fires. A better understanding of the problem will support the development of performance-based standards and tools that explicitly consider realistic fire effects for both the design of new buildings and the assessment and retrofit of existing ones. The Hybrid Fire Testing (HFT) method, through sub-structuring, assumes testing individual members/components (referred to as physical substructure (PS)), while simultaneously accounting for the action of the surrounding structure (referred to as numerical substructure (NS)) which is computationally modelled aside. Thus, HFT allows, through testing individual members, the prediction of the global behaviour of whole structural systems. To date, only few HFT performed on full size structures are found in the literature [1-4]. Small scale HFT are presented in [5,6] and virtual hybrid fire tests (VHFT) (the PS and NS are represented numerically) are presented in [7,8] to validate new methods. These studies have been performed considering standard fire curves (e.g., [9]), while more researchers are currently focused on studying the behaviour of structures exposed to realistic fire scenarios [10-11].

¹Guest Researcher, National Institute of Standards and Technology NIST, Gaithersburg, MD. e-mail: ana.sauca@nist.gov

² Guest Researcher, National Institute of Standards and Technology NIST, Gaithersburg, MD...

e-mail: chao.zhang@nist.gov

³Research Electric Engineer, National Institute of Standards and Technology NIST, Gaithersburg, MD.

e-mail: artur.chernovsky@nist.gov

⁴Research Structural Engineer, National Institute of Standards and Technology NIST, Gaithersburg, MD.

e-mail: mina.seif@nist.gov

Chernovsky, Artur; Sauca, Ana; Seif, Mina; Zhang, Chao. "Hybrid Fire Testing of a Single Degree-of-Freedom Linear System." Paper presented at SiF18 - The 10th International Conference on Structures in Fire, Belfast, United Kingdom. June 6, 2018 - June 8, 2018.

A recent study performed on a prototype building exposed to two simulated realistic fire scenarios [12] showed how much the simulation of individual members could be different than the simulation of the complete structural system in terms of behaviour. Analysis of individual members simulated individual member test behaviour while analysis of the full structure simulated the full-scale testing of the structure as a system. The study highlighted that in some cases, individual members survived the fire, while failure occurred when the complete structural system was modelled. Thus, difference between the individual member testing and full-scale testing can be crucial in terms of safety. HFT is a potential solution that insures the proper accounting of the surrounding system on the individual member behaviour.

To perform a successful HFT, several key components are crucial [13], one of them is the framework for the experimental setup and control. Most testing facilities develop their own framework, while open-source frameworks such as OpenFresco [14] and UI-Simcor [15] have been successfully used in the seismic field. To perform a HFT using frameworks as OpenFresco or UI-SimCor, some modifications need to be implemented, making them more generic, thus allowing them to integrate with certain finite element software that have the capability to solve structures loaded with both mechanical and fire loads (e.g., SAFIR [16]). The main objective of this paper was to validate the HFT concept through developing a generic communication framework and using it to perform a HFT on a single degree-of-freedom (1-DoF) linear system. Some modifications were implemented so that OpenFresco can communicate with SAFIR. A virtual hybrid fire test (VHFT) was performed on a single degree-of-freedom (1-DoF) linear system in order to test these modifications. The communication proved to be successful and this framework can be used in future HFT where the NS is to be represented in SAFIR or other similar software packages. Following the VHFT, a HFT was performed on the 1-DoF linear structural system. National Instrument's LabView [17] is used to establish data exchange between the third-party frameworks OpenFresco (OpenSees) / Matlab / SAFIR. This simplified test setup helped validate the HFT methodology presented in [18], and the interface described above.

2 APLICABILITY OF HYBRID FIRE TESTING

It is important to study behaviour of structures exposed to fire conditions in the most similar conditions with reality. In real fires, it has been observed that the fire spreads from one compartment to another, and in order to understand the real behaviour of structures in these specific conditions, real-scale tests are required to validate the fire models and structural models.

Most of the fire tests are performed on individual structural members exposed to standard fires, based on the assumption that the predictions of these tests are overly conservative. A recent study [12] performed on a prototype 10-story moment resisting frame [19] when exposed to 2 design fire scenarios showed that the results from the simulation of the individual members are not conservative compared with the results of the complete structure simulation. The study highlighted that performing fire tests on individual members might not always be a conservatively safe solution. The beam situated in the fire compartment where the fire initiated was considered as a key element of the structure and it was studied in different possible configurations. The following boundary conditions are possible when testing the beam as a standalone member: f-f (fixed-fixed ends), f-th (fixed-fixed ends with free thermal expansions), f-h (fixed-hinged ends), f-r (fixed-rolling supports), s-s (simply supported beam), h-h (hinged-hinged ends).

Fig. 1 shows the evolution in time of the mid-span displacement of the beam when subjected to the mentioned boundary conditions. It is noted that these results come from finite element analysis and not from real tests. In most cases of the standalone member analysis (e.g., "f-f", "h-h" in *Fig. 1*), the mid-span displacements of the beam are higher than in the case when the full structure assembly was analysed ("full-scale S1 and "full-scale S2" represent the two different fire scenarios). Nevertheless, the standard testing simulation did not lead to the failure of the members, while in the case of the full-assembly simulation, the failure of the structural system occurred in both fire

scenario cases. The standalone member testing (analysis) does not capture the full effect of the surrounding members during the test. The surrounding assembly starts being exposed to fire since the fire spreads from one compartment to another. The capacity of the member is overestimated if the effect of the surrounding members is neglected. This is a clear example of HFT applicability, it is built on the concept of continually accounting for the whole assembly behaviour while realistically testing a member or sub-assembly.



Fig. 1. Mid-span displacement of the beam in different testing configurations

3 HFT ON A SINGLE DEGREE-OF-FREEDOM SYSTEM

As presented in [13], one of the biggest challenges in performing HFT is to ensure a proper communication between the PS and NS. Frameworks for experimental setup and control such as OpenFresco and UI-SIMCOR have been successfully used in seismic hybrid simulations. These frameworks ensure the proper control of the PS, which communicate in real time with the NS during the entire test.

To perform a HFT using frameworks as OpenFresco or UI-SimCor, some modifications are due, making them more generic, thus allowing them to integrate with certain finite element software that have the capability to solve structures loaded with both mechanical and fire loads (e.g., such as SAFIR).

To make the communication between SAFIR and OpenFresco possible, some modifications were implemented in SAFIR and a communication framework was developed in MATLAB. To test this communication interface, first a virtual hybrid fire test VHFT was performed, in which the PS and NS are both modelled numerically. Once the VHFT confirmed the successful continuous communication, a real HFT was performed. The main purpose of the test was to validate the hybrid fire testing methodology presented in [18] and the communication framework mentioned above. The 1-DoF linear system (*Fig. 2*) considered in this example was composed of 2 bars, defined by their stiffness. One of the bars was exposed to elevated temperature, and the thermal expansion induced changes at the interface node 2. For the sake of simplicity, the stiffness of the heated substructure was considered constant during the entire test.

3



Fig. 2. One degree-of freedom linear system

The main characteristics of simplified 1-DoF linear system considered in this example are the following:

- The stiffness of the substructures: $K_P = 11.30$ N/mm (PS) and $K_N = 11.30$ N/mm (NS)
- The length of the substructures: $L_P=260 \text{ mm}$ (PS) and $L_N=260 \text{ mm}$ (NS)
- The thermal coefficient of the material is $\alpha = 0.00012$ 1/K
- The rate of temperature increase of the heated substructure is taken as 200 °C/s (the stiffness of the PS is considered constant).

Prior the real HFT, a VHFT was performed and the results are presented in the following section.

3.1 Virtual Hybrid Fire test (VHFT)

Performing a VHFT (PS and NS are both represented numerically, i.e., separately modelled) is generally useful for the purposes of: (i) testing the framework integrating the different analytical software, and (ii) preparing for the real HFT by helping select the proper time step for the simulation in cases for which no analytical formulations are available.

The communication between SAFIR and OpenFresco is possible because of the developed communication framework in Matlab (see *Fig. 3*). In addition, modifications were done in SAFIR's code in order to allow such communication. The HFT methodology presented in [13] was used in this VHFT, and it was implemented inside the communication framework. The advantage of this particular framework is that it allows the communication between experimental setup and control frameworks with a wider range of software packages that have the capabilities to solve structures loaded with both mechanical and fire load (in this case SAFIR). Also, this communication framework is generic, and can be accessed and modified based on the needs of each case study.

In this VHFT example, the heated substructure was modelled in SAFIR, while the cold substructure was modelled in OpenSees [20]. OpenFresco was used to communicate between SAFIR and OpenSees. In a real HFT, OpenFresco will instead communicate between the tested PS and modelled NS in SAFIR.



Fig. 3. Communication between OpenFresco / OpenSees and SAFIR during the VHFT

Fig. 4 presents the evolution of the interface force versus the interface displacement of the 1-DoF linear system, i.e., the interface force and displacement of the node 2 (see *Fig. 3*). Each of the graphs presents the solution of the global (whole system) analysis of the 1-DoF linear system (referred to as "Global") along with the solution from the VHFT ("NS" is the interface solution of the NS analysed in SAFIR and "PS" is the interface solution of the PS analysed in OpenSees). The

Experimental Research

VHFT solutions are presented for two different considered time steps $\Delta t: \Delta t=1$ s and $\Delta t=10$ s. The time step refers to the time when the displacement is fixed to a constant value (since a displacement control procedure is used). Once the new solution is computed, it needs to be updated on the PS and NS before running the next step. For this example, 1 s is the time needed to update these values on the substructures.

This VHFT showed that the interface solution of the PS and NS were the same, meaning that the equilibrium and compatibility were both satisfied. Moreover, the global solution was reproduced successfully by VHFT (the lines on the plot align on top of each other). The heated substructure (NS) needs to expand but the displacement is constant during one-time step (displacement control procedure), therefore, the interface force is increasing. The force increase of the NS during one time step is obvious in this example for a time step equal to 10 s (*Fig. 4 b*)). Once the interface displacement is updated, the reaction force of the NS (heated substructure) is in equilibrium with the reaction force of the PS.

The stiffness values of the PS and NS are required for the calculation process of the updated boundary conditions (displacement of the node 2). However, as the substructures are exposed to elevated temperatures, their mechanical and material properties (which are temperature-dependent) degrade, i.e., the stiffness varies with temperature. In a real HFT the stiffnesses of the PS and NS can be updated every time step or it can be kept constant. If the stiffness is kept constant, then several iterations are required to converge to the correct solution. For this simplified example, the hypothesis that the stiffness is not affected by the elevated temperatures was adopted. Therefore, the stiffness of the PS and NS were kept constant during this entire VHFT and once the displacement was updated at the interface, the correct solution was achieved along with the equilibrium and compatibility (no iterations were required).



Fig. 4. Interface solution (force versus displacement) for the 1-DoF linear system: a) $\Delta t = 1$ s; b) $\Delta t = 10$ s

This VHFT's communication framework successfully enabled the communication between SAFIR and OpenFresco / OpenSees. The following section shows how the developed communication framework was implemented in a real HFT performed on a 1-DoF linear system. In general, performing a VHFT prior to conducting a real HFT is beneficial for many reasons. It can be used to properly select the optimum value of the required time step. It can also be helpful in determining the influence of using constant stiffness values (versus updating them every time step) of the PS and NS on the results.

3.2 HFT of a 1-DoF linear system

Following the VHFT, a HFT was performed on the single degree-of-freedom (1-DoF) linear structural system shown in *Fig. 5*. In this setup, the PS and NS were represented by springs. Prior the HFT, the stiffness of the PS was measured and had a value of +11.30 N/mm. The stiffness of the NS was assumed to be equal to the stiffness of the PS. All the characteristics of the system are presented in section 3. Since the objective of this experiment was to test and validate the

communication framework, and not necessary the specific structural response to the fire load, the unheated substructure was physically tested (PS) while the heated substructure was modelled (NS) in SAFIR. National Instrument's LabView was used to establish the data exchange between the frameworks PS / OpenFresco / Matlab / SAFIR.



Fig. 5. HFT of the 1-DoF linear system

The test setup is presented in Fig. 5. The actuator is imposing the target displacement which is measured by the string pot. The force is measured by the load cell. The properties of the measurement instruments along with the uncertainty of measurements are presented towards the end of this section.

Fig. 6 presents force versus displacement at the interface, i.e., at node 2 (*Fig.* 5), during the HFT. The global solution results from the global numerical analysis of the 1-DoF linear system (referred to as "Global" on the plot) while the results from the HFT are "PS" (the interface solution of the PS) and "NS" (the interface solution of the NS). It is noted that this work is underway, and the results presented in the *Fig.* 6 are preliminary results. A time step of 10 s was used to update the interface displacement and the time to induce the target displacement was 1 s. During one-time step, the displacement is kept constant at the interface and this induces an increase of the reaction force of the NS reduces and has the same value as the reaction force of the PS. This shows that the compatibility and the equilibrium are both satisfied in each time step. In order to reduce the spikes of the NS's solution, a shorter time step can be used as presented in the VHFT (see *Fig.* 4 *a*)). A slight deviation of the HFT solution from the global analytical solution was observed toward the end of the test which could be due to the spring bowing out of plane once it gets compressed.



Fig. 6. Force-displacement during HFT of the 1-DoF linear system

This simplified HFT showed successful implementation of the developed communication framework. It is noted that this setup could be reversed where the fire load is applied to the PS instead of the NS. In real HFT applications, the fire would potentially be applied on the tested substructure (PS), as the unheated structure could be numerically modelled with less uncertainty than the heated sub-structure. This framework is also ready to be extended to more complicated systems with multiple DoF (requiring multiple actuators).

Although the work described herein is a proof of concept for the purpose of presenting the successful implementation of the developed communication framework, some details of the measurements are worth mentioning. It is noted that the string pot used was a Celesco SP2-50, with a 1270 mm maximum measuring length. Its combined uncertainty was ± 2 %, and the total expanded uncertainty based on coverage factor of two (corresponding to a 95 % confidence interval) was 4 %. The load cell was an OMEGA LCWD-20K, with a 90 kN capacity. Its combined uncertainty was ± 10 %, and the total expanded uncertainty was 20 %. The uncertainties for both measurements were calculated based on [21].

4 CONCLUSIONS

There is a lack of understanding of how structures, as whole systems, perform under realistic fires. The Hybrid Fire Testing (HFT) method, through sub-structuring, and testing individual members/components, while simultaneously accounting for the action of the surrounding structure which is computationally modelled aside, has proven to have great potential for solving this problem. However, a main challenge in HFT is the continuous communication between the tested sub-structure and its numerically modelled surroundings. This paper presented the successful implementation of a newly developed communication framework for HFT. It was performed on a single degree-of-freedom linear system, both in a virtual and in a real setup.

DISCLAIMER

Certain commercial software or materials are identified to describe a procedure or concept adequately; such identification is not intended to imply recommendation, endorsement, or implication by the National Institute of Standards and Technology (NIST) that the software or materials are necessarily the best available for the purpose.

REFERENCES

1. Kiel, M. (1989). "Entwicklung einer intelligenten Prüfmaschine für brandbeanspruchte Gesamttragwerke." Proceedings of the Braunschweiger Brandschutztage 1989, Braunschweig, Germany.

Chernovsky, Artur; Sauca, Ana; Seif, Mina; Zhang, Chao. "Hybrid Fire Testing of a Single Degree-of-Freedom Linear System." Paper presented at SiF18 – The 10th International Conference on Structures in Fire, Belfast, United Kingdom. June 6, 2018 - June 8, 2018

- Korzen, M., Magonette, G., Buchet, Ph. (1999). "Mechanical Loading of Columns in Fire Tests by Means of the Substructuring Method." *Zeitschrift für Angewandte Mathematik und Mechanik*, Vol. 79, pp. S617-S618.
- 3. Robert, F., Rimlinger, S., Collignon, C., (2009). "Promethee, Fire Resistance Facility Taking Into Account the Surrounding Structure." *1st international Workshop on Concrete Spalling due to Fire Exposure*, 2009, 3-5 sept.
- Mostafei, H. (2013a). "Hybrid Fire Testing for Assessing Performance of Structures in Fire Application." *Fire Safety Journal*, Vol. 56, pp. 30-38.
- 5. Whyte, C.A., Mackie, K.R. and Stojadinovici, B. (2016). "Hybrid Simulation of Thermomechanical Structural Response", *Journal of Structural Engineering*, 142(2): 04015107-1 04015107-11.
- Schulthess, P., Neuenschwander, M., Knoblock, M. and Fontana, M. (2016). "Consolidated Fire Analysis – Coupled Thermo-Mechanical Modelling for Global Structural Fire Analysis", 9th International conference on Structures in Fire, 8-10 June, pg. 819-826.
- Tondini, N., Hoang, V. L., Demonceau, J.-F., & Franssen, J.-M. (2013). "Experimental and numerical investigation of high-strength steel circular columns subjected to fire." *Journal of Constructional Steel Research*, 80, 57-81.
- 8. Sauca, A., Mergny, E., Gernay, T., and Franssen, J.M. (2017a). "A method for Hybrid Fire Testing: Development, implementation and numerical application." *Proceedings of Applications of Structural Fire Engineering (ASFI'17)*, September 7-8.
- 9. ASTM E119-16a (2016). Standard test methods for fire tests of building construction and materials, 2016.
- Rackauskaite, E., Kotsovinos, P., Jeffers, A., Rein, G. (2017b). "Structural analysis of multi-storey steel frame exposed to travelling fires and traditional design fires." *Engineering Structures*, 150, 271-287.
- Choe, L., Ramesh, S, Hoehler, M., Seif, M., Gross, J., Zhang, C., and Bundy, M. (2018). "NIST Technical Note 1983. National Fire Research Laboratory Commissioning Project: Testing Steel Beams under Localized Fire Exposure."
- Sauca, A., Zhang, C., Seif, M. (2018). "Stability of Steel Structures at Elevated Temperature: A Hybrid Fire Testing Approach" Proceedings of the Annual Stability Conference, April 10-13
- Sauca, A., (2017). "Development and implementation of a methodology for hybrid fire testing applied to concrete structures with elastic boundary conditions." *Doctoral Thesis*, University of Liege, Liege, Belgium, 2017
- 14. OpenFresco 2016. "Open Framework for Experimental Setup and Control." UC Berkley
- 15. UI-SimCor, University of Illinoi
- Franssen, J.-M. (2005), "SAFIR, A Thermal/Structural Program Modelling Structures under Fire", A.I.S.C. Engineering Journal, 42 (3) 143–158.
- 17. Elliott, C., Vijayakumar, V., Zink, W., Hansen, R. (2007), "National Instruments LabVIEW: A Programming Environment for Laboratory Automation and Measurement" Cytokinetics, Inc., San Francisco, CA.
- Sauca, A., Gernay, T., Robert, F., Tondini, N., & Franssen, J.-M. (2017b). "Hybrid Fire Testing: Discussion on Stability and Implementation of a New Method in a Virtual Environment." *Journal of Structural Fire Engineering* (in press).
- Sadek, F., Main, J. A., Lew, H. S., Robert, S. D., Chiarito, V. P., and El-Tawil, S. (2010). "NIST Technical Note 1669. An Experimental and Computational Study of Steel Moment Connection under a Column Removal Scenario."
- 20. OpenSees. UC Berkley
- Taylor, B., N., Kuyatt, C., E. (1994). "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results." NIST Technical Note 1297

Numerically-Predicted Velocities of C1 and C2 Hydrofluorocarbon Refrigerant Flames with Air

Gregory LINTERIS*, Valeri BABUSHOK

Engineering Laboratory, National Institute of Standards and Technology Gaithersburg, MD, USA Phone: 301-975-2283, Email: linteris@nist.gov

* Corresponding Author

ABSTRACT

Due to their high global warming potentials, many existing working fluids for heating, cooling and refrigeration equipment are being phased out. Their replacements will often be flammable or slightly flammable, and the burning velocity of refrigerant-air mixtures is being used as a metric to rank their flammability. To allow industry to estimate the flammability of new blends of agents, predictive tools for the burning velocity of refrigerants are being developed, and calculating burning velocity requires a kinetic mechanism. The National Institute of Standards and Technology hydrofluorocarbon (HFC) mechanism was developed 20 years ago to describe hydrocarbon-air flames with added trace amounts of hydrofluorocarbon fire retardants (primarily CH2F2, CF3H, CF4, C2H2F4, C2HF5, and C_2F_6). In the present work, the mechanism has been updated slightly to include new HFC compounds, more recent rate data, and rate data for new species. The modified mechanism is used to predict steady, planar, 1D, unstretched burning velocities for mixtures of air with each of the one- and two-carbon saturated HFC compounds R41 (CH₃F), R32 (CH₂F₂), R161 (C₂F₅H), R152 (CH₂F-CH₂F), R152a (CH₃-CHF₂), R143 (CH₂F-CHF₂), R143a (CH₃-CF₃), R134 (CHF₂-CHF₂), and R134a (CH₂F-CF₃), for which existing experimental data were available. Simulation results are present for a range of fuel-air equivalence ratio ϕ , for comparison with the available experimental data. Agreement is reasonable, and major kinetic pathways and radical populations are explored to uncover the general reaction properties of these new flames.

1. INTRODUCTION

Existing refrigerant working fluids in vapor-compression heating/cooling equipment that have high global warming potential (GWP)are being phased out through international treaties (i.e., the Kigali Agreement, an addendum to the Montreal Protocol.). Low-GWP replacements have been developed, primarily by adding double bonds or hydrogen atoms to the molecules, which makes them break down in the troposphere. Unfortunately, these properties also make them more flammable. Flammable refrigerants are a new challenge for the heating, ventilation, and airconditioning and refrigeration industry, and new building standards are required for the safe use of the new compounds. Burning velocity has been adopted as part of the standard to characterize the new refrigerants. The laminar burning velocity is a useful parameter for quantifying fire risk since it is fundamental parameter that combines the effects of energy release, heat and mass transfer, and overall reaction rate. Moreover, predictions of turbulent flame speed are based on the laminar burning velocity, so the overpressure hazard and explosion hazard are both tied to the laminar burning velocity.

To meet the challenges of high efficiency, good volumetric capacity, low toxicity, zero ozone depletion potential, low GWP, and low flammability, industry will use blends of compounds. Analytical methods exist for optimizing the blends for all these properties except for flammability. To allow industry to estimate the flammability of new blends of agents, predictive tools for the burning velocity of refrigerants are being developed. Such a tool would help to accelerate the search for efficient blends that minimize the flammability hazard.

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

There are three parts to the development of the burning velocity predictive ability: 1.) understanding the experimental flame features so that the experimental burning velocity data can be accurately reduced and compared with the appropriate numerical simulation, 2.) applying and developing the necessary numerical simulation tools, and 3.) acquiring or developing the necessary input data to the models so that they can be implemented. The first two parts are dealt with in separate parts of the current project; the third part, obtaining the necessary input data, is the subject of the present manuscript. The necessary input parameters consist of: thermodynamic data (enthalpy and entropy as a function of temperature), transport data (Lennard-Jones parameters), and gas-phase reaction rate data (Arrhenius parameters; i.e., activation energy, pre-exponential, and pressure-dependency term) for all elementary reactions. In addition, the spectral radiation properties of the refrigerants and combustion products, as a function of temperature and pressure for the latter, will eventually be required to account for radiation heat losses in the flames.

As a starting point to obtain the necessary input kinetic parameters for flame modeling, an existing model for hydrofluorocarbon (HFC) flame behavior is adopted in the present work, and updated slightly. The National Institute of Standards and Technology (NIST) hydrofluorocarbon (HFC) mechanism (and its associated transport parameters) was developed 20 years ago to describe the addition of HFC fire suppressants to hydrocarbon-air flames. While some of the one-, two-, and three-carbon HFC compounds are the same as those being considered as refrigerants (as pure compounds or in blends), an assumption in the original model was that the HFC suppressant was added at small concentrations to stable hydrocarbon-air flames. Hence, the predominant species attacking the HFC reactants were the typical hydrocarbon radical pool species (H, O, and OH), and hydrocarbon radicals. For flames of pure refrigerants in air, however, the attack by fluorinated radicals is expected (Babushok et al., 2012) and these reactions must be more thoroughly considered in the reaction set. The original NIST HFC mechanism is currently being updated and extended to apply to new refrigerants added at high concentrations in air, starting with R32 (Burgess Jr et al., 2018), and will likely require additional reactions and species. As a first step in this process, however, the existing NIST HFC mechanism is applied to predict burning velocities of some pure C_1 and C_2 HFC compounds in air, and the results are compared to existing experimental data for burning velocity.

The NIST HFC mechanism was first tested with no modifications. Agreement was initially good for some compounds and poor for others. Consequently, some improvements were made, including addition of new HFC intermediates and their reactions, more recent rate data, and updated thermodynamic data, as described below. The modified mechanism is then used to predict steady, adiabatic, planar, 1D, unstretched burning velocities S_u^0 for mixtures of each refrigerant with air, over a range of fuel-air equivalence ratio ϕ , for comparison with experimental values in the literature. The compounds modeled are the saturated C_1 and C_2 HFC compounds R50 (CH₄), R41 (CH₃F), R32 (CH₂F₂), R170 (C₂H₆) R161 (C₂F₅H), R152 (CH₂F-CH₂F), R152a (CH₃-CHF₂), R143 (CH₂F-CHF₂), R143a (CH₃-CF₃), R134 (CHF₂-CHF₂), and R134a (CH₂F-CF₃), for which existing data were available.

2. KINETIC MODEL

The starting kinetic model is from the NIST C1-C2 HFC model (Burgess Jr et al., 1995a, 1995b). That mechanism had subsequently been updated and expanded to include larger three-carbon HFC's (R-227ea) and other compounds, as described in (Babushok et al., 2015), and to account for new reactions important for the combustion of pure fire suppressants (R23, R125, and R227ea) in air (Babushok et al., 2012). Since the original work, a rather large amount of new kinetic data on the reactions of fluorine containing species has been published. Hence, the kinetic model has been updated to include some new species and recent reaction rate data. The thermodynamic data for fluorinecontaining species in the mechanism haver also been updated using the data of Burcat and co-workers (Goos et al., 2012). For the hydrocarbon sub-mechanism, GRIMech 2.11 (Frenklach et al., 1995) was originally used, and this has been updated to GRIMech 3.0 (Smith et al., 2000). The successive stages of the previously updated NIST HFC mechanism has been validated in numerous studies, comparing predicted and measured laminar burning velocities (Choi et al., 2016; Linteris, 1996; Linteris et al., 1998; Linteris and Truett, 1996; Pagliaro et al., 2016a; Pagliaro et al., 2016b), counterflow diffusion flame extinction conditions (Saso et al., 2000), co-flow diffusion flame extinction conditions (Katta et al., 2006; Takahashi et al., 2015), and intermediate species profiles in low-pressure premixed flames (L'Esperance et al., 1997; L'Esperance et al., 1999; Williams et al., 2000) and flow reactors (Takahashi et al., 2007; Yu et al., 2006).

Although they are not currently used in refrigerant blends, the mono-fluoro alkanes R41 (CH₃F) and R161 (CH₃-CH₂F) are included in the present study for completeness, and because experimental burning velocity data for them

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

are available. Although the compounds were present in the original HFC mechanism, they were not thoroughly treated there because they are not fire suppressants (they are highly flammable), and as trace intermediates in fire suppression studies they are present only at very low concentrations. Not surprisingly, for these two compounds, the predicted peak laminar burning velocity using the original HFC mechanism was in significant error (35 % low, and 22 % high, respectively) as compared to the experimental values. Hence, additions and changes were made to the NIST HFC mechanism to improve the agreement, including, for CH₃F, modifications to its heat of formation and to reactions of CH_2F with O_2 ; and for C_2F_5H , modifications to the reaction rates (within their experimental uncertainty) of some of its initial decomposition products. The final mechanism used in the present work is referred to below as the updated NIST HFC mechanism, and it has 101 species and 915 reactions.

It should be noted that the present kinetic model should be considered as a starting point for further development and refinement. Numerous changes to both the rates and the reactions may be made once a variety of experimental data and theoretical results are available for testing the mechanism.

3. FLAME MODEL

The laminar burning velocities were calculated using the open-source Cantera software package (Goodwin et al., 2016). The equations of mass, species, and energy conservation are solved numerically for the initial gas compositions, temperature (298 K), and pressure (101.33 kPa) corresponding to those in the experiments. The solution assumes isobaric, adiabatic, steady, planar, one-dimensional, laminar flow and neglects radiation and the Dufour effect, but includes thermal diffusion. Molecular diffusion is modeled with the multi-component transport equations. The boundary conditions, corresponding to a freely-propagating flame, are a fixed inlet temperature of 298 K and specified mass flux fractions at the inlet, and vanishing gradients downstream from the flame. The maximum gradient and curvature parameters in the simulation are selected to provide about 150 grid points in the solution, providing the unstretched laminar burning velocity that is grid independent.

4. EXPERIMENTAL DATA

The experimental burning velocity data for the comparisons are from Takizawa and co-workers. For all of the refrigerants, the following experimental arrangement was used, and for some of the refrigerants, additional experiments were conducted. In the first, an electrical spark ignited the premixed fuel and air in a constant volume spherical vessel (3.05 L volume), and a dynamic pressure transducer recorded the pressure rise (Takizawa et al., 2005, 2006). Using the pressure vs. time data, a two-zone thermodynamic model of the burned and unburned gases yielded the burning velocity as a function of temperature and pressure, and curve fits to that surface were used to extrapolate to room temperature conditions (298 K, 101.33 kPa), for which the data are presented. The curve fit parameters are also presented in the references, so experimentally-derived burning velocity data at other pressures and temperatures can be extracted.

For R32, R143, R143a, and R152a, experiments were also conducted in a cylindrical vessel (volume of 3.92 L) with optical access at the ends, which allowed schlieren imaging of the flame (Takizawa et al., 2005). A high-speed camera recorded the increase in flame radius with time (defined as the burned gas burning velocity), and multiplying this by the density ratio of burned and unburned gases (calculated by assuming chemical equilibrium) produced the burning velocity relative to the unburned gases.

For R32, several other experiments were also used. The constant volume, pressure rise method was used with a slightly different chamber (cylindrical, volume 2.92 L), and experiments using this chamber were conducted under both normal gravity (1g) and microgravity (0g) conditions (Takizawa et al., 2013).

5. RESULTS AND DISCUSSION

Figure 1 and Figure 2 show, for a range of fuel-air equivalence ratios (ϕ), the adiabatic flame temperature T_{ad} (upper curves), and the laminar burning velocities S_{μ}^{0} (lower curves) calculated with Cantera. The figures also show the experimental data (points) of Takizawa and co-workers from outwardly propagating spherical flames in constant volume and constant pressure experiments (Takizawa et al., 2013; Takizawa et al., 2005, 2006). In Figure 1 and Figure 2, the open symbols denote experiments in the constant volume apparatus in which pressure rise is measured

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

(Takizawa *et al.*, 2005, 2006) in 1g, while the closed circles in Figure 1 are for the same experiment in 0g (R32 only). The crosses (in both Figure 1 and Figure 2) denote results from experiments in the constant pressure device using schlieren imaging of the flame growth (Takizawa *et al.*, 2013). Although no experimental data are available for pure R134a- or R134-air flames, the laminar burning velocities were calculated for illustration purposes. Table 1 summarizes the peak T_{ad} and S_u^0 from the experiments and simulations for each compound. Also shown are the stoichiometric volume fraction of each compound and the ratio of fluorine to hydrogen atoms in the original mixture, expressed as F/(F+H).



Figure 1: Burning velocity (left scale) and adiabatic flame temperature (right) for C₁ hydrofluorocarbons in air.

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018
2695, Page 5



Figure 2: Burning velocity (left scale) and adiabatic flame temperature (right) for C₂ hydrofluorocarbons in air.

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

2695, Page 6

The stoichiometric concentration of the agents is generally higher than that for hydrocarbons, and increases as the fluorine loading in the molecule increases. The maximum T_{ad} are similar for the flammable HFCs (R161, R41, R152, R152a, R143, R143a), with the peak value in the range 2100 K $\leq T_{ad} \leq 2300$ K, which is comparable to, sometimes higher than, that of hydrocarbons. For the non-flammable refrigerants (R23, R134a, R134, and R125), the peak T_{ad} is somewhat lower (< 1960 K). As illustrated, the agreement in the measured and predicted burning velocity is reasonable for most refrigerants. For R143 and R143a, the simulations predict the peak burning velocity fairly well, although they show a peak value at leaner values of ϕ than measured in the experiments. It should be noted that agreement between predictions and measurements for the mono-fluoro compounds fluoromethane (CH₃F) and fluoroethane (C_2H_5F) using the original NIST model was poor. Modification of the enthalpy of formation for these compounds and kinetic data for several reactions was required to improve the predictions. Nonetheless, it should also be noted that stretch and burned gas thermal radiation have not yet been included in the simulations or in the reduction of the experimental data, and these might affect the results (Burrell et al., 2018; Pagliaro and Linteris, 2016).

In Table 1, burning velocities and adiabatic flame temperatures are listed from highest value of S_u^0 to lowest; both T_{ad} and S_u^0 decrease with increasing fluorine loading in the refrigerant. This is also shown in Figure 3, which presents T_{ad} (Δ symbols) and S_u^0 (• symbols) as a function of the fluorine loading the system. As indicated, the asymmetrical isomers tend to be less flammable than the symmetrical ones, having both lower T_{ad} and S_{u}^{0} .

Refrigerant	Formula	T _{ad} K	S ^{"0} max (Expt.) cm/s	S ^{"0} max (Calc.) cm/s	X _{stoic.} %	F/(F+H)
<u>Flammable:</u>						
R-170	C_2H_6	2265	40.9	43.1	5.66	0.00
R161	C_2H_5F	2265	38.3	41	6.54	0.17
R-50	CH ₄	2230	36.5	38.6	9.5	0.00
R-152	CH ₂ F-CH ₂ F	2278	30.1	32.1	7.75	0.33
R-41	CH₃F	2273	28.3	27.2	12.3	0.25
R-152a	CH ₃ -CHF ₂	2227	23.6	24.9	7.75	0.33
R-143	CH ₂ F-CHF ₂	2248	13.1	13.7	9.5	0.50
R-32	CH_2F_2	2207	6.7	7.3	17.4	0.50
R-143a	CH ₃ -CF ₃	2115	7.1	6.1	9.5	0.50
<u>Non-Flammab</u>	l <u>e:</u>					
R-134	CHF ₂ -CHF ₂	1958		4.6	12.3	0.67
R134a	CH_2F-CF_3	1931		1.8	12.3	0.67
R-125	CHF_2-CF_3	1793		1.56 (at 400K)	17.4	0.83
R-23	CHF ₃	1713		0.57 (at 400K)	29.6	0.75
R-116	C2F6	1389		0	29.58	1.00

Table 1:	: Burning	velocities a	and adiabatic	combusti	on temper	atures for	stoichio	metric re	efrigerant	•air n	nixtures
			(ini	tial temper	ature 298	8 K, 1 bar)).				

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018



Figure 3: Burning velocity (\bullet , left scale) and adiabatic flame temperature (Δ , right scale) as a function of the ratio of F atoms to F plus H atoms in the initial refrigerant-air mixture.

Using the calculated flame structures, the variation in radical concentrations with fluorine loading is determined. The peak volume fraction of chain-carrying radicals (H, O, and OH) is typically near the point of peak heat release in the flame, or the location reaching approximately 95 % of the peak flame temperature. Figure 4 (note semi-log plot) shows the sum of the peak volume fraction of chain-carrying radicals (H, O, and OH) and of F-containing radicals (at location of peak [OH]), as a function of the fluorine loading. As illustrated, the former drops off rapidly as the number of H atoms in the system becomes close to that of F atoms F/(F+H)=0.5, while the F radicals increase rapidly. Hence, at higher fluorine loading, the chemistry becomes dominated by fluorine-containing radicals. Figure 5 shows that with increased fluorine loading, the volume fraction of both F-atoms (at the point of maximum OH volume fraction) and sum of the peak for F-containing radicals increases steadily, as does the equilibrium F atom volume fraction, which becomes higher than the value in the flame zone. For these flames equilibrium F atom volume fraction (far downstream in the calculation domain) can be on the order of 1 %.

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018



Figure 4: Volume fraction of total chain-branching radicals (H+O+OH) and F atoms in the flame reaction zone (i.e., the point of peak [OH]) as a function of the ratio of fluorine loading.



Figure 5: Volume fraction for fluorine radicals (F-radicals) and F atoms in the flame reaction zone, and for F atoms at equilibrium as a function fluorine loading.

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

6. CONCLUSIONS

The unstretched, laminar, planar, 1D, adiabatic burning velocities of saturated C1 and C2 HFC refrigerants (R41, R32, R161, R152, R152a, R143, R143a, R134, R134a were calculated using the original NIST HFC mechanism available in the literature and compared to existing experimental data. The predictions, for a range of fuel-air equivalence ratios, were in significant disagreement for CH₃F and C₂H₅F, and mild disagreement for other compounds. Consequently, the NIST HFC mechanism was modified with additional reactions, using more recent rate data in the literature, and with updated thermodynamic properties. After the changes, the agreement for these refrigerants is reasonable.

The mechanism is then used to examine the properties of the refrigerant-air flames. Adiabatic temperatures of the flammable refrigerant-air flames are comparable to, and sometimes higher than, similar hydrocarbons, whereas T_{ad} of the non-flammable refrigerants is lower. Burning velocity and flame temperature decrease as the fluorine to hydrogen ratio in the reactants increase. The symmetrical isomers of the fluoroethanes (R152, R143, R134) have higher adiabatic flame temperature and laminar burning velocity than the asymmetrical isomers (R152a, R143a, R134a). Analysis of the flame structures revealed that with increasing fluorine to hydrogen ratio, the chainbranching radical concentrations in the flame decrease, and fluorine-containing radicals, particularly F atom, increase. At high enough F/H ratio, the F atom equilibrium values are even higher than those in the flame zone.

While agreement between measured and predicted burning velocities is reasonable using the updated mechanism, work should be done to update and improve the kinetic mechanism. Also, flame stretch and radiation have not been included in either reduction of the experimental data or in the flame modeling, and these should be included in future work to increase the accuracy.

REFERENCES

- Babushok, V. I., Linteris, G. T., Burgess, D. R., & Baker, P. T. (2015). Hydrocarbon flame inhibition by C3H2F3Br (2-BTP). Combust. Flame, 162(4), 1104-1112.
- Babushok, V. I., Linteris, G. T., & Meier, O. (2012). Combustion properties of halogenated fire suppressants. Combust Flame, 159, 3569-3575.
- Burgess Jr, D. R., Manion, J. A., Burrell, R., Babushok, V. I., Hegetschweiler, M. J., & Linteris, G. T. (2018). Development and validation of a mechanism for flame propagation in R-32/air mixtures. 2018 Eastern States Section Meeting of the Combustion Institute State College, PA: The Combustion Institute
- Burgess Jr, D. R., Zachariah, M. R., Tsang, W., & Westmoreland, P. R. (1995a). Thermochemical and chemical kinetic data for fluorinated hydrocarbons. Prog. Energy Combust. Sci., 21(6), 453-529.
- Burgess Jr, D. R., Zachariah, M. R., Tsang, W., & Westmoreland, P. R. (1995b). Thermochemical and Chemical Kinetic Data for Fluorinated Hydrocarbons. (NIST Technical Note 1412). Gaithersburg, MD: National Institute of Standards and Technology.
- Burrell, R., Pagliaro, J. L., & Linteris, G. T. (2018). Effects of stretch and thermal radiation on difluoromethane-air burning velocity measurements in constant volume spherically expanding flames. Proc. Combust. Inst., 37, submitted.
- Choi, B. C., Park, J. S., & Ghoniem, A. F. (2016). Characteristics of outwardly propagating spherical flames of R134a(C2H2F4)/CH4/O-2/N-2 mixtures in a constant volume combustion chamber. Energy, 95, 517-527.
- Frenklach, M., Wang, H., Yu, C.-L., Goldenberg, M., Bowman, C. T., Hanson, R. K., Davidson, D. F., Chang, E. J., Smith, G. P., Golden, D. M., Gardiner, W. C., & Lissianski, V. (1995). GRI-mech: an optimized detailed chemical reaction mechanism for methane combustion. (Gas Research Institute Topical Report No. GRI-95/0058, http://www.me.berkeley.edu/gri_mech). Chicago, IL.
- Goodwin, D. G., Moffat, H. K., & Speth, R. L. (2016). Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes. Version 2.1.1. Pasedena, CA: California Institute of Technology, Retrieved from http://www.cantera.org.

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

- Goos, E., Burcat, A., & Ruscic, B. (2012). Extended third millennium thermodynamic database for combustion and air-pollution use with updates from active thermochemical tables. Retrieved from Ftp://ftp.technion.ac.il/pub/supported/aetdd/thermodynamics/BURCAT.THR
- Katta, V. R., Takahashi, F., & Linteris, G. T. (2006). Fire-suppression characteristics of CF₃H in a cup burner. Combust. Flame, 144(4), 645-661.
- L'Esperance, D., Williams, B. A., & Fleming, J. W. (1997). Detection of fluorocarbon intermediates in low-pressure premixed flames by laser-induced fluorescence. Chemical Physics Letters, 280(1-2), 113-118.
- L'Esperance, D., Williams, B. A., & Fleming, J. W. (1999). Intermediate species profiles in low pressure premixed flames inhibited by fluoromethanes. Combust. Flame, 117(4), 709-731.
- Linteris, G. T. (1996). Numerically Predicted Structure and Burning Velocity of Premixed CO-Ar-O₂-H₂ Flames Inhibited by CF₃H. Combust. Flame, 107(1-2), 72-84.
- Linteris, G. T., Burgess, D. R., Babushok, V., Zachariah, M., Tsang, W., & Westmoreland, P. (1998). Inhibition of premixed methane-air flames by fluoroethanes and fluoropropanes. Combust. Flame, 113(1-2), 164-180.
- Linteris, G. T., & Truett, L. (1996). Inhibition of premixed methane-air flames by fluoromethanes. Combust. Flame, 105(1-2), 15-27.
- Pagliaro, J. L., Bouvet, N., & Linteris, G. T. (2016a). Premixed flame inhibition by CF3Br and C3H2F3Br (2-BTP). Combust. Flame, 169, 272-286.
- Pagliaro, J. L., & Linteris, G. T. (2016). Burning velocities of marginally flammable refrigerant-air mixtures. 2016 Eastern States Section Meeting of the Combustion Institute Princeton, NJ: The Combustion Institute
- Pagliaro, J. L., Linteris, G. T., & Babushok, V. I. (2016b). Premixed flame inhibition by C2HF3Cl2 and C2HF5. Combust. Flame, 163, 54-65.
- Saso, Y., Joboji, H., Koda, S., Saito, N., & Nishioka, M. (2000). Response of counterflow diffusion flame stabilized on a methanol pool to suppressant doping. Proc. Combust. Inst., 28, 2947-2955.
- Smith, G. P., Golden, D. M., Frenklach, M., Moriarty, N. W., Eiteneer, B., Goldenberg, M., Bowman, C. T., Hanson, R. K., Song, S., Gardiner, J. W. C., Lissianski, V. V., & Oin, Z. (2000). GRI Mech 3.0. Retrieved from http://www.me.berkeley.edu/gri_mech
- Takahashi, F., Katta, V. R., Linteris, G. T., & Babushok, V. I. (2015). Combustion inhibition and enhancement of cup-burner flames by CF3Br, C2HF5, C2HF3Cl2, and C3H2F3Br. Proc. Combust. Inst., 35, 2741-2748.
- Takahashi, K., Harada, A., Horigome, S., Cho, R., & Inomata, T. (2007). Thermal decompositons of 1,1,1trifluoroethane and pentafluoroethane in a turbulent flow reactor. Combust. Sci. Technol., 179(7), 1417-1432
- Takizawa, K., Takagi, S., Tokuhashi, K., Kondo, S., Mamiya, M., & Nagai, H. (2013). Assessment of Burning Velocity Test Methods for Mildly Flammable Refrigerants, Part 1: Closed-Vessel Method. ASHRAE Trans., 119(2), 243-254.
- Takizawa, K., Takahashi, A., Tokuhashi, K., Kondo, S., & Sekiya, A. (2005). Burning velocity measurement of fluorinated compounds by the spherical-vessel method. Combust. Flame, 141(3), 298-307.
- Takizawa, K., Takahashi, A., Tokuhashi, K., Kondo, S., & Sekiya, A. (2006). Burning velocity measurement of HFC-41, HFC-152, and HFC-161 by the spherical-vessel method. J Fluorine Chem, 127(12), 1547-1553.
- Williams, B. A., L'Esperance, D. M., & Fleming, J. W. (2000). Intermediate species profiles in low-pressure methane/oxygen flames inhibited by 2-H heptafluoropropane: Comparison of experimental data with kinetic modeling. Combust. Flame, 120(1-2), 160-172.
- Yu, H., Kennedy, E. M., Mackie, J. C., & Dlugogorski, B. Z. (2006). An experimental and kinetic modeling study of the reaction of CHF3 with methane. Environmental Science & Technology, 40(18), 5778-5785.

ACKNOWLEDGEMENT

This work was supported by the Buildings Technologies Office of the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy under contract no. DE-EE0007615 with Antonio Bouza serving as Project Manager.

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

FIRE PERFORMANCE OF LONG-SPAN COMPOSITE BEAMS WITH GRAVITY CONNECTIONS

Lisa Choe¹, Selvarajah Ramesh², Mina Seif³, Matthew S. Hoehler⁴, William Grosshandler⁵, John Gross⁶, Matthew Bundy⁷

ABSTRACT

Prescriptive fire-resistance ratings for structural components provide limited insight into the system-level performance of steel-concrete composite structures in fire. Specifically, long-span composite beam assemblies exposed to fire have vulnerabilities resulting from thermal restraint conditions that can be vastly different from those in short-span assemblies. This paper presents an overview of recent experiments on 12.8 m composite beams with various end support conditions, exposed to combined structural and fire loading. This paper focuses on the results for the specimens with double-angle beam-to-column connections, with and without slab continuity. The experiments showed that the specimens experienced similar thermal and displacement behaviour during the first 40 minutes after ignition regardless of the presence of slab continuity. The specimens exhibited local buckling of the beam near the connection, at the average bottom flange temperature of 400 °C. A thermal gradient in the specimen was observed during the heating and cooling phase. For the specimen with slab continuity, forces in the continuity bars increased when this local buckling occurred but decreased due to concrete fracture. This specimen did not collapse during the fire until its vertical displacement exceeded 1/20 span length. The collapse failure was observed during the cooling phase, which was resulted from weld fracture of the beam end connection due to contraction of the heated beam as it cooled down.

Keywords: Composite beams, double-angle connections, compartment fires, fire tests

1 INTRODUCTION

The span length of composite steel-concrete floor assemblies has increased over the years due to the availability of higher-strength materials, as well as architectural trends. The detailing of the member connections to withstand gravity loads (e.g., bolt hole and gap clearances), however, has

e-mail: william.grosshandler@nist.gov

¹Research Structural Engineer, National Institute of Standards and Technology, Gaithersburg, MD, USA.

e-mail: lisa.choe@nist.gov

² Guest Researcher, National Institute of Standards and Technology, Gaithersburg, MD, USA.

e-mail: selvarajah.ramesh@nist.gov

³Research Structural Engineer, National Institute of Standards and Technology, Gaithersburg, MD, USA. e-mail: <u>mina.seif@nist.gov</u>

⁴Research Structural Engineer, National Institute of Standards and Technology, Gaithersburg, MD, USA. e-mail: matthew.hoehler@nist.gov

⁵Guest Researcher, National Institute of Standards and Technology, Gaithersburg, MD, USA.

⁶Guest Researcher, National Institute of Standards and Technolog, Gaithersburg, MD, USA.

e-mail: john.gross@nist.gov

⁷ Supervisory Mechanical Engineer, National Institute of Standards and Technology, Gaithersburg, MD, USA. e-mail: <u>matthew.bundy@nist.gov</u>

remained largely unchanged. Fire not only leads to degradation of material strength and stiffness, but also introduces additional forces into the assembly as thermal elongation is restrained by the adjoining structure. Since the magnitude of thermal strains is proportional to the heated length, the effect of thermal restraint is more significant on longer beams than shorter ones. These forces can potentially alter the sequence of failures in beam assemblies which are often only designed for limit states relevant to gravity loads at ambient temperature.

A series of 12.8 m composite beams exposed to combined structural and fire loading was recently conducted at the National Fire Research Laboratory [1]. This paper presents an overview of the test program, and the results for the specimens with double-angle connections. The goal of this work is to provide data to help advance performance-based fire design approach for steel frame buildings with composite floor systems.

2 TEST PROGRAM

2.1 Test Matrix and Specimen Design

Table 1 shows the test matrix. Test variables included the type of beam-end connections (double angles versus single plate) and the slab continuity conditions. Test 1 was an ambient-temperature test in which the ultimate moment capacity and the sequence of failure were studied [2]. The remaining specimens were subjected to flexural loading and fire. The applied flexural load was 45 % of the ultimate moment capacity (M_{max}) measured from Test 1. The fire was generated using three natural gas burners with a maximum total heat release rate of 4 MW.

The specimens were designed in accordance with current U.S. codes and standards [3, 4]. Details are presented in [5] and summarized here. *Fig. 1a* shows a cross-section of the specimens, which consisted of a 1.8 m wide by 12.8 m long concrete slab cast on trapezoidal metal decking, and a W18×35 floor beam. The slab was partially-composite with the steel beam via 19 mm headed shear studs spaced at 305 mm (design composite effectiveness 82 %). Welded wire fabric (0.06 mm²/mm in orthogonal directions) was placed at the mid height of 83 mm thick concrete topping. The bottom flange of the beams with double-angle connections was coped at both ends. The steel beam was protected with Sprayed Fire Resistive Materials (SFRM) to achieve a 2 hour fire rating along the member and a 3 hour fire rating at the connections. *Fig. 1b* shows the specimen with the double-angle connections and slab continuity. The 9.5 mm thick angles were bolted to the beam web using three 19 mm high-strength structural bolts (with the minimum tensile strength of 830 MPa) and welded to a sacrificial plate on the flange of the column. The specimens for Test 3 and Test 5 included four no. 4 reinforcing bars (with the minimum tensile strength of 415 MPa) in the hogging moment region for crack control. Slab continuity was achieved by anchoring the reinforcement to a hollow steel section attached to the support column.

The steel beams and angles were made with hot-rolled steel with the minimum yield strength of 345 MPa and 250 MPa, respectively. The slab was cast using lightweight aggregate concrete with a design compressive cylinder strength of 28 MPa. The concrete mix included polypropylene microfibers to minimize spalling during heating. The measured ambient-temperature mechanical properties of all the structural components are reported in [5].

Test	Connection type	Slab continuity	Flexural load	Fire load
1	Double angle	No	M_{max}	
2	Double angle	No	$0.45M_{max}$	4 MW
3	Double angle	Yes	$0.45M_{max}$	4 MW
4	Single plate	No	$0.45M_{max}$	4 MW
5	Single plate	Yes	$0.45M_{max}$	4 MW

Table 1. Test Matrix



Fig. 1. a) Cross section of the specimen; b) Connections of specimen with double-angles and continuity (units: mm)

2.2 Test Setup, Fire Load, and Instrumentation

The test setup is shown in *Figs. 2a* and *2b*. The composite beam specimen was supported by W12×106 columns with bracing modules, using the connections listed in *Table 1*. The distance from the center bolt of the beam connection to the strong floor was 3.6 m. The measured average lateral stiffness of the braced column at this height (i.e. the axial stiffness provided at the beam ends) was 220 kN/mm. The beam specimen was loaded by six equally distributed point loads (*Fig. 2a*). Three loading beams running transverse to the specimen were loaded using an actuator at each of the six ends. The vertical force of each loading beam was transferred to the beam specimen via triangular trusses. The slab was laterally braced at the locations of the three loading beams. The vertical displacements and end rotations of the specimen were measured using four linear position transducers and two inclinometers mounted on the top of the slab, respectively. The thermally-induced axial restraint force at the beam support was measured using the strain gauges attached on the diagonal brace members of the bracing modules. Washer load cells were used to measure the forces in the no. 4 reinforcing bars in Tests 3 and Test 5.

Three natural gas burners (each 1 m wide \times 1.5 m long) distributed along the center of the compartment floor were used to heat the specimen. The fire was designed to achieve an upper layer temperature of around 1000 °C, and avoid combustion outside of the compartment. Peak heat release rate per unit volume was 40 kW/m³, which corresponds to a range of total fuel loads between 550 MJ/m² and 1100 MJ/m² for fire durations between one and two hours. To confine the fire below the specimen, enclosure walls (with a total surface area of 110 m²) were constructed using cold-formed steel framing and sheet steel as shown in *Fig. 2b*. The total ventilation opening area was about 5 m². The fire-exposed wall surfaces were protected with a 50 mm thick layer of ceramic blankets. *Fig. 2c* shows the predicted spatially-averaged temperatures in the upper gas layer and along the lower flange of the steel beam underneath the SFRM layer, using the Fire Dynamic Simulator (FDS) [6] prior to conducting the test series. The heat release rate increases linearly to a maximum of 4 MW during the first 15 minutes and remains steady up to 70 minutes, followed by a linear decrease to zero at 112 minutes.

Fig. 3 show the thermocouple layout. Four Inconel shielded thermocouples (K-type, 24 gauge) were used to measure the average bottom flange temperature. Glass-sheathed thermocouples were mounted at various locations of the steel beam and the connection elements. However, there was considerable difficulty in getting reliable temperatures of the steel beam beyond the first fifteen minutes of the heating. Glass shielding of thermocouples burned off and made secondary junctions when exposed to hot gas temperatures. In some cases, contact between the thermocouple beads and the steel surface could not be maintained as the beam heated and deformed. In other cases, fissures in the insulation layer permitted hot gas to penetrate to the thermocouple on steel surface. All these



effects could lead to higher temperature readings during heating. Table 2 summarizes the estimated total expanded measurement uncertainties with a coverage factor of 2 as defined in [7].

Fig. 2. a) Longitudinal section of the setup (units: cm); b) Photograph of inside the compartment; c) Temperature predictions using the FDS



Bundy, Matthew; Choe, Lisa; Gross, John; Grosshandler, William; Hoehler, Matthew; Ramesh, Selvarajah; Seif, Mina. "Fire Performance of Long-Span Composite Beams with Gravity Connections." Paper presented at SiF18 - The 10th International Conference on Structures in Fire, Belfast, United Kingdom. June 6, 2018 - June 8, 2018.

Measurand	Range (max)	Total expanded uncertainty (coverage factor of 2)
Mechanical load (Actuators)	220 kN	±10 %
Displacement	760 mm	±15 %
Strain	50,000 με	±6 %
Heat release rate	4 MW	±7 %
Steel Temperature (Inconel sheathed thermocouples)	1200 °C	±25 %

Fig. 3. Location of thermocouples along the beam length, in the beam cross section and connection region (units: mm)

RESULTS AND DISCUSSION 3

Tests 2 (T2) and Test 3 (T3) studied the behaviour of the composite beam with the double-angle connection subjected to combined flexural loading and compartment fire described above. The experiments were conducted as follows: i) A total mechanical load of 106 kN was applied and held constant at ambient temperature, ii) the heat release rate was increased to 4 MW over a period of 15 minutes and held constant until the mechanical load was eliminated, and iii) the cooling phase was initiated by linearly decreasing the heat release rate over 30 minutes.

Fig. 4a shows plots of the total mechanical load and the burner heat release rate. Fig. 4b shows the average gas temperature measured using thermocouple probes located 810 mm below the slab and the heat release rate measured with the calorimeter. Elimination of the mechanical loads was triggered by the pre-set limit of actuators. In Test 2, the actuators were depressurized when the total mechanical load dropped below 90 kN (about 85 % of its original value applied at ambient temperature) around 30 minutes, when the beam web near the angle connection buckled.

For Test 3, the pre-set load limit was removed and a displacement limit was implemented instead. The mechanical load was removed around 65 minutes after ignition when the vertical displacement of the specimen reached 620 mm. The heat release rates and resulting gas temperatures were comparable until 40 minutes (Fig. 4b). The gas temperature reached 900 °C at 15 minutes after ignition. Peak average temperatures exceeded 1000 °C around 40 minutes. The maximum measured temperatures were as much as 100 °C above the values predicted by FDS model (Fig. 2c).



Bundy, Matthew; Choe, Lisa; Gross, John; Grosshandler, William; Hoehler, Matthew; Ramesh, Selvarajah; Seif, Mina. "Fire Performance of Long-Span Composite Beams with Gravity Connections." Paper presented at SiF18 - The 10th International Conference on Structures in Fire, Belfast, United Kingdom. June 6, 2018 - June 8, 2018.

Fig. 4. a) Heat release rate based on fuel flow and total mechanical load applied using actuators; b) Heat release rate measured using the calorimeter and gas temperature inside the compartment

3.1 Thermal Response

Fig. 5 shows the range of a) the average steel temperatures at various locations, including the bottom flange of the steel beam in the middle of compartment and the connection region, and b) the slab temperatures near midspan. Since similar fire conditions were maintained in Tests 2 and 3 up to 40 minutes (*Fig.* 4), the thermal response of the specimens for these tests was comparable.

Although the upper layer gas temperature was approximately constant with an average standard deviation of 30 °C, a thermal gradient still existed in the specimen along the beam length and across the section. The metal deck rapidly heated with increasing gas temperatures, exceeding 700 °C in 1 hour after ignition. The temperature increase in the connection region with thicker SFRM and inside the slab above the steel beam was slower. In Test 3, the measured peak average temperature of the beam web next to the edges of the angle and the angle itself were around 400 °C and 200 °C, respectively. The temperatures in the slab above the steel beam remained below 300 °C.



Fig. 5. a) Steel temperatures; b) Concrete temperatures measured in Test 2 and 3

3.2 Structural Response and Failure

The specimens with the double-angle connections exhibited diverse behaviours during the heating and cooling. *Fig.* 6 shows the vertical displacement of the specimen measured 1.1 m from midspan and the average axial restraints induced at the beam ends as a function of the beam lower flange temperature. The positive values of displacements indicate downward displacements; the negative values of axial restraints represent compressive axial force in the specimen.

As shown in *Fig. 6a*, for both Tests 2 and 3, the specimen deflected downward after ignition. The beam web near the angle connection buckled around 30 minutes, at the corresponding bottom flange temperature of 400 °C. The peak axial restraint of 370 kN was observed around this temperature (*Fig. 6b*). From this point through the cooling phase, there was no further failure observed in Test 2 since the mechanical load was removed. During the 30-minute period at which the heat release rate linearly decreased, the vertical displacement continued to increase due to the weight of the specimen and loading fixtures. For Test 3, in which this pre-set load limit was eliminated, the loaded specimen continuously bent down with increasing temperatures following buckling of the beam web around 400 °C. The mechanical load was removed at the vertical displacement of 620 mm (1/20 span length), at the bottom flange temperature of 760 °C. The heated specimen continued to deflect downward while the heat release rate was decreased. The measured peak vertical displacement was 690 mm (1/18 span length). The vertical displacement decreased after the burner was shut off. *Figs 6c-d* show photographs of the specimens after cool-down.



Fig. 6. a) Vertical displacement; b) Axial restraint force as a function of average bottom flange temperature; c-d) Photograph of buckled beam web (Test 2) and deformed specimen (Test 3) after cool-down

The concrete slab in Test 3 had no. 4 reinforcing bars anchored at slab ends (*Fig. 1*). *Fig. 7a* shows a comparison of the tensile forces developed in the bars, located at 335 mm from the longitudinal centreline of the slab (*Fig. 2a*), along with the range of the vertical displacement of the specimen. *Fig. 7b* shows a photograph of fracture after the test. The tensile forces in the rebars significantly increased after buckling of the beam web occurred around 30 minutes. The magnitude of tensile forces became larger as the vertical displacements increased from 200 mm to 350 mm, followed by a rapid drop indicating the development of concrete failure near the reinforcement. The triangular shape of concrete fracture was observed after cool-down, as shown in *Fig. 7b*.



Fig. 7. a) Forces in the continuity bars; b) Slab at west end after cool-down

7

This specimen did not collapse during the fire. Around five hours after the fire was extinguished, however, the specimen collapsed by weld fracture of the east angle connection. The angles at both ends pried due to contraction of the specimen during cool-down, but there was neither weld failure of the west angle connection nor failure of shear studs.

4 CONCLUSIONS

The fire experiments on the loaded steel-concrete composite beam specimens with double-angle connections (Tests 2 and 3) showed:

- Similar thermal and structural responses during the first 40 minutes after ignition;
- A thermal gradient along the beam length and across the section during the heating and cooling phase in Test 3;
- Local buckling of the beam near the connection and increase of tensile forces in the continuity bars around 30 minutes after ignition, at the average bottom flange temperature of 400 °C;
- Vertical displacements exceeding 1/20 times the span length during the fire in Test 3; and
- Collapse failure by weld fracture of the double angle connection at east end during the cooling phase in Test 3.

ACKNOWLEDGMENTS

The authors would like to acknowledge the great support of NIST colleagues for their hard work to safely complete the experiments. They would also like to thank the expert panel members including William Baker, Glenn Bell, Craig Beyler, Luke Bisby, Ian Burgess, Charlie Carter, Charles Clifton, Michael Engelhardt, Graeme Flint, Nestor Iwankiw, Kevin LaMalva, Roberto Leon, Jose Torero, and Amit Varma for their support throughout this project.

REFERENCES

- 1. Bundy, M., Hamins, A., Gross, J., Grosshandler, W., and Choe, L. (2016). *Structural fire experimental capabilities at the NIST National Fire Research Laboratory*. Fire Technology, 52(4), 959-966.
- Choe, L., Ramesh, S., Hoehler, M., Gross, J. (2018). Experimental Study on Long-Span Composite Floor Beams Subject to Fire: Baseline Data at Ambient Temperature. *Proc. of Structures Congress* 2018, Fort Worth, Texas, April 19-21, 2018.
- 3. ANSI/AISC-360 (2016). *Specificication for Structural Steel Buildings*. American Institute of Steel Construction, Chicago, Illinois.
- 4. ASCE/SEI 7 (2016). *Minimum Design Loads for Buildings and Other Structures*. American Society of Civil Engineers, Reston, Virginia.
- Ramesh, S., Choe, L., Hoehler, M., Grosshandler, W., Gross, J. (2018). Design and Construction of Long-Span Composite Beam Specimens for Large Structural-Fire Tests. *Proc. of Structures Congress* 2018, Forth Worth, Texas, April 19-21, 2018.
- McGrattan, K., Hostikka, S., McDermott, R., Floyd, J., Weinschenk, C., and Overholt, K. (2013). Fire Dynamics Simulator, Technical Reference Guide, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, and VTT Technical Research Centre of Finland, Espoo, Finland, sixth edition, September 2013.
- Taylor, B. N., and Kuyatt, C. E. (1994). Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. National Institute of Standards and Technology, NIST Technical Note TN-1297, Gaithersburg, Maryland.

EXPERIMENTAL FIRE-SIMULATOR FOR POST-FLASHOVER COMPARTMENT FIRES

Daniel Brandon¹, Joachim Schmid², Joseph Su³, Matthew S. Hoehler⁴, Birgit Östman⁵, Amanda Kimball⁶

ABSTRACT

The number of tall timber buildings around the world is rapidly increasing as a result of changes in regulations and the development of new engineered timber products. However, due to the combustibility of timber, the fire safety of tall timber buildings has been questioned. Building regulations for structural elements are based on fixed periods for which specimens shall resist standard fire exposure in a fire resistance furnace. Because no distinction is made between the exposure in fire resistance tests of combustible and non-combustible specimens, less conventional testing methods have been used for research of timber structures. This study aims to identify aspects that are important to simulate realistic fire conditions relevant to assess the structural performance of timber in post-flashover fires. A test method is developed aiming to replicate conditions in compartment fire tests using an experimental fire simulator that results in similar damage types and rates of damage in the timber specimen. Based on observations from these tests and results obtained using other testing methods, the applicability of standard fire resistance tests and other nonconventional tests for investigating the fire performance of timber are discussed.

Keywords: Cross laminated timber, compartment fires, fire resistance tests, simulation

1 **INTRODUCTION**

The number of tall timber buildings around the world is rapidly increasing as a result of changes in regulations and the development of new engineered, often glued, timber products; such as Cross Laminated Timber (CLT). This increase in building height creates new challenges for fire safety. Various methods have been used to study the performance of timber products in post flashover fires; such as standard fire resistance tests [1] and tests with a radiant heat source in an open environment (e.g. [2]). Differences in results from these tests have led to various hypotheses regarding the important aspects to achieve realistic fire exposures relevant for timber structures [3,4].

This paper focusses on key parameters that must be controlled to produce results relevant for the structural assessment of timber in post-flashover compartment fires. While full scale compartment

¹Researcher, Fire Research, RISE Safety, Sweden.

e-mail: daniel.brandon@ri.se

² Research assistant, PhD Candidate, IBK, ETH Zürich, Switzerland.

e-mail: schmid@ibk.baug.ethz.ch

³ Principal Research Officer, National Research Council of Canada, Ottawa, Canada.

e-mail: joseph.su@nrc-cnrc.gc.ca

⁴ Research Structural Engineer, National Institute of Standards and Technology, Gaithersburg, USA. e-mail: matthew.hoehler@nist.gov

⁵ Senior Advisor, Linneaus University, Växjö, Sweden.

e-mail: birgit.ostman@lnu.se

⁶Research Director, Fire Protection Research Foundation, Quincy, MA, USA.

e-mail: akimball@nfpa.org

Brandon, Daniel; Hoehler, Matthew; Kimball, Amanda; Ostman, Birgit; Schmid, Joachim; Su, Joseph. "Experimental Fire-Simulator For Post-Flashover Compartment Fires." Paper presented at SiF18 - The 10th International Conference on Structures in Fire, Belfast, United Kingdom. June 6, 2018 - June 8, 2018.

fire tests can be used to represent real building fires, these tests can be time consuming and costly. Therefore, we present an intermediate scale method to replicate relevant conditions of a specific compartment fire test [5]. Successful replication of the compartment fire should result in similar damage to that observed in a comparable full scale test.

2 REFERENCE COMPARTMENT FIRE TEST

An extensive study led by the Fire Protection Research Foundation involved six full scale compartment fire experiments, performed by the National Research Council Canada and the National Institute of Standards and Technology to study the contribution of exposed Cross Laminated Timber (CLT) to the fuel load of the fire [5]. The study involved compartment tests with various configurations of exposed CLT surfaces and ventilation. Due to weakening of the CLT adhesive in the fire, fire induced delamination (hereafter *delamination*) occurred during the compartment fire tests with exposed CLT. In the compartment tests, delamination led to increased combustion and subsequent fire growth or continuation of the fully-developed fire [5]. Delamination depends on the adhesive used to produce CLT. To study whether delamination would have occurred had other adhesives been used, an intermediate scale fire testing method was developed as discussed herein. For this, compartment Test 1-4 described by Su et al. [5] was chosen for reference. A short summary of the test setup is discussed below.

The compartment was made of 175 mm thick 5-ply CLT walls and ceiling. The interior of the compartment was 9.1 m \times 4.6 m \times 2.7 m (depth \times width \times height) and there was an opening of 1.8 m \times 2.0 m in one of the 4.6 m \times 2.7 m walls. Three layers of 15.9 mm type X gypsum boards were applied on the CLT walls and the CLT ceiling was left exposed. Typical residential furniture with a fuel load density of 550 MJ/m² was used as fuel.

After ignition, flashover occurred after approximately 11 min and the exposed ceiling became fully involved in the fire. A significant fire plume was observed during the fully-developed stage of the fire. Two layers of the exposed CLT delaminated. Delamination of the second layer occurred during the decay phase of the fire, leading to a secondary flashover. The results from Test 1-4 are used in Section 3 to benchmark the fire conditions in the intermediate scale tests.

3 EXPERIMENTAL FIRE SIMULATOR FOR POST-FLASHOVER FIRES

An experimental method to replicate fire conditions was needed to study whether other adhesives would have performed better under the conditions of Test 1-4 (discussed above). For the purpose of this study, a successful replication of the fire conditions of the compartment fire test in a furnace should lead to similar damage in both tests if the same product is tested. Recent studies have indicated that the oxygen concentration of ambient gasses has a significant influence on the rate of damage of timber exposed to fire. Timber tested in fire conditions involving oxygen concentrations over 10 % had significant char oxidation leading to the gasification of protective char layers and more thermal damage of the timber [6]. Therefore, oxygen concentrations, in addition to plate thermometer temperatures, measured in Test 1-4 were used to replicate the conditions of a flashover compartment fire in a furnace. This setup is referred to here as the 'experimental fire simulator'.

To study whether delamination would occur for an arbitrary adhesive, the delamination induced temperature increase and the reduction of oxygen concentration after 120 min in Test 1-4 were ignored (Figure 1). The target oxygen content and temperature in the final stage of the test (from 120 min to 180 min) were extrapolated based on a similar compartment test without exposed CLT (Test 1-1 in [5]).

Two replicates (denoted A and B) of CLT panels constructed using five different adhesives were tested. In this paper, the results of two polyurethane adhesives (referred to as PU1 and PU2) are discussed. PU1 is the same adhesive as used in Test 1-4. The other polyurethane adhesive (PU2) was enhanced to avoid delamination.



Figure 1: Target and typical temperature and oxygen concentration in fire simulator tests

The experimental fire-simulator had interior dimensions of $1.0 \text{ m} \times 1.0 \text{ m} \times 1.0 \text{ m}$ (Figure 2). The CLT floor specimen was 1400 mm × 600 mm, of which 1000 mm × 600 mm was exposed. One of plate thermometers was positioned at 100 mm distance from the CLT surface directed away from the surface (similar to the plate thermometer in Test 1-4) and an additional plate thermometer was positioned 150 mm from the surface. The oxygen concentration was measured in the exhaust of the furnace. There was an inlet for air and nitrogen to control the oxygen concentration and the temperature during the cooling phase. Temperatures were measured in the adhesive bond line of the CLT in a similar fashion as was measured in the compartment fire tests discussed above. The incident radiant heat flux was determined in two different ways:

- 1. From the plate thermometer temperature and the gas temperature measured by a thermocouple near the plate thermometer.
- 2. From a water cooled heat flux gauge flush with the surface of the CLT specimen. The temperature of the cooled sensor was measured using a thermocouple. The heat flux was estimated for corresponding to different convection coefficients.

Delamination was indicated using: (1) temperature measurements in the lamellae and in at the bond line of the CLT, (2) video recordings of the specimen surface during the test, (3) charring rates determined from temperature measurements, (4) final charring depth measured after the test, and (5) measurements of the mass loss.

AN EXPERIMENTAL FIRE-SIMULATOR OF POST-FLASHOVER COMPARTMENT FIRES



Figure 2: Cross section of the test setup

3.1 Results and evaluation of the furnace tests

In this section, the damage and temperatures for compartment Test 1-4 and the intermediate scale tests are compared. A successful replication of the fire conditions of a compartment test should lead to a similar rate of charring for the same CLT product.

Figure 3 shows the average char depth measured after Test 1-4 and the estimated average char depth from the intermediate scale furnace tests after the same time (160 min). Because the furnace test was 20 min longer, the corresponding average char depth was estimated using linear interpolation of measurements of the char depth at the end of the test and the depth at the time at which the thermocouples read 300 °C. A distinct difference can be seen between the charring depth of CLT with non-delaminating adhesives and CLT with a delaminating adhesive (PU1). The average char depth determined after Test 1-4 only deviates a few millimeters from the average charring depth determined from the furnace tests. Based on a study of the uncertainty of thermocouple locations and temperature measurements, the expanded total uncertainty for the interpolated char depth is estimated to be ± 0.9 mm (95 % confidence). It should be noted that there was significant spatial (24 mm) variation in char depth across the ceiling, as measured after Test 1-4. For both intermediate scale tests of the same CLT product (PU1), the spatial variation of char depth measured was 9 mm. The spatial variation of charring depth measured in tests with non-delaminating polyurethane adhesives (PU2-A and PU2-B) was significantly less (2 mm).

Figure 4 shows the temperatures measured in the CLT of the compartment fire test together with the furnace tests. For the comparisons, only temperature measurements of the furnace tests with the same CLT product as the compartment test are shown (PU1). It can be seen that the temperature increased in the decay phase of PU1-A, despite the reducing furnace temperature. Due to delamination and the high oxygen content, the temperature could not be controlled in this phase. The temperature increase is, however, similar to that of the compartment fire. Test replicate PU1-B had the same specimen, but only had small parts of lamellae falling during the decay phase. At the end of the test, however, most of the second lamella fell-off.



Figure 3: Estimated average charring depth at 160 min using linear interpolation

The temperature jumps measured by the thermocouple indicate delamination, which is confirmed by video recordings of the surface. The times at which delamination occurs is determined from temperature measurements in the adhesive bond line and in the falling lamellae. Times at which temperature increases of 100 °C within 1 minute were measured in the fully-developed phase and times at which temperature increases of 100 °C within 5 min were measured in the decay phase, were found to correspond well with the times of delamination determined from video recordings. Box plots of the times at which this temperature increase was observed in all thermocouples are shown in Figure 5. It can be seen that the time at which delamination of the second layer was observed in Test 1-4 is within the range of delamination times observed in the furnace tests.



Figure 4: Temperature measurements at specified depths of PU1-A, PU1-B and compartment Test 1-4. In PU1-B thermocouples were installed at 50 mm instead of 70 mm depth

The incident radiant heat flux determined of the compartment and the replicating tests are shown in Figure 6. The incident heat flux was calculated using the following equation, assuming different convection coefficients of 0 W/m²K and 25 W/m²K (see Figure 6, "PT and TC h_c =0" or " h_c =25"):

$$\dot{q}_{inc} = \sigma T_{PT}^4 - \frac{h_c}{\varepsilon_{PT}} (T_g - T_{PT})$$

(1)

Where: σ is the Stefan Boltzmann constant; T_{PT} is the temperature measured by the plate thermometer; T_g is the gas temperature approximated by a thermocouple; h_c is the convection coefficient; ε_{PT} is the emissivity. The thermocouple probes measuring the plate thermometer temperatures were compliant with IEC 60584-2 requirements for class 1 K-type thermocouples, having a tolerance of 0.4 % of the measured temperature for temperatures exceeding 375 °C and ± 1.5 °C for lower temperatures. The similarity between the incident heat flux of the compartment tests and the furnace tests, suggest that the incident heat flux can also be used to control the furnace test instead of the plate thermometer temperature.



Figure 5: Time of delamination determined from thermocouple measurements



Figure 6: Incident heat flux by radiation corresponding to non-delaminating CLT with PU2 adhesive

4 DISCUSSION OF OTHER TESTING METHODS

Results of Section 3 suggest that certain relevant fire conditions of compartment fire Test 1-4 can be replicated in a furnace in which the oxygen concentration is controlled. The results of this paper indicate that this can be done by replicating the plate thermometer temperature or the incident radiant heat flux as well as the oxygen concentration of the compartment test in the furnace. In this

section, the relevance of test methods in which the oxygen concentration is not similar to that of post-flashover compartment fires is discussed.

4.1 Test setups in open environments

Recently, new test setups have been introduced in which the thermal exposure is controlled by a radiant heat source in open environments with atmospheric oxygen concentration. A test published in [3] involved a testing method developed in [8] and had significantly lower incident radiant heat fluxes than those measured during Test 1-4. If the incident heat flux alone would be enough to describe reasonable fire exposure, the test with a significantly lower incident heat flux should show lower charring rates (less damage). However, despite that the incident heat flux was more than twice as high in the compartment for approximately most of the first 30 min, it was found that the charring rates corresponding to this period of both tests were similar. Due to the high oxygen concentration, the char oxidizes significantly in the radiant panel test. Additionally, surface flaming occurs which causes additional radiation and convection to the timber, which does not occur in the fully developed phase of a compartment fire due to the low oxygen content. Therefore, results of tests in which the oxygen concentration does not correspond to a post-flashover fire condition, should be considered with care for the assessment of timber in post-flashover fires.



Figure 7: Setup of the position controlled radiant panel test [3] (left) & incident rad. heat flux of a radiant panel test, compartment test and replicating furnace test and corresponding charring rates

4.2 Standard Fire Resistance Tests

Standard fire resistance tests are performed in furnaces typically with low oxygen concentrations. This corresponds to the fully developed phase of compartment fire tests, which is often the most damaging phase of a fire for structures [10]. However, the decay phase with increasing oxygen concentration and increasing char oxidation of a compartment fire is, commonly, not represented in standard fire resistance tests. The validity of fire resistance tests is generally limited to a fully developed fire according to a recent publication [10], as the cooling phase cannot be described with standard fire resistance tests. It should be noted, that the temperatures of the ISO 834 timetemperature curve are significantly lower than temperatures measured during the fully developed phase of recent compartment tests [5]. Whether timber elements would allow for a cooling phase (including burn-out) is dependent on the interaction of the element with the fire compartment. It may be needed to require withstanding burn-out for tall buildings and buildings which the fire service cannot easily reach. If withstanding burn-out is required, it can be stated that the comparison of different materials in standard fire resistance tests can only be fair if the fire exposure (fire

7

temperature to be measured by plate thermometer in front of the exposed surface or incident radiant heat flux, and oxygen concentration) of post-flashover fires of compartments, with these materials exposed, would be similar. Recent studies indicate that this is the case if (1) a limited amount of timber is exposed and (2) delamination of exposed engineered timber and fall-off of the base layer of gypsum boards are avoided [e.g. 5]. If that is not achieved using prescriptive regulations, performance-based methods such as proposed in [11] are needed.

5 CONCLUSIONS

This paper shows that relevant fire conditions of a post-flashover fire can be successfully replicated in fire testing furnaces. Hereby, a successful replication of a compartment fire in a furnace, should lead to the same rate of damage as was observed in the replicated compartment fire test. It was shown that controlling (1) the plate thermometer temperature or the incident heat flux and (2) the oxygen concentration, so that it resembles the conditions measured in the compartment test, leads to a successful replication. Fire tests of timber elements performed in an open environment with unlimited oxygen can lead to significantly higher charring rates than observed in compartment tests due to oxidation and surface flaming.

ACKNOWLEDGMENT

This paper includes work funded and managed by the Fire Protection Research Foundation which is acknowledged by the authors. Additionally, the framework COST Action FP1404 is acknowledged for highlighting the need of the discussions of the validity of test methods.

REFERENCES

- 1. ISO 834-1 (1999), Fire resistance tests, Geneva, Switzerland.
- 2. Lineham S, Thomson D, Bartlett A, Bisby L, Hadden R (2016) Structural response of fireexposed cross-laminated timber beams under sustained loads. Fire Safety Journal 85: 23-34.
- 3. Brandon D, Maluk C, Ansell MP, Harris R, Walker P, Bisby L, Bregulla J (2015) Fire performance of metal-free timber connections. Proceedings of the Institution of Civil Engineers: Construction Materials 168(4): 173-186.
- 4. Schmid, J., Daniel, B, Werther, N., Klippel, M., Thermal exposure of wood in standard fire resistance tests, Fire Safety Journal, 2017.
- 5. Su J, Lafrance PS, Hoehler M, Bundy M (2018) Cross Laminated Timber compartment fire tests for research on fire safety challenges of Tall Wood Buildings Phase 2. FPRF/NFPA.
- Schmid J, Santomaso A, Brandon D, Wickström U, Frangi A (2017) Timber under real fire conditions – the influence of oxygen content and gas velocity on the charring behaviour. Journal of Structural Fire Engineering.
- 7. Wickström (2016) Temperature Calculation in Fire Safety Engineering. Springer Switzerland
- Maluk C, Bisby L, Krajcovic M, Torero J (2016) A Heat-Transfer Rate Inducing System (H-TRIS) Test Method. Fire Safety Journal (In press).
- 9. Brandon and Just (2018) Analysis of fire damages and limitation of fire spread. SP Report, Borås, Sweden.
- 10. Schmid J, Lange D, Sjöström J, Brandon D, Klippel M, Frangi A. The Use of Furnace tests to describe real fires of timber structures, SIF 2018.
- 11. Brandon D (2018) Engineering methods for performance-based design -Fire safety challenges of Tall Wood Buildings Phase 2 report. FPRF/NFPA.

An Empirical Model for Refrigerant Flammability **Based on Molecular Structure and Thermodynamics**

Gregory LINTERIS^{1,*}, Ian BELL², Mark McLINDEN²

¹National Institute of Standards and Technology, Gaithersburg, MD, USA ²National Institute of Standards and Technology, Boulder, CO, USA linteris@nist.gov * Corresponding Author

ABSTRACT

Screening methods for refrigerant blend flammability using metrics that can be easily calculated are of great interest to the refrigerant industry. Existing flammability metrics such as heat of combustion are not adequate for hydrofluorocarbon blends. Alternative metrics are needed that can be used to assess the flammability of refrigerant blends without requiring time-consuming experimental measurements.

In this work we study the combination of the maximum adiabatic flame temperature and the fluorine-substitution ratio as metrics for characterizing the flammability of refrigerant blends. The combination of these metrics yields an estimate of the flammability class of refrigerants (both blends and pure fluids) containing hydrofluorocarbon and hydrocarbon components. The calculations of adiabatic flame temperature are carried out with the open-source chemical kinetics software package Cantera using a mechanism available in the literature.

1. INTRODUCTION

In the search for low-global-warming-potential (GWP) replacements for refrigerants, the common performance metrics of coefficient of performance (COP) and volumetric capacity (Q_{vol}) remain important for both pure fluids and blends. Rigorously determined values for the GWP have been determined for only about 100 compounds, but an estimation method suitable for screening purposes has been developed by Kazakov et al. (2012). Using these techniques, pure compounds have been screened for their potential suitability (McLinden et al., 2017). Blends of existing or new compounds have also been identified and optimized (Bell et al., 2018). In order to complete the screening or optimization, however, a metric for flammability is also required.

2. FLAMMABILITY METRICS

The flammability of a refrigerant is classified by ANSI/ASHRAE Standard 34 (ASHRAE, 2016) and ISO Standard 817 (ISO, 2014) based on its heat of combustion, lower flammability limit, and laminar burning velocity. The classes range from "1" (fluids exhibiting "no flame propagation") to "3" ("higher flammability"-fluids with a heat of combustion greater than 19 MJ/kg or a lower flammability limit less than 0.10 kg/m³). Fluids of "lower flammability" are assigned to class "2." Class 2 fluids have a heat of combustion less than 19 MJ/kg and a lower flammability limit greater than 0.10 kg/m3. There is a further subclass "2L" for class 2 fluids that also meet the additional condition of a maximum burning velocity less than 10 cm/s. Flame propagation and the lower flammability limit are determined by the test method specified in ASTM E-681 (ASTM, 2015), with slight modifications.

Although the distinct flammability classes of 1, 2L, 2, and 3 might suggest that there is a clear boundary between "flammable" and "nonflammable" refrigerants, flammability is, in fact, a continuum (Williams, 1974). While methane, for example, is clearly flammable and carbon dioxide is clearly nonflammable, many other fluids are somewhere in between. What constitutes "flammable" is a function of the test method and test conditions. The ASTM E-681 test (as modified for refrigerants) is carried out in a 12 L glass flask with premixed fuel and air, with ignition provided by an electric spark. The criteria for "flame propagation" is that the flame move upwards and outwards from the spark, extend to the walls of the flask, and subtend an angle equal to or greater than 90°, as measured from the point of ignition. Thus, a refrigerant exhibiting a weak flame with a flame angle less than 90° in the E-681 test would

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

Bell, Ian; Linteris, Gregory; McLinden, Mark. "An Empirical Model for Refrigerant Flammability Based on Molecular Structure and Thermodynamics." Paper presented at 17th International Refrigeration and Air Conditioning Conference at Purdue, West Lafayette, IN, United States. July 9,

be considered "nonflammable" under ASHRAE Standard 34. The test conditions can also affect flame propagation. Higher temperatures will generally enhance a flame, and for highly fluorinated species, tests in humid air will enhance the flammability compared to tests in dry air (this point is discussed further in section 2.2).

Among the flammability criteria considered in the ASHRAE and ISO standards, only the heat of combustion is easily calculated, provided that enthalpies of formation are available for the refrigerant and combustion products and that the products of combustion can be determined. The flammability limits are dependent on the specifics of the test method and cannot presently be calculated from first principles, and they must be determined from experiments. The burning velocity can be calculated from first principles very accurately for some hydrocarbons (Egolfopoulos et al., 2014). However, it has only recently been calculated for pure HFCs with air (e.g., R-32 (Burgess Jr et al., 2018); C₁ and C₂ HFCs (Linteris and Babushok, 2018)), and for blends of HFCs with hydrocarbons (Linteris, 1996; Linteris et al., 1998; Linteris and Truett, 1996; Pagliaro et al., 2016a; Pagliaro et al., 2016b). For accurate predictions however, additional research is necessary to understand the important effects of flame stretch (Pagliaro and Linteris, 2016), radiation heat losses (Burrell et al., 2018), and buoyancy (Takizawa et al., 2013). Also, the calculation of the laminar burning velocity of HFC compounds in air requires a detailed kinetic mechanism, and while these are in active stages of development (Babushok and Linteris, 2017; Needham and Westmoreland, 2017; Papas et al., 2017), more work is required (Linteris and Babushok, 2018). Thus, while work is proceeding to calculate burning velocity, presently it must be measured.

An easily computed estimate of flammability would speed industry's screening for and optimization of new refrigerants and blends. General estimates of flammability have been made in previous work, including the very early estimates based on heat of combustion. The heat of combustion can be easily calculated, and has been used as a metric for flammability in refrigerant screening tests, but it is too crude a metric for marginally-flammable hydrofluorocarbon (HFC) compounds. A similar metric, the adiabatic flame temperature, has been used in previous work, and an extended version of that metric is used in the present work.

2.1 Adiabatic Flame Temperature

The adiabatic flame temperature (T_{ad}) is the temperature reached by a reacting mixture of fuel (e.g., refrigerant) and oxidizer (e.g., air) that undergoes an exothermic (heat-releasing) reaction to its most stable equilibrium products, under adiabatic conditions. Invoking the first and second laws of thermodynamics, this parameter is calculated for conditions of either constant enthalpy and constant pressure (HP) or conditions of constant internal energy and constant volume (UV). For flames in unconfined spaces, the constant HP calculation is more appropriate, while for pressure vessels, a constant UV calculation is used. Conceptually, the difference in the bond energy between the reactant and product species manifests itself as a temperature rise in the products. In the constant HP case, the PV work done by the expanding gases is also accounted for. Using the difference in the heats of formation of the products and reactants, and their gas-phase specific heats (C_p or C_v), one could, in principle, readily calculate the temperature rise in the products. The equilibrium composition is temperature dependent, however, and for combustion conditions a wide range of actual products (including radical species) is present in the final gases. Hence, in practice, automated numerical approaches (Goodwin et al., 2016; Gordon, 1996; Lutz et al., 1998) are used that are based on minimization of the Gibbs Free Energy (G) (Turns, 2000); these use standard combustion databases for the entropy and enthalpy as a function of temperature (Kee et al., 1989; McBride et al., 2002). In the present work we use Cantera (Goodwin et al. 2016), an open-source suite of object-oriented software tools for problems involving chemical kinetics, thermodynamics, and transport processes.

The adiabatic flame temperature has been used previously as a metric for flammability. The approach builds upon very early work using the heat of combustion. For example, Le Chatelier and Boudouard (1898) found that for typical hydrocarbons, flammability requires a heat of combustion above about 4.2 kJ/mol. Nonetheless, recent examination for HFCs has found that, while there is a trend of heat of combustion varying inversely with the lower flammability limit, this approach shows considerable scatter (Kazakov et al., 2012). The use of adiabatic flame temperature is an improvement over heat of combustion. Burgess and Wheeler (1911) found that for a wide range of hydrocarbons, flammability requires T_{ad} above 1600 K, and White (1925), Zabetakis (1965), and Weinberg (1971) expanded upon this work to show that the effects of differing initial temperatures, as well as most inert diluents, can be explained by their influence on the T_{ad} .

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

Bell, Ian; Linteris, Gregory; McLinden, Mark. "An Empirical Model for Refrigerant Flammability Based on Molecular Structure and Thermodynamics." Paper presented at 17th International Refrigeration and Air Conditioning Conference at Purdue, West Lafayette, IN, United States. July 9,

2.2 Fluorine Substitution Ratio

Supplementing the adiabatic flame temperature, the present method adds a metric that can account for changes to the chemical mechanism with varying composition of the reactants. The high temperature requirement for flammability of hydrocarbons results from the radical chain-branching reaction $\bullet H + O_2 \rightarrow \bullet H + \bullet OH$, which although it has a high activation energy and requires a high temperature, greatly increases the concentration of the chain-carrying radicals (•H, •O, and •OH), and is generally responsible for the rapid reaction in these systems. For systems with lowered chain-branching, for example hydrocarbons with halogens substituted for some of the hydrogens, a different, higher temperature is required for rapid reaction, and the activation energy is different. For HFCs, as the hydrogen-to-fluorine substitution (also known as fluorine loading) in the system increases, the importance of the normal hydrocarbon radical pool (•H, •O, and •OH) decreases, and the importance of fluorinated radicals increases (Babushok et al., 2012; Linteris and Babushok, 2018). This shift in the kinetic mechanism from one dominated by the typical hydrocarbon radical pool to one dominated by fluorinated-species radical attack can be captured via the molar ratio of F to H atoms in the reactants. A convenient parameter capturing this is the molar ratio of F atoms to the sum of F and H atoms F/(F + H)in the reactants (the ratio F/H has also been used, but becomes infinite for perfluorinated compounds). Typically, all sources of hydrogen in these systems react rapidly, and have the net effect of increasing hydrogen-containing radical species (e.g., •H or •OH) in the flame. Thus, added water vapor can increase the overall reaction rate by creating •H atoms and increasing the importance of the usual chain-branching reactions (Takahashi et al., 2015; Takahashi et al., 2017).

The importance of the F to H ratio to the flame reaction pathways has been shown in many works. Linteris (1994), Nyden et al. (1994), and Linteris and Gmurczyk (1995) studied HF formation in premixed and diffusion flames inhibited by HFCs and showed that the equilibrium products and flame kinetics change dramatically as F/H increases above unity (F/(F + H) > 0.5), and that premixed flame stability is markedly decreased for F/H near unity (Linteris and Gmurczyk, 1995). For methane-air and propane-air systems with added R-32, R-125, R-227ea, or R-116, premixed flames (i.e., those resulting from premixed reactants) could not be stabilized for values of F/H greater than unity (Linteris and Gmurczyk, 1995). In experimental studies of burning velocity of various pure HFC compounds in moist and dry air (Linteris, 1999), the F/H ratio was found to be a key parameter determining flame stability and burning velocity, and highlighted the importance of water vapor in the air. Similarly, in experimental studies with premixed flames of R-32, R-143, R-143a, or R-152a with air, Takizawa et al. (2005) found the burning velocity to drop off rapidly as F/(F + H) increased above 0.5. In numerical premixed flames simulations employing detailed kinetic mechanisms (Linteris and Babushok, 2018), calculated burning velocities were also found to drop rapidly as F/(F + H) approached and exceeded 0.5, where the fundamental reaction kinetics shifted from typical hydrocarbon radical pool species (•H, •O, and •OH) to fluorinated species. Hence, the parameter F/(F + H) is expected to be useful for describing the shift in kinetics and changes to flame stability for refrigerant blends.

3. CALCULATION OF ADIABATIC FLAME TEMPERATURE

For the calculation of the adiabatic flame temperature, the initial reaction mixture is given by:

$$\phi$$
 · Fuel + $\Gamma \left[O_2 + 3.76 N_2 + \frac{X_{H_2O}}{\left(\frac{1.0}{4.76}\right)(1 - X_{H_2O})} H_2O \right]$ (1)

in which ϕ is the fuel-air equivalence ratio, and has a value of <1 for lean combustion, >1 for rich combustion, and 1 for stoichiometric combustion. The fuel is given by the sum of the molar fractions x_i of its constituent refrigerants (i), and Γ is the molar oxygen requirement for stoichiometric (ϕ =1.0) combustion of one mole of fuel (refrigerant blend). The nitrogen in the air is relatively inert in the combustion reactions, although it does absorb heat because some of the heat of reaction goes to heating the nitrogen to T_{ad} .

The water vapor volume fraction in the oxidizer (air) is denoted $X_{H,O}$. The air is modelled as having an N₂:O₂ ratio of 3.76:1.0 such that the molar composition of the humid air is:

$$N_2$$
: (3.76/4.76)(1- X_{H_2O}),

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

Bell, Ian; Linteris, Gregory; McLinden, Mark. "An Empirical Model for Refrigerant Flammability Based on Molecular Structure and Thermodynamics." Paper presented at 17th International Refrigeration and Air Conditioning Conference at Purdue, West Lafayette, IN, United States. July 9,

O₂:
$$(1.0/4.76)(1-X_{H_2O})$$

H₂O: X_{H_2O}

These conditions correspond to a dry air O_2 volume fraction of 0.21. The volume fraction of water vapor $X_{H,O}$ is selected to be 0.014, corresponding to air at 23 °C and 50 % relative humidity in accordance with the flammability tests specified in ASHRAE Standard 34.

 Γ is approximated as the mole-fraction-weighted average of the individual stoichiometric air requirements for each constituent v_i ,

$$\Gamma = \sum_{i=1}^{N_{comp}} x_i \gamma_i \tag{2}$$

The value of γ_i , for a given pure compound is determined by conducting an equilibrium calculation (constant enthalpy and pressure) for a range of fuel-air ratios, and then examining the equilibrium products. From these, it is possible to write a stoichiometric reaction with appropriate coefficients. For example, for R-134a (C₂H₂F₄), the equilibrium products indicate an approximate stoichiometric reaction given by:

$$C_{2}H_{2}F_{4} + 3/2 (O_{2} + 3.76 N_{2}) \rightarrow CO_{2} + 2HF + COF_{2} + 3/2(3.76) N_{2}$$
(3)

such that $y_i = 1.5$ for R-134a. Similarly, for R-125, the stoichiometric reaction is:

$$C_2HF_5 + 1 (O_2 + 3.76 N_2) \rightarrow \frac{1}{2} CO_2 + COF_2 + HF + \frac{1}{2} CF_4 + 3.76 N_2$$
(4)

and $y_i = 1.0$ for R-125. The values of y_i are listed in Table 1 for stoichiometric combustion. For hydrocarbons, $y_i = n_c + n_u/4$, in which n_c is the number of carbon atoms in the fuel molecule, and n_u is the number of hydrogen atoms.

The value of Γ given by Eq. (2) is an initial estimate, as it can vary with the reactants. For example, while the combustion of R-134a with air yields COF₂ as one of the products, a mixture of R-134a and propane may not. This is because some of the hydrogen atoms from the propane (that would have formed H₂O) can react with the fluorine atoms from the R-134a to give HF in preference to COF2. The final set of combustion products for the given reactants is one of the key outputs from the Cantera equilibrium calculations.

For the final determination of T_{ad} for a given "fuel", the equilibrium calculation is performed again for the estimated value of Γ , over a range of $0.5 \le \phi \le 2.0$ in Eq. [1], and the peak value of T_{ad} is selected. This value of ϕ for the peak value is typically within a few percent of $\phi = 1.0$, validating the approximation of Γ as a mole-fraction-weighted combination of the individual γ_i . Note also that while the estimation of γ_i and Γ are based on dry air, the final estimates of F/(F+H) and T_{ad} are based on moist air as specified in ASHRAE Standard 34. Since the final T_{ad} are determined as the peak value over a range of ϕ , any variation in γ_i and Γ due to the presence of the moisture in the air is accounted for in the final value of peak T_{ad} .

As Table 1 shows, the flammable HFCs have T_{ad} close to those of hydrocarbons, and as the fluorine loading increases, T_{ad} decreases mildly and then more strongly as F/(F + H) increases above 0.5. For T_{ad} near 2000 K, values of F/(F + H) of about 0.66 denote the boundary between flammable and non-flammable (for these compounds).

It is worth commenting on values of T_{ad} as high as 1928 K among the "nonflammable" refrigerants listed in Table 1; indeed, what does a "flame temperature" mean for a nonflammable fluid? This has to do with reaction kinetics (i.e., the rate of chemical reaction). The concept of the adiabatic flame temperature is based on equilibrium thermodynamics for an adiabatic system (infinite time, zero heat losses). Hence, while the equilibrium products of R-134a in the presence of air are CO₂, HF, and COF₂, at T_{ad} = 1940 K, as indicated by Eq.(3), this state is typically not reached under normal conditions because the reaction kinetics are too slow; i.e., R-134a in air is "nonflammable" because the reaction rates are too slow.

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

Refrigerant (i)	Formula	γi	F /(F + H)	T _{ad} (K)	Π		
A3, Flammable							
R-290	C_3H_8	5	0.000	2262	100		
R-1270	C_3H_6	4.5	0.000	2338	100		
R-600	C_4H_{10}	6.5	0.000	2265	100		
R-600a	iso-C ₄ H ₁₀	6.5	0.000	2261	100		
R-601	$C_{5}H_{12}$	8	0.000	2265	100		
R-601a	iso-C ₅ H ₁₂	8	0.000	2265	100		
A2, Flammable							
R-152a	CH ₃ -CHF ₂	2.5	0.316	2208	55.8		
A2L, Mildly flan	nmable_						
R-32	CH_2F_2	1	0.484	2190	35.6		
R-143a	CH ₃ -CF ₃	1.5	0.479	2100	28.3		
R-1234yf	CH ₂ =CFCF ₃	2.5	0.633	2046	6.9		
R-1234ze(E)	CFH=CHCF ₃	2.5	0.633	2046	6.9		
A1, Nonflammable							
R-134a	CH ₂ F-CF ₃	1.5	0.645	1928	-8.1		
R-125	CHF ₂ -CF ₃	1	0.815	1788	-34.7		
R-227ea	C ₃ HF ₇	1.5	0.853	1791	-35.5		
R-218	C_3F_8	1	0.983	1581	-62.4		
R-744	CO_2	n.a.	n.a.	n.a.	n.a.		

Table 1: Properties of selected pure refrigerants in ASHRAE Standard 34. The adiabatic flame temperature T_{ad} and F/(F + H) are based upon a volume fraction of H₂O in the air of 0.014, and the initial temperature is 296.15 K.

4. RESULTS—ESTIMATION OF FLAMMABILITY

The present approach calculates, for a given set of premixed reactants, the adiabatic flame temperature $T_{\rm ad}$ and the fluorine loading, expressed as F/(F + H). The existing ASHRAE Standard 34 flammability designations are then used to produce a map of the flammability as a function of T_{ad} and F/(F + H). The dataset included in this comparison are mixtures of the pure compounds listed in Table 1. The pure compounds include: flammable (class 3) hydrocarbons; flammable (class 2) R-152a; mildly flammable (class 2L) HFCs and hydrofluoroolefins (HFOs), and the nonflammable (class 1) HFCs and CO₂.

Figure 1(a) shows the T_{ad} and F/(F + H) values for each of the refrigerants listed in ASHRAE Standard 34 which contain the compounds in Table 1 at various compositions. The pure compounds are shown as closed circles, and the blends, open circles. The dotted lines were manually drawn to separate the different flammability regions. The origin of those lines is at $T_{ad} = 1600$ K and F/(F + H) = 0. This point corresponds to hydrocarbons at their flammability limit, which is appropriate, since as described above, 1600 K is an empirically observed flammability limit for hydrocarbon flames over a wide range of hydrocarbons, inert diluents, and flame configurations.

We define a "flammability index" Π based on the angle that a point in [F/(F + H), T_{ad}] coordinates makes with the origin point at (0, 1600 K), in which the temperature difference (T_{ad} – 1600 K) is normalized by a reasonable upperlimit T_{ad} , which we select to be 2500 K:

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

$$\Pi = \arctan \left\{ \left[\frac{T_{ad} - 1600}{(2500 - 1600)} \right], \left[\frac{F}{(F+H)} \right] \right\} \cdot \left(\frac{180}{\pi} \right), \tag{5}$$

where $\arctan 2(y, x)$ gives the four-quadrant arctangent angle in the domain $[-\pi, \pi]$, and the angle is then shifted to the domain [-180°, 180°]. A normalized flammability index Π is then defined by

$$\bar{\Pi} = \frac{\Pi - \Pi_{1,2L}}{90 - \Pi_{1,2L}} \cdot 100, \tag{6}$$

with $\Pi_{1,2L}$ is equal to 34.78. This results in a normalized flammability index which is zero at the 1/2L boundary and 100 for the highly flammable hydrocarbons. Values less than zero indicate that the mixture is probably nonflammable according to the ASHRAE Standard 34 criteria.

The blends shown in Figure 1(a) are calculated for the compositions listed in ASHRAE Standard 34-the so-called "nominal" compositions. In fact, the flammability classification of blends is based on a slightly different composition. Standard 34 specifies a composition tolerance for each component in a blend to account for variations in blending by the refrigerant producer. The "most flammable" composition resulting from a blend at the extreme of the tolerances is denoted as the "worst case of formulation for flammability" (WCF). This WCF composition is then analyzed for possible composition shifts resulting from defined leakage scenarios, yielding a "worst case of fractionation for flammability" (WCFF). The flammability testing of blends is then carried out at the WCF and WCFF compositions. Sometimes, for multicomponent blends, multiple possible WCF and WCFF compositions are tested to determine the worst case. We extracted the flammability test data from the applications to the Standard 34 committee for new refrigerant designations. The flammability test data in each application at those WCF and WCFF compositions are plotted as a function of T_{ad} and F/(F + H) in Fig. 1(b). This was done for the applications submitted for R-444A through R-462A (corresponding to applications submitted from 2012 to 2016). As per the ASTM E-681 test method, we classify the flammability of the mixture based on the maximum flame angle observed in the test; maximum flame angles $> 90^{\circ}$ are "flammable" and those $< 90^{\circ}$ are "nonflammable."

The overall results from the WCF and WCFF compositions are substantially the same as those from the nominal compositions, and the resulting boundary between class 1 and class 2L is essentially the same. There is more scatter seen with the WCF and WCFF compositions, and we drew the boundary to be conservative. The flammability data from the applications is particularly valuable in defining the 1/2L boundary because many of the blends were formulated to be close to the boundary. For example, some blends include a hydrocarbon to improve oil solubility characteristics; a higher concentration of hydrocarbon would be advantageous for the oil solubility, and it would be set as high as possible without crossing into the flammable region.

As indicated, the parameters adopted provide a reasonable estimate of the flammability for these compounds evaluated via the ASHRAE criteria. Moreover, the slope of the line between that data point and the origin (F/(F + H) = 0, F/(F + H)) $T_{\rm ad} = 1600$ K), *i.e.*, the flammability index defined above, can provide a relative measure of the flammability of a refrigerant. Such an approach can provide a first estimate of the flammability of a new blend, which increases as the slope increases. This method of flammability estimation has been used in the screening work described in Bell et al. (2018).

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018





Figure 1: ASHRAE Standard 34 flammability classification (A1, A2L, A2, or A3) shown as a function of T_{ad} and F/(F + H) for blends (open circles) and pure compounds (closed circles). The pure compounds are listed in Table 1. (a) ASHRAE Standard 34 flammability rating (1: blue, 2L: green, 2: orange, 3: red); (b) inset with measured flame angles for WCF and WCFF compositions from ASHRAE 34 applications (where available).

5. CONCLUSIONS

The present empirical estimate for the flammability of hydrofluorocarbon refrigerants is based on historical work showing the influence of the adiabatic flame temperature on hydrocarbon flammability limits, and the influence of fluorine-containing species on the combustion chemistry of hydrocarbon flames. These two effects are described by T_{ad} and the fluorine loading in the refrigerant, expressed as the molar ratio of F to (F + H) in the reactants. The adiabatic flame temperature is determined, using the Cantera open-source kinetics solver, as the peak value over a range of fuelair ratios for the given refrigerant (pure compound or blend) in moist air at 296 K.

Using the ASHRAE flammability designations, a map of T_{ad} vs. F/(F + H) is developed showing the regions of these parameters corresponding the flammable (A2 and A3), mildly flammable (A2L), and nonflammable (A1) regions. A metric is developed based on these parameters to rank the flammability (according to the ASHRAE Standard 34 criteria) for an arbitrary blend of constituents for which the chemical composition is known, and the adiabatic flame temperature can be calculated. The agreement with the existing data in the ASHRAE Standard 34 refrigerant database is good.

It should be noted that the present flammability ranking system is based on the ASHRAE Standard 34 test methods. For flammability configurations that require a different (i.e., more or less stringent) flammability requirement, the present method may not be accurate. Also, for application to new compounds or blends, it is required that the new constituent compounds for which T_{ad} and F/(F + H) are to be calculated are chemically similar to the compounds used to develop the model.

REFERENCES

ASHRAE. (2016). ANSI/ASHRAE Standard 34-2016, Designation and Safety Classification of Refrigerants. Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

ASTM. (2015). ASTM E681-09: Standard test method for concentration limits of flammability of chemicals (vapors and gases). ASTM Fire Standards. West Conshohocken, PA: American Society of Testing and Materials.

Babushok, V. I., & Linteris, G. T. (2017). Kinetic mechanism of 2,3,3,3-tetrafluoropropene (HFO-1234yf) combustion. J Fluorine Chem, 201, 15-18.

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

Bell, Ian; Linteris, Gregory; McLinden, Mark. "An Empirical Model for Refrigerant Flammability Based on Molecular Structure and Thermodynamics." Paper presented at 17th International Refrigeration and Air Conditioning Conference at Purdue, West Lafayette, IN, United States. July 9,

- Babushok, V. I., Linteris, G. T., & Meier, O. (2012). Combustion properties of halogenated fire suppressants. Combust Flame, 159, 3569-3575.
- Bell, I., Domanski, P. A., Linteris, G. T., & McLinden, M. O. (2018). Evaluation of binary and ternary refrigerant blends as replacements for R134a in an air-conditioning system. In 17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018 West Lafayette, IA: Purdue University.
- Burgess Jr, D. R., Manion, J. A., Burrell, R., Babushok, V. I., Hegetschweiler, M. J., & Linteris, G. T. (2018). Development and validation of a mechanism for flame propagation in R-32/air mixtures. 2018 Eastern States Section Meeting of the Combustion Institute State College, PA: The Combustion Institute
- Burgess, M. J., & Wheeler, R. V. (1911). The lower limit of inflammation of mixtures of the paraffin hydrocarbons with air. Journal of the Chemical Society, 99, 2013-2030.
- Burrell, R., Pagliaro, J. L., & Linteris, G. T. (2018). Effects of stretch and thermal radiation on difluoromethane-air burning velocity measurements in constant volume spherically expanding flames. Proc. Combust. Inst., 37, submitted.
- Egolfopoulos, F. N., Hansen, N., Ju, Y., Kohse-Höinghaus, K., Law, C. K., & Qi, F. (2014). Advances and challenges in laminar flame experiments and implications for combustion chemistry. Prog. Energy Combust. Sci., 43, 36-67
- Goodwin, D. G., Moffat, H. K., & Speth, R. L. (2016). Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes. <u>http://www.cantera.org</u>, Version 2.1.1. Pasedena, CA: California Institute of Technology Retrieved from http://www.cantera.org.
- Gordon, S., McBride, B.J. (1996). Computer program for calculation of complex chemical equilibrium compositions and applications. (NASA Reference Publication 1311). Cleveland, OH: NASA Glenn Research Center,.
- ISO. (2014). ISO/DIS 817, Refrigerants—Designation and safety classification. Currently in final draft International Standard stage. Geneva, Switzerland: International Organization for Standardization.
- Kee, R. J., Rupley, F. M., & Miller, J. A. (1989). CHEMKIN-II: A fortran chemical kinetics package for the analysis of gas phase chemical kinetics. (SAND89-8009B). Livermore, CA, USA.

Le Chatelier, H. L., & Boudouard, O. (1898). Limits of inflammability of combustible vapors. Comptes Rendus(126).

- Linteris, G. T. (1994). Acid Gas-Production in Inhibited Flames. In A. Miziolek & W. Tsang (Eds.), Halon Replacements: Technology and Science (225-242). Washington, DC: American Chemical Society.
- Linteris, G. T. (1996). Numerically Predicted Structure and Burning Velocity of Premixed CO-Ar-O₂-H₂ Flames Inhibited by CF₃H. Combust. Flame, 107(1-2), 72-84.
- Linteris, G. T. (1999). [Unpublished data.].
- Linteris, G. T., & Babushok, V. I. (2018). Numerically-Predicted Burning Velocities C1 and C2 Hydrofluorocarbon Refrigerants with Air. In 17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018 West Lafayette, IA: Purdue University.
- Linteris, G. T., Burgess, D. R., Babushok, V., Zachariah, M., Tsang, W., & Westmoreland, P. (1998). Inhibition of premixed methane-air flames by fluoroethanes and fluoropropanes. Combust. Flame, 113(1-2), 164-180.
- Linteris, G. T., & Gmurczyk, G. W. (1995). Prediction of HF formation during suppression. In R.G.Gann (Ed.), Fire suppression system performance of alternative agents in aircraft engine and dry bay laboratory simulations (201-318). Gaithersburg, MD: National Institute of Standards and Technology.
- Linteris, G. T., & Truett, L. (1996). Inhibition of premixed methane-air flames by fluoromethanes. Combust. Flame, 105(1-2), 15-27.
- Lutz, A. E., Rupley, F. M., Kee, R. J., Reynolds, W. C., & Meeks, E. (1998). EQUIL: A CHEMKIN implementation of STANJAN for computing chemical equilibria. Reaction Design, Inc., 6500 Dublin Boulevard, Dublin, CA 94568. Software and manual authorized by Ellen Meeks and Fran Rupley.
- McBride, B. J., Zehe, M. J., & Gordon, S. (2002). NASA Glenn coefficients for calculating thermodynamic properties of individual species. (NASA TP-2002-211556). Cleveland, OH: NASA Glenn Research Center.
- McLinden, M. O., Brown, J. S., Brignoli, R., Kazakov, A. F., & Domanski, P. A. (2017). Limited options for lowglobal-warming-potential refrigerants. Nat. Commun., 8, 9.
- Needham, C. D., & Westmoreland, P. R. (2017). Combustion and flammability chemistry for the refrigerant HFO-1234vf (2,3,3,3-tetrafluroropropene). Combust. Flame, 184, 176-185.
- Nyden, M. R., Linteris, G. T., Burgess, D. R. F., Westmoreland, P. R., Tsang, W., & Zachariah, M. R. (1994). Evaluation of Alternative In-Flight Fire Suppressants for Full-Scale Testing in Simulated Aircraft Engine Nacelles and Dry Bays. (NIST SP 861). Gaithersburg MD.
- Pagliaro, J. L., Bouvet, N., & Linteris, G. T. (2016a). Premixed flame inhibition by CF3Br and C3H2F3Br (2-BTP). Combust. Flame, 169, 272-286.

17th International Refrigeration and Air Conditioning Conference at Purdue, July 9-12, 2018

Bell, Ian; Linteris, Gregory; McLinden, Mark. "An Empirical Model for Refrigerant Flammability Based on Molecular Structure and Thermodynamics." Paper presented at 17th International Refrigeration and Air Conditioning Conference at Purdue, West Lafayette, IN, United States. July 9,

2018 - July 12, 2018

- Pagliaro, J. L., & Linteris, G. T. (2016). Burning velocities of marginally flammable refrigerant-air mixtures. 2016 Eastern States Section Meeting of the Combustion Institute Princeton, NJ: The Combustion Institute
- Pagliaro, J. L., Linteris, G. T., & Babushok, V. I. (2016b). Premixed flame inhibition by C2HF3Cl2 and C2HF5. Combust. Flame, 163, 54-65.
- Papas, P., Zhang, S., Kim, W., Zeppieri, S. P., Colket, M. B., & Verma, P. (2017). Laminar flame speeds of 2,3,3,3tetrafluoropropene mixtures. Proc. Combust. Inst., 36, 1-10.
- Takahashi, F., Katta, V. R., Linteris, G. T., & Babushok, V. I. (2015). Combustion inhibition and enhancement of cup-burner flames by CF3Br, C2HF5, C2HF3Cl2, and C3H2F3Br. Proc. Combust. Inst., 35, 2741-2748.
- Takahashi, F., Katta, V. R., Linteris, G. T., & Babushok, V. I. (2017). A computational study of extinguishment and enhancement of propane cup-burner flames by halon and alternative agents. Fire Saf. J., 91, 688-694.
- Takizawa, K., Takagi, S., Tokuhashi, K., Kondo, S., Mamiya, M., & Nagai, H. (2013). Assessment of Burning Velocity Test Methods for Mildly Flammable Refrigerants, Part 1: Closed-Vessel Method. ASHRAE Trans., 119(2), 243-254.
- Takizawa, K., Takahashi, A., Tokuhashi, K., Kondo, S., & Sekiya, A. (2005). Burning velocity measurement of fluorinated compounds by the spherical-vessel method. Combust. Flame, 141(3), 298-307.

Turns, S. R. (2000). An introduction to combustion : concepts and applications. Boston: WCB/McGraw-Hill.

Weinberg, F. J. (1971). COMBUSTION TEMPERATURES - FUTURE. Nature, 233(5317), 239-&.

White, A. G. (1925). Limits for the propagation of flame in inflammable gas-air mixtures. Part III. The effect of temperature on the limits. Journal of the Chemical Society, 127, 672-684.

Williams, F. A. (1974). A unified view of fire suppression. J. Fire Flamma., 5, 54-63.

Zabetakis, M. G. (1965). Flammability characteristics of combustible gases and vapors. (627). Washington, D.C.

ACKNOWLEDGEMENT

This work was supported by the U.S. Department of Defense, Strategic Environmental Research and Development Program (SERDP), project WP-2740. We thank Angela McFarlin and Brian Cox of ASHRAE for providing us access to the Standard 34 applications. The mechanisms and tools used in the present work were developed through concurrent support from the Buildings Technologies Office of the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy under contract no. DE-EE0007615 with Antonio Bouza serving as Project Manager.

The Industrial Ontologies Foundry Proof-of-Concept Project

Boonserm (Serm) Kulvatunyou¹, Evan Wallace¹, Dimitris Kiritsis², Barry Smith³, Chris Will⁴

¹Systems Integration Division, National Institute of Standards and Technologies *{serm, evan.wallace*}@nist.gov ²EPFL, Switzerland *dimitris.kiritsis@epfl.ch* ³State University of New York, Buffalo *phismith@buffalo.edu* ⁴Dassault Systemes *chris.will@3ds.com*

Abstract. The current industrial revolution is said to be driven by the digitization that exploits connected information across all aspects of manufacturing. Standards have been recognized as an important enabler. Ontology-based information standard may provide benefits not offered by current information standards. Although there have been ontologies developed in the industrial manufacturing domain, they have been fragmented and inconsistent, and little has received a standard status. With successes in developing coherent ontologies in the biological, biomedical, and financial domains, an effort called Industrial Ontologies Foundry (IOF) has been formed to pursue the same goal for the industrial manufacturing domain. However, developing a coherent ontology covering the entire industrial manufacturing domain has been known to be a mountainous challenge because of the multidisciplinary nature of manufacturing. To manage the scope and expectations, the IOF community kicked-off its effort with a proof-of-concept (POC) project. This paper describes the developments within the project. It also provides a brief update on the IOF organizational set up.

Keywords: Smart manufacturing, Industrie 4.0, Ontology, IOF, Industrial Ontologies Foundry

1 Introduction

The current industrial revolution is said to be driven by the digitization that exploits connected information across all aspects of manufacturing [1]. Standards have been recognized as an important enabler; meanwhile, ontology is considered as the next generation standard for connected information. Although there have been ontologies developed in the industrial manufacturing domain, they have been disparately developed

with inconsistent principles and viewpoints. Hence, existing industrial ontologies are incoherent and not suitable for the connected information goal.

With successes in developing coherent ontologies in the biological, biomedical, and financial domains [2-4], an effort called Industrial Ontologies Foundry (IOF) has been formed to pursue the same goal for the industrial manufacturing domain [5]. Modern manufacturing, particularly with today's complex cyber-physical products and materials, however, requires diverse disciplines of engineering, information technology, and management. This nature makes the scoping and development of coherent ontology for the industrial manufacturing domain a mountainous challenge. To manage scope and expectations, the IOF community has devised a proof-of-concept (POC) project with the aim to prove the viability and values of the endeavor. This paper describes the developments within the POC project.

The paper first describes the IOF formation, organizational structure, and aims of its subgroups. It then describes the POC process, discusses current results, and finally concludes with future plans.

2 IOF Formation and Organization

The first IOF workshop was organized in December 2016 at the National Institute of Standards and Technology (NIST), Gaithersburg, USA [6]. Thereafter, the community has had weekly conference calls and yearly workshops [7, 8].

Following the workshop in 2017, the IOF charter has been drafted and became available on its web site as a community draft [9]. One of the most important messages from the charter that makes IOF unique from other standard organizations that have published engineering ontologies is the intention of the IOF for its ontologies to be freely open. Although the current charter stops short at indicating a particular intellectual property licensing agreement, the community recognizes the need for such an encompassing ontology to be freely reusable, so that the corresponding information can be truly connected. The intention of the community has always been gravitating toward one of the royalty free licenses (e.g., variations of the creative commons licenses [10], which only require some recognitions).

The charter also outlines goals that include not only publishing the freely available ontologies, but also providing principles, guidelines, and governance processes such that a suite of ontology modules can grow in an interoperable fashion. One of the components to enable this is the organizational structure of the IOF community.

The community has devised three kinds of committees: a governance board (GB), a technical oversight board (TOB), and working groups (WGs). There is one GB, one TOB, and as many WGs as deemed necessary by the community. To ensure interoperability, these boards have overlapping personals as shown in Fig. 1, where each circle represents the membership of each board.



Fig. 1. IOF organizational structure

The primary role of the GB is to maintain the health and effective operation of the IOF organization. It sets the overall policy and manages legal aspects of the business. The other important role for the GB is to resolve conflicts unresolvable by the TOB.

TOB members are responsible for setting ontology principles and design guidelines used across the WGs. They have an important role to ensure that modules of the IOF ontologies developed by each WG are interoperable and consistent.

Each WG develops an ontology or a suite of ontologies of the IOF ontologies vetted by the TOB. Some WGs may be responsible for developing or adapting cross-cutting, domain independent ontologies such as for time or units of measurement. Fig. 2 shows types of ontologies anticipated within the IOF ontologies. A WGs may exist for the foundation ontology, each of bubbles, or a group of bubbles in the figure. The top two layers reflect specializations of the IOF ontologies for a particular use that may be private or licensed (developing such specializations is considered out of scope for the IOF).



Fig. 2. Architecture of the IOF ontologies [9]

We can observe from Fig. 2 that it is a challenge to establish the scope of each WG, particularly at the Domain Specific Reference Ontologies level as information about a

manufactured product overlaps across its life cycle phases and operational areas involved in development and manufacturing of the product. On the other hand, starting from bottom up looking at the choices of foundation ontology (e.g., BFO [11], DOLCE [Error! Reference source not found.]) and Domain Independent Reference Ontologies carry a risk of developing something don't meet the requirements of the domain. To manage such risk and challenge, the IOF community has agreed to start with a proofof-concept (POC) project, which set the scope based on a consensus rather than the life cycle or operation areas of manufacturing and devised both bottom-up and top-down groups to test the feasibility and values of IOF ontologies. The next section describes this ongoing activity.

3 IOF POC

The IOF proof-of-concept (POC) project was intended to test the feasibility of IOF goals. Therefore, the objectives of the POC included not only producing a small, initial ontology, but also testing the organizational structure (described above) and producing and testing drafts principles and guidelines.

To set the scope, the POC started by asking for most interested manufacturing-related terms from the community. To set a very low bar, textual definitions were optional. Terms submitted should have a use case behind them so that a proof-of-value (POV) can be performed once the ontology is available. Each submission could include up to 50 terms in virtually any form of structured or unstructured file format. After collecting all the submissions, 20 terms were to be identified based on the frequencies of matches across the submissions. Each term has a (synonym) set of closely-matched terms; therefore, we call the output of this step the top-20 sets.

Getting the top-20 sets

At the time of this writing, 23 submissions were received, some of which went over the 50-term limit. Submission topic areas included product life cycle management, general manufacturing, manufacturing process, material (in the sense of material science), supply chain management, logistics, shop floor automation, manufacturing resources, production system engineering and analysis, and additive manufacturing.

To identify the top-20 sets, first submissions were transcribed into a web-based collaborative ontology editing tool [Error! Reference source not found.] using a Simple Knowledge Organization System (SKOS) representation [14]. In a few cases, submissions were provided in an Web Ontology Language (OWL) [15] format. The tool was able to import these, then SKOS concept assertions were manually added to their class URIs. In this way, all terms could be mapped using the SKOS *closeMatch* relationship.

This led to the second step, creating a rough mapping. The objective of this mapping was not to do semantic alignment, but rather to set a scope for the POC. Hence, mappers were instructed to be quite liberal with their mapping and to not be concerned with the concept hierarchy. One relative rich submission was copied as a starting point for a canonical terminology (i.e. the IOFPOC skos:ConceptScheme) for mapping. The mapping task was divided among four individuals who performed the mapping consecutively, and the target for completion was a few weeks.

Even with the rough mapping in mind, it was not a trivial work for the last individual because the canonical grew as unmapped terms were added to it. The last individual reported that the search functionality in the tool became very handy. At the finish, the canonical had over 600 terms. Table 1 below shows the resulting top-20 sets. Due to space limitation, only the terms and counts are shown. Table 2 shows definitions provided for the term set with the most matches. Complete result will be available on the IOF web site [5].

	Term Set	Count
1	Product, Physical Product, Product Material, Manufactured Product	13
2	Material, Material Object, Engineered Material	12
3	Manufacturing Machine, Processor, Machinery, Machine Tool, Ma-	10
	chine, Mechanism, Machine, Workstation	
4	Tool, Tools, Tooling, Manufacturing Tool	9
5	Assembly, Part, Composition	8
6	Part, Physical Part, Sub Assembly, Product Component, Component	8
7	Process, Transformation	8
8	Supplier, Supplier Provider or Vendor, Material Supplier	8
9	Transportation Process, Move, Movement, Transfer, Transport, Act	8
	of Transportation, Transport	
10	Quality, Indicator, General KPI, Quality, Engineering Quality, Phys-	7
	ical Quality	
11	Requirement, Requirement Specification, Control	7
12	Assembly Process, Assembly Operation, Technological Pair Posi-	6
	tioning, Joining, Act of Assembly	
13	Customer, Business Customer	6
14	Feature, Materials Property, CAD Model Feature	6
15	Process Plan, Work Instructions, Manufacturing Method, Operation,	6
	Process Plan, Manufacturing Process Plan	
16	Resource	6
17	Task, Activity, Operation	6
18	Design, Design Process	5
19	Equipment, Machinery	5
2.0	Fixture Work Holder	5

Table 1. The top-20 sets of closely matched terms

Kiritsis, Dimitris; Kulvatunyou, Boonserm; Smith, Barry; Wallace, Evan; Will, Chris. "The Industrial Ontologies Foundry Proof-of-Concept Project." Paper presented at APMS 2018 International Conference Advances in Production Management Systems (APMS 2018), Seoul, Republic of Korea. August 25, 2018 - August 30, 2018.
Term	Textual Description	
Physical Product	Subclass of spatial region (derived from the class axiom)	
Product	Product (for manufacturing industry) is a Material Object,	
	manufactured to satisfy a need of the market (e.g. to be sold	
	in order to provide profit and support customers by covering	
	their needs).	
Product	the output of a manufacturing process	
Product	a material entity or service that is developed to be sold	
Product	This is a tangible object manufactured to satisfy a need of the	
	market. For the specific mould maker, common products are:	
	moulds, dies, and high precision parts	
Product	material and/or service sold to others. Note that in manufac-	
	turing enterprises 'product' often refers to a product type or	
	class, which in supply chains may be differentiated by pack-	
	aging, but 'product' may also refer to a product instance.	
Product	A goods, idea, method, information, object or service created	
	as a result of a process and serves a need or satisfies a want. It	
	has a combination of tangible and intangible attributes (bene-	
	fits, features, functions, uses) that a seller offers a buyer for	
	purchase.	
Product Material	A material entity produced by man or machine, including raw	
	material, parts, semi-finished product, and finished product	
Manufactured	a product that is created via a manufacturing process.	
Product		
Product	A product is the subject of the activity.	
Product	desired output or by-product of the processes of an enterprise.	
	Note 1 to entry: A product can be an intermediate product, end	
	product, or finished goods from a business perspective.	
	[SOURCE: IEC 62264-1:2013-01, 3.1.27]	
Product	No definition provided.	

Table 2. Textual descriptions for the most frequent set

Discussion

It can be seen from Table 1 that judging only by the label, the terms in each set are quite semantically close to each other. However, Table 2 demonstrates that each submission in the set gave a variety of definitions; and a few distinct notions (or concepts) may be refactored from the set. All but three of the submissions provided textual definitions; though two of those three provided subsumption hierarchies. Due to the richness of the semantics provided in the submission, the mapping result frequently yielded closely related notions within each set. That is, we observed that notions in the same set were often either close to each other in a subsumption hierarchy, an action or process of another notion in the set, or a mereologically related notion to another notion in the set. There are however other complex cases where similar terms are used for overlapping notions and similar terms are used for subtly distinct notions. An example of the former case can be observed in Table 2 where 'product' was used exclusively for physical object, both physical object and service, only desired (or designed) output, and both desired and undesired (by-product) output. An example of the latter case is 'material'. In one notion, it is a chemical composition, while in the other, it is a part, assembled component, or raw substance (e.g., metal powder) supplied to the manufacturing activity.

Next step

In the next step, the IOF community is creating a WG to formalize these top-20 sets. Two kinds of WGs are in the plan, top-down and bottom-up. At the time of this writing, a top-down WG has been formed. Its approach is to classify the notions in the top-20 set using the Basic Formal Ontology (BFO) [11] and conversely to see if BFO needs changes.

The formations of the bottom-up WGs are driven by use cases. At this time, the TOB is soliciting use cases from the community. With use case information, the TOB will cluster use cases together to form bottom-up WGs.

Both types of WGs will allow the submitters an opportunity to complete their submissions with detail textual definitions before the WGs proceed. It is expected that each bottom-up WGs would consider notions beyond the top-20 set. They will take the focal point at ensuring that the harmonized definitions or ontology in each bottom-up domain board satisfy their functional requirements and viewpoints. It is unclear at this point whether each bottom-up WG will produce a more formal definition (or model) such as class diagram or even an ontology. They will however perform gap analysis between their notions and the output from the top-down group by classifying their notions with respect to the top-down view. It is yet to be determined whether the gap analysis will be performed concurrently across all the bottom-up groups or sequentially. The community seems excited about what is coming before them.

4 Conclusion and Future Plans

The IOF is a growing community. After a round of inactivity elimination, over 60 participants have registered and are actively involved with representative across all populous continents except Africa. Both the contingent governance (5 seats) and technical oversight (12 seats) boards consist of diverse representatives from private companies, research institutes, academia, and standard development organizations. Both boards have a lot of deliverables lying ahead. The governance board should set up membership policies, development infrastructure, and intellectual property policy. The technical oversight board should draft ontology design rules (e.g., naming convention, minimal ontological commitment, URIs, versioning) and design principles (e.g., modularity, interaction with existing standards). It is anticipated that the proof-of-concept

(POC) project, which is still ongoing, will be the platform for developing these documents. At the present time there is no fee to participate in the IOF; however, the governance board is tasked with developing a business model to sustain the community given a successful POC. Interested individual are invited to submit a request for participation on the IOF web site.

Acknowledgement and a Disclaimer

The authors wish to thank all the Industrial Ontologies Foundry (IOF) members who have contributed their expertise and content to the founding of IOF and the POC project. Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST

Reference

- 1. The Economist. April 21 issue on the Third Industrial Revolution (2012).
- 2. Open Biological and Biomedical Ontology Foundry Web Site http://obofoundry.org, last access 2018/07/11.
- 3. Financial Industry Business Ontology press release http://www.edmcouncil.org/downloads/20171026_FIBO_Release.pdf, last access 2018/07/11.
- 4. FIBO spec page https://www.edmcouncil.org/financialbusiness, last access 2018/07/11. 5. Industrial Ontologies Foundry (IOF) Website -
- https://sites.google.com/view/industrialontologies/home, last access 2018/07/11. Kulvatunyou, B. and Morris, K.: Working Towards an Industrial Ontology Foundry to 6.
- Facilitate Interoperability http://blog.mesa.org/2017/03/working-towards-industrialontology.html, last access 2018/07/11.
- 7. Ivezic, N., et al.: 2017 NIST/OAGi Workshop: Enabling Composable Service-Oriented Manufacturing Systems. NIST Advanced Manufacturing Series 100-15 (2018).
- 8. 2018 NIST/OAGi Workshop Enabling Composable Service-Oriented Manufacturing Systems - https://www.nist.gov/news-events/events/2018/04/2018-nistoagi-workshopenabling-composable-service-oriented-manufacturing, last access 2018/07/11.
- 9. IOF Charter https://sites.google.com/view/industrialontologies/about/charter, last accessed 2018/07/11
- 10. Creative Common Licenses https://creativecommons.org/licenses/, last accessed 2018/07/11.
- 11. Smith, B., et al. Basic Formal Ontology Home Page http://ontology.buffalo.edu/bfo/, last accessed 2018/07/11.
- 12. Gangemi A, et al.: Sweetening Ontologies with DOLCE. LNCS, vol 2473. Springer, Berlin, Heidelberg (2002).
- 13. Mobi a decentralized, federated and distributed graph data platform for teams and communities - https://github.com/matonto/matontohttps://creativecommons.org/licenses/, last accessed 2018/07/11.
- 14. W3C: Simple Knowledge Organization System Reference (2009)
- 15. OWL Document Overview https://www.w3.org/TR/owl2-overview/, last accessed 2018/07/11

Kiritsis, Dimitris; Kulvatunyou, Boonserm; Smith, Barry; Wallace, Evan; Will, Chris. "The Industrial Ontologies Foundry Proof-of-Concept Project." Paper presented at APMS 2018 International Conference Advances in Production Management Systems (APMS 2018), Seoul, Republic of Korea. August 25, 2018 - August 30, 2018.

Dimensioning Wireless Use Cases in Industrial Internet of Things

Yongkang Liu*, Richard Candell[†], Mohamed Kashef*, Lotfi Benmohamed*

*Advanced Network Technologies Division [†]Intelligent Systems Division National Institute of Standards and Technology, Gaithersburg, Maryland, USA Email: {yongkang.liu, richard.candell, mohamed.hany, lotfi.benmohamed}@nist.gov

Abstract—Industrial wireless is rapidly developing to enhance data dissemination in the Industrial Internet of Things (IIoT) for leveraging the visibility, control, and safety in industrial environments. In this paper we propose a generic framework that enumerates physical and cyber factors in an IIoT use case that may impact the performance of wireless transmissions and challenge wireless system designs. These factors are discussed in a conceptually tiered structure using domain knowledge from operational technology (OT) applications, data services, and information technology (IT) infrastructures. The correlations between intra- and inter-domain factors are also discussed in search of efficient wireless design solutions. Given the diversity and variety of IIoT practices, such a framework that identifies general and specific wireless requirements in individual use cases would be of high practical value to industrial wireless design and implementation.

Index Terms—Industrial Internet of Things (IIoT), industrial wireless networks.

I. INTRODUCTION

The Industrial Internet of Things (IIoT), which makes use of generic Internet of Things (IoT) techniques [1], connects the massive number of industrial devices and enables machineto-machine data communications in the manufacturing sector to monitor, learn, and manipulate physical processes and objects [2]. In the IIoT paradigm, communication systems play important roles that connect various industrial "things" such as sensors and actuators, and convey system status and control information within open and closed-loop processes. Compared to wired connectivity, wireless networks have advantageous features, such as connection flexibility and cost efficiency, allowing for reaching distributed nodes and serving mobile objects such as robots, raw materials, and parts moving between work cells. Wireless networks are thus becoming increasingly popular in IIoT practices.

Wireless use cases in the IIoT vary considerably in terms of traffic pattern, system operation, and deployment environment. As a result, wireless problems that are formulated in individual use cases usually differ vastly from one to another. Following the bottom-up design style that is commonly used in the literature, wireless solutions that function well under the use case conditions they were designed for may only yield limited value in different use cases. To replicate the wireless success in more plants and new applications, wireless adoption in the IIoT requires guidance from design studies to be able to identify various factors shaping wireless use cases and to weigh their impact on the wireless system design.

This paper provides a glimpse of an ongoing effort among the earliest attempts to classify wireless use cases in a named structure and identify wireless design issues that are associated with individual factors. As one of the major contributions, the proposed framework decomposes the analysis of complex IIoT use cases into separate technical domains, i.e., operational technology (OT) applications, IIoT data services, and information technology (IT) infrastructures, each of which comprises structured factors that define wireless solutions for target use cases. By adopting such a framework, OT engineers, IT administrators, and network planners are able to address the design issues in their domain expertise and collaboratively contribute to the wireless solutions.

The remainder of this paper is organized as follows. Section II reviews related work that motivates the framework design to classify various wireless use cases in industrial environments. Section III introduces the proposed framework and elaborates the relation between individual factors and their impact on wireless system design. Section IV concludes the paper and identifies further work.

II. RELATED WORK

Research and implementation efforts have been taken in industrial wireless applications, such as collecting massive sensor data and enabling real-time control, with or without infrastructure support [3]. A number of air interfaces are introduced to build industrial wireless networks including both commercial solutions, e.g., cellular and Wi-Fi, and industrial variants, e.g., IEEE 802.15.4 time division multiple access (TDMA) radios developed for endurability and determinism [4], [5]. Radio resource management (RRM) schemes adapt network changes, such as in user traffic and radio channel characteristics, to provide reliable data services in dynamic and diverse wireless environments. Standardization efforts are also making progress in parallel to integrate wireless systems into the unified factory information architecture, such as OPC UA [6].

Diverse requirements on wireless system performance in different IIoT use cases have been observed. Such diversity mostly lies in work patterns of industrial plants, usage of the transmitted OT data, and factory IT infrastructure supports on data communication and compute resources. Some recent work

U.S. Government work not protected by U.S. copyright



Fig. 1: Reference Framework of classifying impact factors in wireless use cases

has addressed the design of open and scalable IT platforms for generic IoT applications, such as smart grid and smart city [7]. However, IIoT deployments have unique features in terms of data structures, service requirements, and coordination between IIoT devices in factory operations, which requires tailored design principles. Moreover, as cloud computing has been in industrial control and business analytics, wireless networks are also expected to facilitate more timely and reliable communications between field devices and remote controllers [8].

Recently, some classification work has been reported in the literature that distinguishes wireless use cases by different criteria, such as the type of industrial sites, the scale of wireless connectivity, or the application of wireless data [3]. However, there is still a lack of a comprehensive review that enumerates all necessary factors in a generic use case, traces logical connections between them, and weights their values of guidance in the design of effective and efficient systems. As IIoT practices are greatly expanding the working range of wireless links and increasing the influence of wireless data services in cyber-physical systems, it is urgent to systematically revisit the classification criteria of wireless use cases. Motivated by the necessity of such a study, we propose a new dimensioning framework to classify wireless use cases and identify their particular impacts and requirements on wireless system design.

III. FACTORS DIMENSIONING WIRELESS USE CASES

A. Factor Domains

In this section, a framework is proposed that recognizes various factors in wireless use cases and elaborates their own roles in wireless system design following a layered architecture. Factors in wireless use cases, named *dimensioning factors (DF)*, are identified and organized by wireless-related functions and features in an information-centric infrastructure. The information update for OT applications is carried as various data services in the IT systems so that plant controllers and process analyzers can subscribe sensor updating services to collect data from sensors. The data can then be distributed

through broadcast/multicast links along with control messages to actuators for the target applications. For each service type, the IT infrastructure will provide necessary network support on data communications, e.g., various air interfaces and communication protocols, as well as the other IT functions including data storage, computing, and IT control functions. Accordingly, factors can be represented by their roles in these services so that network planners can innovate using separate domain expertise in modularized design solutions.

Fig. 1 illustrates the proposed conceptual framework that groups DFs in three domains, i.e., OT applications, IIoT data services, and IT infrastructures, which form the top layer of the classification framework. Each domain further hosts the affiliated factors in functional clusters specified by domain knowledge. Before elaborating individual factors along with their respective impacts in wireless use cases, the features of the proposed framework are addressed first: (1) The domain division secures the independence and transparency of intradomain design while offering the opportunities of inter-domain collaborations; (2) The DFs capture diverse structural features and operations of wireless use cases so that the domain expertise can be employed to address design requirements and concerns in respective dimensions; (3) A wireless use case can accommodate multiple domain instances in a complex industrial scenario, i.e., the multi-tenancy case; (4) The proposed classification framework is open and scalable, which clearly illustrates the basic rules behind wireless use cases that assist problem formulation and solution development of wireless networks in the IIoT. In the remaining part of this section, factors in each domain are enumerated with discussions to reveal their impacts on practical wireless problems.

B. Factor Analysis in OT Domain

Factors in the domain of OT applications address functions and activities of manufacturing systems and physical environments in a wireless use case, which vary with application tasks and plant sites. By the covered issues in the OT domain, the factors are classified into two clusters, i.e., *environmental* and *engineering* DF.

Benmohamed, Lotfi; Candell, Richard; Hany, Mohamed; Liu, Yongkang. "Dimensioning Wireless Use Cases in Industrial Internet of Things." Paper presented at 14th IEEE International Workshop on Factory Communication Systems, Imperia, Italy. June 12, 2018 - June 15, 2018.

1) Environmental DF: The cluster of environmental factors depict the working environment of wireless networks and are divided into topographic and mobility factors. Topographic factors regulate the shape of the plant floor where production takes place, which covers environmental features such as (indoor/outdoor) site types, layers of floors, richness of blocks and aisles, and materials of the supporting walls and beams. The mobility factors focus on depicting the impacts of objects' movements in wireless use cases. The environmental factors are generally resource-oriented, i.e., they claim their influence on the availability and effectiveness of utilizing radio resources rather than explicitly raising requirements on data services. The topographic factors mainly function in the radio frequency (RF) environment while mobility factors shape the resource needs to support mobile nodes as discussed above.

2) Engineering DF: The other cluster of OT domain factors are engineering factors that address the OT engineering activities in the plant. Generally, these factors can be grouped into three sets which mirror the three facets in the design of industrial systems and production operations, i.e., depth of intelligence, function pyramid, and layouts and schedules. These three facets indicate the data-related features and operations of industrial applications in wireless use cases. Unlike environmental factors, engineering factors do not explicitly address their preference or requirements on the way of transmitting data, e.g., using wireless networks or optical communication links. Given the relative independence between OT and IT engineering efforts, the proposed framework defines a standard interface, i.e., data services, that ensures data transparency while encouraging design collaborations in cyber-physical systems. As shown in Fig. 1, engineering factors would influence wireless systems through defining and customizing various data services in wireless links, which introduces the factors in the second domain, i.e., HoT data services. C. Factor Analysis in Data Service Domain

Factors in the domain of data services are classified into three clusters, each of which represents one aspect of defining a data service, i.e., data semantization, utility indexing, and flow attributes.

1) Data Semantization: It refers to adding general markups and notifications into the formatted data so that data are annotated with domain knowledge and context information about the served applications/processes. Such add-on information, including domain knowledge and measurement system, enables different devices/systems to better interpret heterogeneous data that are shared in the IIoT. Specifically, domain knowledge is used in the semantic annotation to associate the plain data, e.g., numerical values, with their target metrics in physical processes/systems, and provides necessary system background and context information to interpret the data. Domain knowledge varies with industry and regulation. Measurement system regulates the units of measurement in target OT applications as well as identifying the representation format in the annotation.

2) Utility Indexing: Utility measures in data services share a set of perceptible performance metrics between the OT and IT systems in wireless use cases, which match various data requests of OT applications with service capabilities of IT

resources at some common quality of service (QoS) level. Utility measures that depict different levels of transmission performance can be classified into three facets, i.e., capacity, elasticity, and agility, which regulate the performance boundaries, evaluate the change rates, and test the recognition and reaction speed to service changes, respectively. Utility measures mainly affect individual service metrics to identify the QoS in wireless transmissions. Some typical QoS metrics can be found in different service dimensions, such as the maximum flow rate (by capacity), received signal strength of edge user (by elasticity), and latency threshold (by agility).

3) Flow Attributes: Resembling the interactions between data semantization and OT applications, data services also have various flow attributes that serve as the control port to the IT infrastructure domain. As data flows may encounter conflicts between service demand and resource supply, as well as uneven distributions of IT resources, flow attributes regulate data flows in the end-to-end connections from data sources to sinks, specify the forwarding strategy at intermediate relay nodes, and check the transmission quality in each relay hop to ensure the service satisfaction over heterogeneous IT resources. There are two factors identified as flow attributes, i.e., dispatch and tunneling. Flow dispatch refers to the selection of flow paths for service data beyond a single IT system, while tunneling of data flows configures various IT resources in the end-to-end paths to provide consistent transmission performance fitting data service requirements.

Fig. 1 indicates that factors in the data service domain are indispensable in wireless use cases where they have revealed the logical connections between OT and IT systems, i.e., the OT-to-IT conversions from application data to abstract service requests and the reverse ones from heterogeneous resources to unified channels, which closes the cycle of cyber-physical system design. OT engineering factors that are identified in Section III-B influence wireless system design through data services. Meanwhile, port factors in semantization and flow attributes have addressed case-specific adaptiveness and kept system-wide interoperability in the design of IIoT data services.

D. Factor Analysis in IT Infrastructure Domain

The IT infrastructure generally comprises necessary resources, such as core network connections, compute resources, and storage resources, to support data services in collecting, distributing, and processing industrial application data. Meanwhile, it also applies various management functions and policies over these resources in a particular architecture. The impact of IT infrastructures can be identified in the operations, which are divided into two types, i.e., service provisioning and architecture management.

1) Service Provisioning: Service-oriented operations in the IT infrastructure refer to the admission, customization, and management of data services that vary diversely with their features as discussed in Section III-C. According to the processing order in a data service, these operations can be further classified into subscription and customization operations. The former one occurs before starting a new service type or adding a new service flow by user demand; and the latter one covers

	On-premises	Cloud (X as a Service)		
		Infrastructure (IaaS)	Platform (PaaS)	Service (SaaS)
Virtualization	Network appliances	+ Compute and storage	+ Software API	+ Software (good-to-go)
Planner	Plant IT team	Third-party cloud solution providers and network operators		
Services	End-to-end	Metal (MaaS), IaaS	PaaS	SaaS
Openness	Low	Medium	High	High
CapEx	High	Low (activation fee)	Low (activation fee)	Low/none
OpEx	Medium (utility +	Variable (subscription fee & charge per use)		per use)
Capacity/Elasticity	System updating)	Florible		
Capacity/Elasticity	Lingitz d			
Scalability	Linnied	Demand-specific		
Data exposure	System capacity	Site layout and shift	Traffic pattern	Information interest
Exemplary usage	Proprietary IT infrastructure	"Plug-n-play" radio access networks (RAN)	Mobile Apps	Online data subscription (e.g., air quality, utility market price)

TABLE I: Feature comparison between IT infrastructure options

the life cycle of active data services to allocate and adjust resources for varying service demands.

2) Architecture Management: In the IT infrastructure, architecture management operations organize IT resources and functions structurally by their physical connections (e.g., relay, bridge, and switch) and logical relations (e.g., affiliation, supplement, and competition in conceptual domains) in between. Resilience factors focus on the architectural support on IT management functions and specify the way of maintaining and orchestrating the IT system. For example, a wireless use case needs to specify its capability of using software-defined networking (SDN) and network function virtualization (NFV) technologies to tackle IT function diversity and resource heterogeneity. Provider factors address the policies that regulate the ownership and authority in the IT deployment, such as economic considerations about purchasing on-premises IT appliances and renting cloud computing and storage to build the IT architecture. As wireless networks are cooperating with factory IT architectures in IIoT practices, the wireless network design should follow the rules in the IT architecture to achieve interoperability between access and core network resources and service consistency.

E. Case Study: Process Automation

We choose a typical process automation application in a chemical production plant to demonstrate the framework usage. In the OT domain, the plant works in a typical indoor industrial RF environment with rare mobility in operations. The repeated system status measurements and actuation commands form a field control process that may work on a 24/7 schedule. As a result, such a process control application can deploy the publisher/subscriber data services to enable periodic process and control message transmissions between field nodes and the programmable logic controller (PLC). The details about the chemical process and one wireless network design based on the IEEE 802.15.4 radios can be referred to [9] and [10], respectively. When the wireless data service is integrated into the factory IT infrastructure, there would be multiple options as illustrated in Table I following the identified IT factors.

IV. CONCLUSIONS

A new comprehensive framework has been proposed in this paper which reviews the critical factors in industrial wireless use cases. Such a framework has practical value as it standardizes the design cycle of wireless systems for industrial applications and reduces the time and cost involved in customizing the design problems for different use cases. Useful directions for future work include automating the problem formulation process of industrial wireless use cases and standardizing the recommendation criteria on wireless techniques and solutions for the identified problem.

DISCLAIMER

Certain commercial equipment, instruments, materials, or systems are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

References

- [1] U. Sendler, The Internet of Things: Industrie 4.0 Unleashed, Springer Vieweg, 2017.
- [2] L. Xu et al., "Internet of Things in Industries: A Survey", IEEE Trans. Ind. Informat., vol. 10, no. 4, pp. 2233-2243, Nov. 2014.
- V. C. Gungor and G. P. Hancke, "Industrial Wireless Sensor Networks: Challenges, Design Principles, and Technical Approaches", IEEE Trans. *Ind. Electron.*, vol. 56, no. 10, pp. 4258-4265, Oct. 2009. [4] IEC 62591 Ed. 1.0 b:2010, "Industrial Communication Networks
- Wireless Communication Network and Communication Profiles WirelessHARTTM," 2010.
- ISA, "Wireless Systems for Industrial Automation: Process Control and [5] Related Applications", ISA-100.11a-2009 Standard, 2009. D. Schilberg et al., "Interoperability in Smart Automation of Cyber
- [6] Physical Systems", in Industrial Internet of Things - Cybermanufacturing Systems, S. Jeschke et al., Eds. Springer Nature, 2017, pp. 261-287.
- [7] J. Gubbi et al., "Internet of Things (IoT): A vision, architectural elements, and future directions", Future Generation Computer Syst., vol. 29, no. 7, pp. 1645-1660, Sept. 2013.
- [8] A. Leonardo et al., "A Wireless Cloud Network Platform for Industrial Process Automation: Critical Data Publishing and Distributed Sensing", IEEE Trans. Instru. Meas., vol. 66, no. 4, pp. 592-603, Apr. 2017.
- J. J. Downs and E. F. Vogel, "A Plant-wide Industrial Process Control Problem", Comput. Chem. Engng., Vol. 17, No. 3, pp. 245-255, 1993.
- [10] Y. Liu, R. Candell, K. Lee, and N. Moayeri, "A simulation framework for industrial wireless networks and process control systems," in Proc. IEEE WFCS'16, May 2016, pp. 111.

COMPARATIVE ANALYSIS OF BEARING HEALTH MONITORING METHODS FOR MACHINE TOOL LINEAR AXES

N. Jordan Jameson and Gregory W. Vogl Engineering Laboratory National Institute of Standards and Technology (NIST) 100 Bureau Drive Gaithersburg, MD 20899 Telephone: (301) 975-8966 noel.jameson@nist.gov

Abstract: The study of rotating machinery ball bearing diagnostics and prognostics is quite mature and an abundance of methods/algorithms are available to perform these functions. However, extending these algorithms to other ball bearing applications is challenging and may not yield usable results. This work used a linear axis testbed to study the ability of an inertial measurement unit to measure changes in geometric error motions. Faults were introduced on the recirculating ball bearings of one carriage truck with increasing severity. The inertial measurement unit data was analyzed using a variety of methods proposed and used in the rotating machinery community, including auto-regressive filtering, minimum entropy deconvolution, and spectral kurtosis. The results reveal an ineffectiveness of the methods for the induced faults, for this one experiment, which have low signal-to-noise ratio and/or weaker periodicity than faults in rotating machinery.

Keywords: Linear axis, ball bearing, degradation, diagnostics, smart manufacturing

Introduction: Linear translational axes are important systems in manufacturing, often installed as a subsystem of a computer numerical control (CNC) machine. They are used to move cutting tools and workpieces to their desired positions for part production [1]. In a typical machine, there can be multiple such axes, each of which can degrade, producing inaccuracies that affect the quality of the parts produced. A single linear axis is shown in Figure 1(a), with each of the main components labeled. Four trucks (also called "linear motion guides") are used to guide the carriage along the axis rails using recirculating ball bearings.

As a linear axis operates over time, abrasion and adhesion between parts causes material fatigue, pitting, cracking, and wear. This can result in faults developing in a variety of components such as the axis rails, rolling element bearings, and/or ball screw [2, 3]. If not properly mitigated, these faults will develop to such an extent as to affect the quality of the parts produced and/or precipitate a failure of the machine [4]. As demands for versatility and batch volume increase for manufacturing processes, machines are experiencing higher loads, and as a result the potential for faults and failures is becoming more common.



Figure 1: (a) Linear axis. The carriage travels along the axis of the lead screw and four trucks guide the carriage along the two rails (or guideways). (b) Angular and linear error motions of a carriage commanded to move along a (nominal) straight-line trajectory parallel to the *X*-axis.

The health assessment process consists of measuring the error motions of the linear axis machine and comparing those with specifications for allowable errors. In a linear axis, there are six degrees of freedom corresponding to six error directions, as shown in Figure 1(b), where the second variable in the subscript denotes the direction of travel (in this case, the carriage is traveling in the X-direction). In the schematic, E_{XX} denotes the linear positioning error motion of the X-axis, E_{YX} denotes the straightness error motion in the Y-direction, E_{ZX} denotes the straightness error motion around the X-axis (roll), E_{BX} denotes the angular error motion around the Z-axis (yaw).

Ideally, after a linear axis is installed on a machine tool, all errors are zero, but in practice nonzero errors exist resulting in the errors on the workpiece. Hence, the allowable errors are dependent upon the tolerance needed for the workpiece. As the machine ages or crashes occur as a result of programming mistakes, these errors can increase, leading to the workpieces becoming more and more out of tolerance. Instructive machine tool monitoring coupled with proper maintenance regimes can be used to mitigate this accumulation of degradation and minimize the costs imposed by imperfect production and scrapped parts.

Mature methods for the fault detection and diagnostics of linear axes are manual, time consuming, and often cost prohibitive. The state-of-the-art instruments in linear axis error measurement (the basis for diagnostics) are explained in the International Organization for Standardization (ISO) 230-1 [5]: straightedge and linear displacement sensor, microscope and taut wire, alignment telescope, alignment laser, and laser straightness interferometer. These time-consuming instruments require the shutdown of the machine and thus cannot provide insitu diagnostics [6].

To ease the burden of implementing the traditional monitoring technologies, manufacturers need automated and efficient methods for diagnosing machine tool linear axes without halting production. In 2010, Teti *et al.* [7] identified that new sensors and sensor systems, advanced sensor signal data processing, and intelligent sensor monitoring need to be further developed

to help achieve decreased machine downtime, higher productivity, higher product quality, and enhanced knowledge of manufacturing processes.

Previous work has shown that a possible advance in sensor technology lies in the use of an inertial measurement unit (IMU) consisting of a 3 degree-of-freedom (DOF) accelerometer and a 3 DOF rate gyroscope [8-10] as shown in Figure 2. The bandwidths and noise properties of the sensors used in the IMU are shown in Table 1. Data from the IMU can be used to detect changes in the positioning, straightness, and angular error motions. IMU measurements can be made quickly and with little intrusion into the operation of the machine, resulting in data that provides insight into the condition of the linear axis. It has been shown to be effective at detecting rail degradation to similar levels of accuracy delivered by a laser interferometer [8]. However, a challenge remains in verifying the IMU's capabilities for detecting degradation in the truck bearings of a linear axis.



Figure 2: (a) Isometric view of industrial IMU and (b) top view of industrial IMU without its cover.

Sensor	Bandwidth [*]	Noise
Accelerometer	0 Hz to 400 Hz	69 (μ m/s ²)/ \sqrt{Hz}
Rate Gyroscope	0 Hz to 200 Hz	35 (μ rad/s ²)/ \sqrt{Hz}

Table 1: Specified properties of sensors used in the IMU

frequencies correspond to half-power points, also known as 3 dB points

One complication that presents itself in this task is the multiplicity in the rolling element bearings within the bearing system, which causes a low signal-to-noise ratio (that can be less than 1) in the health monitoring data. In the linear axis shown in Figure 1(a), there are four trucks, each with two loops of ball bearings, and each loop has 32 ball bearings. The ball bearings rotate in and out of contact with the rail, with about 13 ball bearings per loop contacting the rail at any given instance. The two loops interact with different raceways (or grooves) in the rails, one inner and one outer. Hence, at any given instance, there are about 104 bearings (13 bearings per loop \times 2 loops per truck \times 4 trucks) in contact with the rails. The convolution of so many bearing ball signatures into the sensed effects on the carriage make it difficult to isolate small indications of damage from a single bearing ball. This interaction

between the bearings and rails is illustrated in Figure 3(a) and Figure 3(b). It was with this in mind that an experiment was designed to examine the sensitivity of the IMU-based error motions to artificially-induced damage on the bearings.



Figure 3: (a) Schematic of rails and carriage assembly of one-axis testbed; (b) underside view of one truck with loops of ball bearings exposed, where the "inner loop" contacts the inner raceway and the "outer loop" contacts the outer raceway; and (c) view of rail, carriage, lead screw, and trucks assembly showing the attachment of the trucks to the rails.

In this study, ball bearings from a single truck were removed sequentially, one-at-a-time, intentionally degraded, then replaced in their original position within the truck. The balls were degraded one at a time to progressively alter the error motions of the carriage. IMU data was collected after each ball bearing was damaged and the data was analyzed to develop diagnostic features. The data analysis procedure progressed from less to more complex culminating in a series of algorithms recommended by Randall and Antoni in their 2011 bearing diagnostics tutorial [11].

Experimental Setup: Truck 2 was modified to allow the removal of one ball at a time. This permitted the opportunity to inflict damage on to a single ball and then replace it into the truck, while leaving the carriage/truck/rail/lead screw system effectively unchanged. This was necessary because if the trucks were removed and then reassembled at each stage of degradation, then the load placed upon each of the trucks would have changed from one assembly procedure to the next, which would have changed the error motions as well. Hence, to eliminate the need for disassembly and reassembly of the trucks during the experiment and thus ensure that changes in error motions are due only to bearing ball damage in truck 2, the

truck was modified with an access hole and slot, as shown in Figure 4(a). The set screw covered the access hole from which balls were removed to introduce damage. Bearing balls were removed one-at-a-time and a "flat" was introduced by dragging the ball across sandpaper. The flat damage for one bearing is shown in Figure 4(b). After damaging the bearing, it was replaced in Truck 2 at its original position within the outer ring.



Figure 4: (a) Truck 2 as modified for this experiment with a slot for viewing the bearing movement and an access hole covered by a set screw, which allowed for the removal of single bearing ball; and (b) diagram showing the damage introduced to the balls in Truck 2.

Over the course of the experiment, 26 of the 32 bearing balls in the outer loop of Truck 2 were damaged. Figure 5 shows the pattern of twenty-six metal balls with six (black) silicon nitride balls. The pattern was chosen so that at any time, visual inspection of the balls through the slot, as seen in Figure 4(a), yields a unique pattern of black and silver (metallic) balls for identification. The twenty-six metal balls were made of chromium steel with a nominal diameter of 3.962 mm, while the six ceramic balls had a nominal diameter of 3.969 mm. The average depth of material removed for the chromium steel balls was 30 µm with a standard deviation of 2 µm. The maximum depth was 35 µm and the minimum depth was 26 µm. The Truck 2 outer loop balls were degraded in a specific order, as shown in Figure 5, that attempts to increase general frequency content of defect-induced impulses.



Figure 5: Truck 2 outer loop bearing degradation arrangement. The filled in circles represent ceramic bearing balls that were not degraded, while the numbers in the open circles denote the order in which the remaining metal balls were damaged.

After a single bearing was degraded and placed back into Truck 2, 50 runs of IMU data was gathered and processed to yield the estimated error motions for the carriage. Each run consists of moving the carriage over the linear axis, forward and backward, at three different speeds: 0.02 m/s, 0.1 m/s, and 0.5 m/s. These three speeds are meant to capture different ranges in the spatial frequency spectrum, with the slowest speed capturing the highest spatial frequencies and the highest speed capturing the lowest spatial frequencies. With the rate gyroscope and accelerometer used, the available spatial frequency range is between 0 and 2 cycles/mm (this

can also be written as 2000 m⁻¹). (For more information on how the rate gyroscope and accelerometer data was processed to achieve the error motion data, see [8, 9].) Hence, for each of 27 degradation "stages", there are fifty (50) data runs to be analyzed. Stage 1 is collected prior to any balls being degraded, and data collection ended at stage 27 with 26 of the balls having been damaged. The first sample for stages 1, 9, 18, and 27 are shown in Figure 6 for all six degrees of freedom.



Figure 6: Unfiltered error motions of the carriage at four stages of bearing degradation. The error components are noted in the bottom left corner of each sub-figure.

Analysis Methods: The simplest analysis methods use only statistical and frequency domain features to analyze each instance of data. The statistical time-domain features used in the analysis of this experiment are given in Table 2, where \bar{y} denotes the signal average and N denotes the number of data points in the signal. For the current implementation, each 'signal' of Table 2 is an error motion or some transformation of the sampled error motion, e.g., via filtering. Each of the values provided was computed for every instance of error motion data generated at the incrementally increasing stages of degradation.

Table 2: Description of statistical time-domain features used in bearing
diagnostic analysis [12-15]

Feature name	Brief description	Formula
Peak value	The peak value measures the amplitude of the signal, y. As bearing faults develop, there are more impacts producing higher average amplitude, hence the peak value is expected to increase.	$PV(y) = \frac{1}{2} (\max(y) - \min(y))$
RMS	The root mean square (RMS) of a signal is measure of the energy of the signal.	$rms(y) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}y_i^2}$
Standard deviation	The standard deviation measures the dispersion of the signal around the mean value, which is another measure of the amplitude and energy in the signal.	$\sigma(y) = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})^2}$
Skewness	Skewness quantifies the asymmetry of the signal with respect to its probability density function.	$\gamma_1(y) = \frac{\frac{1}{N} \sum_{i=1}^N (y_i - \overline{y})^3}{\sigma^3}$
Kurtosis	Kurtosis quantifies the "tailedness" of a signal. As a bearing degrades, more impacts are expected. This leads to a signal having more frequent and extreme deviations than a signal from a healthy bearing.	$\gamma_2(y) = \frac{\frac{1}{N} \sum_{i=1}^N (y_i - \overline{y})^4}{\sigma^4}$
Crest factor	Crest factor is a ratio of the peak value to the RMS. If there are impacts from a degraded bearing, then this value will be high, due to the discrepancy between the high peak produced from the impacts and the nominal amplitude of the waveform.	$CF(y) = \frac{PV(y)}{rms(y)}$
Shape factor	Shape factor is a ratio between the RMS and the average absolute value. Since RMS measures the square of the signal, then it will be more sensitive to outliers, such as impacts. So as bearings get more degraded, the shape factor is expected to increase.	$SF(y) = \frac{rms(y)}{\frac{1}{N}\sum_{i=1}^{N} y_i }$

Jameson, Jordan; Vogl, Gregory. "Comparative Analysis of Bearing Health Monitoring Methods for Machine Tool Linear Axes." Paper presented at Society for Machinery Failure Prevention Technology 2018 Annual Conference, Virginia Beach, VA, United States. May 15, 2018 - May 17, 2018.

 Table 2: Description of statistical time-domain features used in bearing diagnostic analysis [12-15]

Feature name	Brief description	Formula
Impulse factor	As bearings degrade, they are expected to produce more impacts, which will greatly affect the peak value, but will have a smaller effect on the average absolute value. Hence, as bearings degrade, the impulse factor is expected to increase.	$IF(y) = \frac{PV(y)}{\frac{1}{N}\sum_{i=1}^{N} y_i }$
Clearance factor	The clearance factor is also a measure that relates the peak value to a measure of the average amplitude. As the bearings degrade and more impulses are produced, the numerator should rise while the denominator remains almost constant.	$CLF(y) = \frac{PV(y)}{\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{ y_i }\right)^2}$

Beyond computing these statistical features from the error motions, many signal processing procedures have been proposed in the bearing community. Most of the proposed techniques aim to increase the kurtosis of the signal, in order to isolate changes due to impulses from degradation. These techniques include autoregressive (AR) filtering, self-adaptive noise cancellation (SANC), minimum entropy deconvolution (MED), and spectral kurtosis (SK). The AR filtering and SANC are both methods of separating the deterministic portion of the signal from the random portion of the signal [16]. In both cases, a prediction is computed, then the residual between the predicted signal and the real signal is considered as the bearing signal. In this study, only AR filtering was used.

The AR filter predicts the deterministic pattern of the signal, but is not capable of adapting to the sudden impulses caused by bearing faults. Consequently, the fault signal is expected to remain in the residual of the AR filter, which is the difference between the measured signal and the estimated signal (i.e., the output of the AR filter). An AR(L) (i.e., an L^{th} order AR) process is one in which the signal is assumed to be given as a linear combination of L past values and some input, e_n , as shown in (1):

$$x_n = -\sum_{l=1}^{L} a_l x_{n-l} + e_n$$
 (1)

By assuming the input is unknown, the signal can be approximately predicted from the linear summation of past samples. This is computed as shown in (2).

$$\hat{x}_n = -\sum_{l=1}^{L} a_l x_{n-l}$$
(2)

The residual is then computed as $e_n = x_n - \hat{x}_n$. The model coefficients can be computed through a variety of methods, though perhaps the most used is the Yule-Walker equations, which relate the model coefficients to the autocovariance of the signal [16]. The order of the AR model (i.e., the number of model coefficients) can be chosen using a variety of criteria, such as Akaike information criteria (AIC), Bayesian information criteria (BIC), or cross validation. However, in the present study, the AR model order was chosen as the number of coefficients that maximized the kurtosis of the residual, in an attempt to isolate the defectcaused terms from the error motion data.

After separating out the bearing signal via AR (or SANC) filtering, the impulsive fault signal can be enhanced using MED, which was originally proposed by Wiggins [17] to extract information from seismic recordings. MED operates under the assumption that the original excitation was impulsive (with high kurtosis), then designs a filter that attempts to reconstruct the original signal. This is a means to overcome the effect of the transmission path through which the impulses reach the sensor(s). Since impulsive signals are highly structured (low entropy), the desired filter should minimize the signal entropy. Sawalhi et al. [18] applied MED to bearing faults, Endo and Randall [19] used it with gear faults, and Randall and Antoni [11] recommended its use as part of a "semi-automated bearing diagnostic procedure". Lee and Nandi [20] and McDonald et al. [21] have provided comprehensive methods for computing MED. The procedure used herein is shown below.

Step 1: Assume a centered impulse filter, $f = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^T$, of length

Step 2: Compute the $L \times L$ autocorrelation matrix \mathbf{R}_{x} ; a Toeplitz matrix with the first row as the first L terms of the autocorrelation of \boldsymbol{x}

- Step 3: Compute filter output, $\mathbf{y} = \mathbf{X}_0^T \mathbf{f}$ Step 4: Compute new filter: $\mathbf{f} = \frac{\sum_{i=1}^N y_i^2}{\sum_{i=1}^N y_i^4} \mathbf{R}_x^{-1} \mathbf{X}_0 \mathbf{y}^3$
- Step 5: Repeat steps 3-4 until stopping criterion are satisfied

Here, y^3 is an element-wise cubing of the filtered signal, and X_0 is an $L \times N$ matrix of the delayed input signal:

$$\boldsymbol{X}_{0} = \begin{bmatrix} x_{1} & x_{2} & x_{3} & \cdots & x_{N} \\ 0 & x_{1} & x_{2} & \cdots & x_{N-1} \\ 0 & 0 & x_{1} & \cdots & x_{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & x_{N-L+1} \end{bmatrix}$$

The stopping criterion can be based upon the change in the filtered signal kurtosis (i.e., once the change in filtered signal kurtosis reaches a sufficiently small value) as proposed by McDonald et al. [21]. Or the stopping criterion can be based upon changes in the filter as proposed by Lee and Nandi [20], stating that when $E[\epsilon]$ gets below a certain threshold, the routine is complete.

$$\epsilon = (f^{(i)} - \mu f^{(i-1)}) / \mu f^{(i-1)}$$

$$\mu = \sqrt{\frac{E[(f^{(i-1)})^2]}{E[(f^{(i)})^2]}}$$
(3)

In (3), the expressions are shown for the *i*th iteration, $(f^{(i)})^2$ is an element-wise squaring of the vector $f^{(i)}$, and $E[\cdot]$ denotes the expected value. In this study, the criterion proposed by Lee and Nandi was used with the threshold for $E[\epsilon]$ set to $5(10^{-4})$.



Figure 7: Frequency and decomposition levels for construction of kurtogram.

The last step in the process is to use spectral kurtosis and the kurtogram to locate the frequency band and frequency resolution that maximizes the signal kurtosis. The kurtogram is a method that uses a finite impulse response (FIR) filter bank to decompose the signal into a set of subsignals with different frequency bands [22]. This decomposition consists of taking the signal, then low-pass- and high-pass-filtering to the frequency bands [0, 1/4] and [1/4, 1/2], respectively (in normalized frequency). This is considered as Level 1 decomposition. The original signal is also decomposed into the frequency bands [0, 1/6], [1/6, 1/3], and [1/3, 1/2]. This is considered as Level 1.6 decomposition. Then the resulting signals are sent through further replicate filter banks, producing the tree-like structure shown in Figure 7, where $S_{k,j}$ represents the signal decomposed at level k in the *j*th frequency band, and f_s is the sampling frequency. After the signals are decomposed, the kurtosis of each $S_{k,j}$ can be computed, and the level and frequency band that maximizes kurtosis chosen.

Experimental Analysis Results: The error motions were first high-pass filtered to frequencies above 500 samples/m, which was done to eliminate any data that simply captured the shape of the rails. Bearing fault signatures are assumed to occur with wavelengths less than about 12.4 mm (circumference of a single ball), therefore sampling between 0.5 samples/mm and 2 samples/mm is sufficient to capture any impacts from bearing faults. Most of the statistical metrics yielded no indications of the increasing number of faults being introduced to the balls of truck 2. As an example, the kurtosis of each sample is shown in Figure 8. There are 27 total stages, and at each stage, there are 50 runs, translating to 50 values of kurtosis. At each stage,

the gray box represents the middle 50% of these 50 values, and the whiskers extend to the largest value that falls below $q_3 + 1.5(q_3 - q_2)$, or the smallest value that falls above $q_2 - 1.5(q_3 - q_2)$, where q_2 and q_3 are the 25th and 75th percentile respectively. Any value that falls outside this range is classified as an outlier, and the outliers are plotted as red circles. The black circle in the center of each box is the median value for all 50 runs in a given degradation stage.



The statistical features can be evaluated using the Pearson correlation coefficient between the median of each distribution of 50 samples and the degradation stages (1-27). Because Pearson correlation measures the strength of a linear relationship between two variables, it is expected that as a feature is more responsive to bearing degradation, it will register a higher Pearson correlation coefficient magnitude. The sign of the Pearson correlation coefficient indicates the direction of the line, where +1 indicates the line is increasing and -1 indicates the line is decreasing. The Pearson correlation coefficient for each feature and DOF is shown in Figure 9.

From Figure 8 and Figure 9, it is apparent that E_{YX} trends the most as the bearings of Truck 2 outer loop are degraded, since most of its statistical features have Pearson correlation coefficients with magnitudes significantly above 0.5. Furthermore, E_{XX} has a moderate trend in kurtosis and E_{ZX} has a moderate trend in skewness, each with Pearson correlation coefficient magnitudes around 0.5. Moreover, in the case of E_{YX} , the kurtosis starts as lower than that of the normal distribution (which has a kurtosis of 3) and ends with most runs still below a

kurtosis of 3, but having a higher kurtosis than at Stage 1. Generally, when bearings are faulty, the impacts that the bearings have with the runways produce kurtosis values much higher than that of normal distribution (see e.g., [18, 23]). This physical trend should be further examined for further understanding. Also, it must be noted that the error motion data under examination are high-pass filtered and not unprocessed.



Figure 9: Pearson correlation coefficient for each statistical feature computed.

After performing the procedure recommended by Randall and Antoni, the results for one response are shown in Figure 10. The data was first high-pass-filtered to above 500 m⁻¹, then an AR(26) filter was used to eliminate the deterministic portion of the signal. The length of the AR filter was found by exhaustively searching all error directions. Each signal was filtered using filter lengths between 2 and 400, and the kurtosis was measured for each filtered signal. For all error directions, the optimal length that maximized filtered signal kurtosis was between 25 and 28, with an average of 26. After AR filtering, the MED filter of length 10 was used. The length of the MED filter was also chosen according to the length that maximized signal kurtosis. Finally, the kurtogram was used on the MED filtered signal.

After processing all six DOF error motion data in the manner shown in Figure 10, the kurtosis values are shown using boxplots in Figure 11, à la Figure 8. In most cases, the kurtosis achieved using the MED filter was very close to the kurtosis achieved using the kurtogram. More importantly, like the simple statistical analysis results shown in Figure 8, the straightness errors in the Y-direction exhibit the most movement in Figure 11 as the ball bearings were degraded. Also, as shown in comparison of Figure 11 to Figure 8, this signal kurtosis was increased by decomposing the MED filtered signal to Level 1 with center frequency of $f_s/8 = 2500 \text{ m}^{-1}$. However, it appears that the kurtogram may have decreased the robustness for monitoring changes in degradation, because the kurtogram increased the variance of the kurtosis of E_{YX} , as evidenced by the significantly longer gray boxes for E_{YX} in Figure 11 compared with those for E_{YX} in Figure 8.

These results show that certain metrics lack no significant trends with increasing degradation, which could result from several possibilities. First, the faulty bearings occurred in only one of

two bearing loops in one of four trucks on the linear axis, so it is possible that the degradation in one bearing ball loop does not significantly affect certain error motions due to the influence of seven healthy bearing ball loops and the geometry of the raceways. Second, the signal processing techniques used were designed for temporal measurements (i.e., vibration as a function of time) that typically have many cycles for each dataset due to the rotary nature of the applications (e.g., spindles and motors). On the other hand, each run collected by the IMU has on the order of only one cycle of ball motion, in which the balls move in and out of contact between the truck and rail. Furthermore, because of the approximately one cycle of ball motion and recirculation of balls, the spatial locations of the defect-caused impulses along the rail are constantly moving, sometimes to outside of the spatial measurement range. Third, the ceramic balls were nominally 7 μ m larger in diameter than the metal balls, which means that as more metal balls were degraded, the ceramic balls were bearing more of the load, which could account for apparent approaching of relative equilibria in Figure 8.



Figure 10: Semi-automated bearing diagnostic procedure for degradation stage 16, run 6, in *y*-axis translation, with the respective kurtosis values noted in the upper right corner of each plot. (a) Error response filtered to spatial frequencies above 500 m⁻¹. (b) Residual of signal after *AR*(26) filtering of the signal from (a). (c) Signal after MED (filter length of 10) of signal from (b). (d) Magnitude of complex envelope of the signal from (c) which maximized the kurtogram (level 1 decomposition) and its 0.1% significance threshold (shown in red).



Finally, the results raise many questions for future work:

- 1. What is the lowest amount of damage that can be detected with the IMU via the methods explored or any other?
- 2. Why is the *Y*-direction most sensitive to changes in the bearings?
- 3. What signal processing techniques could be used to increase the signal to noise ratio for earlier detection of degraded linear axis bearings?
- 4. How can signal processing techniques be adapted to dealing with spatial measurements, rather than temporal measurements?

Conclusion: The linear axis is used in a variety of machining processes and their bearings are subject to wear. This paper presented an experiment wherein an inertial measurement unit (IMU) was employed to detect degraded bearings in a linear axis. The six DOF error motions for the carriage were first analyzed for statistical features and then signal processing techniques were used to amplify the signal of degraded bearings. The results showed the error motions, determined from processing IMU data, experienced some evolution over time, particularly in the *Y*-direction (error motion described as E_{YX}). Pearson correlation coefficients showed additional trends detected by kurtosis for the translation in the *X*-direction (E_{XX}) and skewness for translation in the *Z*-direction (E_{ZX}). However, certain metrics lacked no significant trends with increasing degradation, which could have resulted from the experimental setup (lack of cycles per run), physical phenomena (effects of degradation, recirculation of balls), and the analytical techniques themselves. The results motivate further research into fault detection and

diagnostics for linear axes, with an emphasis on identifying the optimal techniques suited to the unique problem of linear translation.

NIST Disclaimer: Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References:

- [1] Y. Altintas, A. Verl, C. Brecher, L. Uriarte, and G. Pritschow, "Machine tool feed drives," CIRP Annals - Manufacturing Technology, vol. 60, no. 2, pp. 779-796, // 2011.
- Y. Li, X. Wang, J. Lin, and S. Shi, "A Wavelet Bicoherence-Based Quadratic [2] Nonlinearity Feature for Translational Axis Condition Monitoring," Sensors, vol. 14, no. 2, pp. 2071-2088, 2014.
- [3] Y. Zhou, X. Mei, Y. Zhang, G. Jiang, and N. Sun, "Current-based feed axis condition monitoring and fault diagnosis," in 4th IEEE Conference on Industrial Electronics and Applications, ICIEA 2009, Xi'an, China, 2009, pp. 1191-1195: IEEE Computer Society.
- [4] E. Uhlmann, C. Geisert, and E. Hohwieler, "Monitoring of slowly progressing deterioration of computer numerical control machine axes," Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, vol. 222, no. 10, pp. 1213-1219, 2008.
- ISO 230-1 Test code for machine tools Part 1: Geometric accuracy of machines [5] operating under no-load or quasi-static conditions, 2012.
- A. W. Khan and W. Chen, "Calibration of CNC milling machine by direct method," in [6] 2008 International Conference on Optical Instruments and Technology: Optoelectronic Measurement Technology and Applications, Beijing, China, 2009, vol. 7160, p. 716010: SPIE.
- R. Teti, K. Jemielniak, G. O'Donnell, and D. Dornfeld, "Advanced monitoring of [7] machining operations," CIRP Annals - Manufacturing Technology, vol. 59, no. 2, pp. 717-739, 2010.
- [8] G. W. Vogl, M. A. Donmez, and A. Archenti, "Diagnostics for geometric performance of machine tool linear axes," CIRP Annals - Manufacturing Technology, vol. 65, no. 1, pp. 377-380, 2016.
- [9] G. W. Vogl, M. A. Donmez, A. Archenti, and B. A. Weiss, "Inertial Measurement Unit for On-Machine Diagnostics of Machine Tool Linear Axes," presented at the Annual Conference of the Prognostics and Health Management Society 2016, Denver, CO, October 3 - 6, 2016, 2016.
- [10] G. W. Vogl and M. E. Sharp, "Diagnostics of machine tool linear axes via separation of geometric error sources," presented at the Annual Conference of the Prognostics and Health Management Society, St. Petersburg, FL, October 2-5, 2017, 2017.
- R. B. Randall and J. Antoni, "Rolling element bearing diagnostics-A tutorial," [11] Mechanical systems and signal processing, vol. 25, no. 2, pp. 485-520, 2011.

Jameson, Jordan; Vogl, Gregory. "Comparative Analysis of Bearing Health Monitoring Methods for Machine Tool Linear Axes." Paper presented at Society for Machinery Failure Prevention Technology 2018 Annual Conference, Virginia Beach, VA, United States. May 15, 2018 - May 17, 2018.

- [12] W. Caesarendra and T. Tjahjowidodo, "A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing," *Machines*, vol. 5, no. 4, p. 21, 2017.
- [13] I. Howard, "A Review of Rolling Element Bearing Vibration'Detection, Diagnosis and Prognosis'," Defence Science and Technology Organization Canberra (Australia)1994.
- [14] B. Sreejith, A. Verma, and A. Srividya, "Fault diagnosis of rolling element bearing using time-domain features and neural networks," in *Industrial and Information Systems, 2008. ICIIS 2008. IEEE Region 10 and the Third international Conference* on, 2008, pp. 1-6: IEEE.
- [15] C. T. Yiakopoulos, K. C. Gryllias, and I. A. Antoniadis, "Rolling element bearing fault detection in industrial environments based on a K-means clustering approach," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2888-2911, 2011/03/01/2011.
- [16] R. Randall, N. Sawalhi, and M. Coats, "A comparison of methods for separation of deterministic and random signals," *International Journal of Condition Monitoring*, vol. 1, no. 1, pp. 11-19, 2011.
- [17] R. A. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, vol. 16, no. 1, pp. 21-35, 1978/04/01/1978.
- [18] N. Sawalhi, R. B. Randall, and H. Endo, "The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2616-2633, 2007.
- [19] H. Endo and R. B. Randall, "Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 906-919, 2007.
- [20] J.-Y. Lee and A. Nandi, "Extraction of impacting signals using blind deconvolution," *Journal of Sound and Vibration*, vol. 232, no. 5, pp. 945-962, 2000.
- [21] G. L. McDonald, Q. Zhao, and M. J. Zuo, "Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection," *Mechanical Systems and Signal Processing*, vol. 33, pp. 237-255, 2012.
- [22] J. Antoni, "Fast computation of the kurtogram for the detection of transient faults," *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 108-124, 2007/01/01/ 2007.
- [23] J. Antoni and R. Randall, "The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines," *Mechanical Systems and Signal Processing*, vol. 20, no. 2, pp. 308-331, 2006.

Towards a Platform for Smart Manufacturing Improvement Planning

SangSu Choi¹, Thorsten Wuest², and Boonserm (Serm) Kulvatunyou³

 ¹ IGI, LLC, Clarksburg MD 20871, USA Sangsu.choi@igiamerica.com
 ² Industrial and Management Systems Engineering, West Virginia University, Morgantown, WV 26506-6070, USA thwuest@mail.wvu.edu
 ³ Systems Integration Division, National Institute of Standards and Technology, Gaithersburg, MD, USA 20899-1070, USA boonserm.kulvatunyou@nist.gov

Abstract. The manufacturing industry is transitioning towards smart manufacturing systems (SMS). Small and medium size manufacturers (SMMs) are particularly behind in this transition, plagued by lack of knowledge and resources. Several smart manufacturing capability assessment and maturity models exist to guide the transition. However, support for choosing the right one is lacking. This paper proposes a web-based, open source platform for smart manufacturing assessment to support particularly SMMs in this transition. The platform allows for risk-free self-assessments of the current maturity levels and developments of continuous improvement plans that are customized to the manufacturers' unique characteristics. The platform also allows for sourcing of third-party technologies and services relevant to the improvements. More importantly, it will learn, rate, and recommend improvements and services based on past data. The platform is designed to be extensible and scalable to ultimately serve manufacturing enterprises of all industries.

Keywords: Smart Manufacturing, Smart Factory, Industry 4.0, Cyber-Physical Production System, Smart Manufacturing Assessment, Smart Manufacturing Readiness

1 Introduction

The world is moving towards the next industrial revolution where manufacturing enterprises are seeking for performance improvements through a new level of interconnectivity within their own production environments and value chains. The core of the current industrial revolution is the development of smart manufacturing systems (SMS). SMS are built on the intersection of cutting-edge information technology (IT) and Operation Technology (OT) such as internet of things (IoT), big data, cloud computing, machine learning, additive manufacturing, CNC machine tools, and collaborative robots [1, 2].

A smart manufacturing environment is ultimately realized by integrating several of these core technologies in a customized and value-adding way. SMS must be developed harmoniously based on a combination of IT/OT technologies that fit each individual company's unique manufacturing environment. A recent study by West Virginia University (WVU) [3] on Small and Medium-Size Manufacturers' (SMMs) smart manufacturing adoptions found that the vast majority of SMMs do not possess the required resources and are struggling in taking the critical first step of their own smart manufacturing journey. They need systematic assessment methodologies to better understand their needs and build a solid foundation for their own, value-adding SMS. [4, 5]

Several smart manufacturing and Industrie 4.0 manufacturing system assessment methods and capability maturity models are available. These methods and models evaluate a manufacturing enterprise from various perspectives with differing applicability to various sizes and maturity of the enterprise. There are also still limited wealth of information supporting the usages and derived values of these methods and models. SMMs typically lack resources to sort through them. Using outside consultants can also be cost prohibitive and too risky for many SMMs.

This paper introduces a smart manufacturing improvement planning system that aims to systemize smart manufacturing consulting services, to be easily utilized by manufacturing companies. In the next section, related works are provided. Then, an overview of the platform is given, followed with discussion of the future work and finally conclusion.

2 **Related Works**

Over the past decade, frameworks or models for business and process maturity assessment have been developed within the manufacturing sector. Technology readiness level (TRL) [7] represents the maturity of a technology for commercial adoption. Similarly, manufacturing readiness level (MRL) [8] reflects the characteristics of manufacturing process technology. Supply chain readiness level (SCRL) [9] provides a way to assess the ability of the supply chain to operate and to achieve specific operational performance goals. It is associated with characteristics within fifteen categories that discretely provide an improvement roadmap for design and operation of a supply chain.

More recently as the new industrialization paradigms such as smart manufacturing and Industrie 4.0 have been developed, models for respective readiness and maturity assessments have also been proposed. Weber et al. [10] presented a model for data driven manufacturing in the context of Industrie 4.0 that consists of six maturity levels (0-Nonexistent IT Integration, 1-Data and System Integration, 2-Integration of Cross-Life-Cycle 3-Data and Service-Orientation, 4-Digital Twin, 5-Self-Optimizing Factory). Lee et al. [11] developed an assessment framework in the context of smart factory that consists of Leader-ship, Process, System and Automation, and

2

Performance Criteria. In addition, they introduced the effects based on 20 case studies of South Korean SMEs.

Carolis et al. [12] introduced the Digital REadiness Assessment MaturitY (DREAMY) model that focuses more on the digitalization of an enterprise. It aims to assess a manufacturing enterprise readiness level for the digital transformation and to develop a roadmap for prioritizing investments. Smart manufacturing system readiness level (SMSRL) [6] was developed by the National Institute of Standards and Technology (NIST) as an index that measures a manufacturing company's readiness for improving manufacturing system performance using data intensive smart manufacturing technologies. It measures the readiness in four aspects including ICT, performance management, organization, and information connectivity. Li et al. [13] introduced an assessment model using MESA MOM Capability Maturity Model (CMM) and suggested a way to reduce the time and resources consumption for applying the CMM to manufacturing enterprises by replacing the lengthy yes/no questionnaires with multiple-choice and providing various kinds of improvement strategies.

Although various smart manufacturing assessment models are being developed, their applicability issues remain, particularly regarding SMMs special requirements. SMMs typically need trainings to conduct assessments internally. Alternatively, consulting services may be used. Both options add more loads to the SMMs scarce resources. For these reasons, SMEs need a clear value proposition and a mechanism to evaluate the trustworthiness of consulting services.

To solve these problems, it is necessary to develop an assessment platform that enables manufacturers to either conduct a self-assessment with a very intuitive, stepby-step user interface, or provide access to assessment models and communityevaluated consulting services. The platform should allow users to develop subsequent improvement plans and share success values.

3 **Factory Assessment and Improvement System**

3.1 **FAISYS Architecture**

This assessment system is being developed based on the MEAN stack (MongoDB, Express.js, Angular 4, and Node.js). It will be an open source, web-based, free selfassessment system for manufacturing companies. It will provide a platform to host various SMS assessment methodologies [e.g., 6, 10, and 11], reflecting the diversity of manufacturing companies and their individual, often domain specific requirements. Manufacturing enterprises can intuitively compare their current smart manufacturing readiness to reference models and peers, adopt improvement strategies, get technologies and services recommendations that have been community-vetted, and generate improvement roadmap systemically on the platform.

Fig. 1 shows the system architecture. Professionals can add assessment methods into FAISYS by a convenient procedure. A bundle that consists of a detailed description, questionnaires, computation logics, and charts are defined using a standard XML schema shown in Fig. 2 (a). The mounting of an assessment model is completed

3

uploading XML files using the neutral language interface of FAISYS. Manufacturing users select one of the assessment models to conduct their assessment as shown in Fig. 2 (b). Intuitive charts indicating the current status are provided as shown in Fig. 2 (c). Improvement plans are automatically presented based on the assessment results. Administrators can analyze accumulated data and generate industry, regional, or national level benchmarking reports without revealing individual manufacturer's identity.

4



Fig. 1. Architecture of FAISYS



Choi, Sangsu; Kulvatunyou, Boonserm; Wuest, Thorsten. "Towards a Platform for Smart Manufacturing Improvement Planning." Paper presented at APMS 2018 International Conference Advances in Production Management Systems (APMS 2018), Seoul, Republic of Korea. August 26, 2018 - August 30, 2018.

(a) XML Schema Material and energy control The functions of materials and energy control typically include: a) managing inventory, transfers, and quality of material and energy materials and energy based on short- and long-term requirements; c) calculating and reporting inventory balance and losses of raw receiving incoming material and energy supplies and requesting quality assurance tests; e) notifying purchasing of accepted material receiving incoming material and energy supplies and requesting quality assurance tests; e) notifying purchasing of accepted material receiving incoming material and energy supplies and requesting quality assurance tests; e) notifying purchasing of accepted material receiving incoming material and energy supplies and requesting quality assurance tests; e) notifying purchasing of accepted material receiving incoming material and energy supplies and requesting quality assurance tests; e) notifying purchasing of accepted material receiving incoming material and energy supplies and requesting quality assurance tests; e) notifying purchasing of accepted material receiving incoming material and energy supplies and requesting quality assurance tests; e) notifying purchasing of accepted material and energy supplies and requesting the supplication of the supp (1) Is this task being performed? O Yes I do not know (2) Is there any expert in charge of this task? Yes
 No
 I do not know (3) Have related performance indicators been defined for decision making? O Yes O I do not know (4) Are the performance indicators being managed O Yes I do not know (5) Are the performance indicators based on international standards? O Yes Ó № O I do not know (b) Questionnaire \leftrightarrow \rightarrow C \triangle (1) localhost:4200/#/self-assessment/smsrl-assessment

Home Smart Manufacturing Sessment Methodology (20) Self-Assessment v (3) About

1. Smart Manufacturing Maturity Index



Fig. XML Schema and FAISYS Functions

3.2 Assessment Methodologies

Two assessment methods are currently equipped in FAISYS. The first model is SMSRL [6] that seeks to provide an evaluation of how manufacturers

manage/improve their performances and the available infrastructure to support related activities. SMSRL method is based on a formal reference model for factory improvement activities called FDI [14] and evaluates manufacturers in four aspects including organizational maturity, performance management maturity, IT maturity, and information connectivity maturity as shown in Fig. 3. Consolidating these four categories into a single measure creates a maturity index reflecting a manufacturer's readiness to deploy smart manufacturing solutions or to participate in a network of smart manufacturers.



Fig. 3. Four aspects of SMSRL

The second unique methodology is developed by WVU's Smart Manufacturing Lab and focuses specifically on the needs of manufacturing SMEs. The WVU assessment and maturity model is unique in several aspects: i) it reflects the lower baseline several manufacturing SMEs have when it comes to Smart Manufacturing (adding a 'level 0'), ii) it merges a readiness index with a maturity model, and iii) integrate building blocks and toolkits that can be combined in a modular fashion based on the individual maturity and assessment outcome. The model will be set around five organizational dimensions and five maturity levels. The model is informed by a recently conducted in-depth study of the current state of the art of Industry 4.0 / Smart Manufacturing maturity/ assessment/ readiness models and the learning from WVU's recent SME study [3].

3.3 Discussion

FAISYS will be developed as a scalable open source platform with a high degree of openness and collaboration as shown in Fig. 4 [15]. Manufacturing users will intuitively compare different SMS technologies and services on the platform. Suitable solutions are matched with individual manufacturers requirements by a machine learning algorithm. The machine learning will use data from a peer review system and success stories associated with the past technologies and services adoptions. Thereby the effectiveness of suggestions will improve with scale and over time. FAISYS will allow manufacturers to assess their smart manufacturing level for free by themselves. Guidelines including tools, advice and counsel on how to invest wisely will be provided separately by industry.

6



Fig. 4. Concept of Platform

FAISYS services will sustain by charging advertising and/or transaction fees from trainings, consulting services, technologies offered on the platform. They will be propelled by developments and standardizations of assessment methodologies and increasing utilizations.

4 Conclusions

A smart manufacturing assessment system called FAISYS for manufacturing enterprises, especially SMMs, is being developed. FAISYS will be released in the near future as an open source platform. Through this platform, technologies, knowledge and information related to SMS will be widely shared. It will also provide appropriate assessment methodologies and best practices for various industries and national users. Accumulated big data from manufacturing users will be analyzed, and it can be used for developing appropriate government support policy and SMS technology roadmap. This platform will help manufacturing companies improve SMS understanding and build their own smart factories. Creating a coherent suite of assessment methods or showing their differences and relationships remain a challenge. In our future work, we plan to investigate the possibility of using the three smart manufacturing dimensions, namely smart factory, value chain, and digital thread proposed by [16] as a guidance.

Disclaimer

Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST.

7

References

- 1. Thoben, K. D., Wiesner, S., Wuest, T.: Industrie 4.0 and Smart Manufacturing-A Review of Research Issues and Application Examples, International Journal of Automation Technology 11(1) (2017).
- 2. Kang, H. S., Lee, J. Y., Choi, S., Kim, H., Park, J. H., Son, J. Y., Noh, S.D.: Smart manufacturing: Past research, present findings, and future directions. International Journal of Precision Engineering and Manufacturing-Green Technology 3(1), 111-128 (2016).
- 3. Wuest, T., Schmid, P., Lego, B. & Bowen, E.: Overview of Smart Manufacturing in West Virginia. Bureau of Business & Economic Research West Virginia University.
- 4. Sommer, L.: Industrial revolution-industry 4.0: Are German manufacturing SMEs the first victims of this revolution? Journal of Industrial Engineering and Management, 8(5), 1512 (2015).
- 5. Moeuf, A., Pellerin, R., Lamouri, S., Tamayo-Giraldo, S., Barbaray, R.: The industrial management of SMEs in the era of Industry 4.0. International Journal of Production Research, 56(3), 1118-1136 (2018).
- 6. Jung, K., Kulvatunyou, B., Choi, S., Brundage, M. P.: An Overview of a Smart Manufacturing System Readiness Assessment. In IFIP International Conference on Advances in Production Management Systems, pp. 705-712. Springer, Cham (2016).
- 7. Dubos, G. F., Saleh, J. H., and Braun, R.: Technology readiness level, schedule risk, and slippage in spacecraft design. Journal of Spacecraft and Rockets 45(4), 836-842 (2008).
- 8. Ward, M. J., Halliday, S. T., and Foden, J.: A readiness level approach to manufacturing technology development in the aerospace sector: an industrial approach. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 226(3), 547-552, (2012).
- 9. Tucker, B., and Paxton, J.: SCRL-model for human space flight operations enterprise supply chain. In Aerospace Conference 2010, pp.1-9. IEEE (2010).
- 10. Weber, C., Königsberger, J., Kassner, L., Mitschang, B.: M2DDM-A Maturity Model for Data-Driven Manufacturing. Procedia CIRP, Vol.63, pp.173-178 (2017).
- 11.Lee, J., Jun, S., Chang, T. W., Park, J.: A Smartness Assessment Framework for Smart Factories Using Analytic Network Process. Sustainability, 9(5), 794 (2017).
- 12.De Carolis, A., Macchi, M., Negri, E., Terzi, S.: A maturity model for assessing the digital readiness of manufacturing companies. In IFIP International Conference on Advances in Production Management Systems, pp.13-20, Springer, Cham (2017).
- 13.Li, Q., Brundage, M., Kulvatunyou, B., Brandl, D., Noh, SD.: Improvement Strategies for Manufacturers Using the MESA MOM Capability Maturity Model. In IFIP International Conference on Advances in Production Management Systems, pp. 21-29, Springer, Cham (2017).
- 14.Jung, K., Choi, S., Kulvatunyou, B., Cho, H., Morris, K. C.: A reference activity model for smart factory design and improvement. Production Planning & Control 28(2), pp.108-122 (2017).
- 15.Menon, K., Kaerkkaenen, H., Wuest, T.: Role of openness in industrial internet platform providers' strategy. PLM 2017, Seville, Spain (2017).
- 16.Leiva, C. (February 2018). Three perspectives converge on smart manufacturing. Internet article available at http://advancedmanufacturing.org/17797-2/.

Choi, Sangsu; Kulvatunyou, Boonserm; Wuest, Thorsten. "Towards a Platform for Smart Manufacturing Improvement Planning." Paper presented at APMS 2018 International Conference Advances in Production Management Systems (APMS 2018), Seoul, Republic of Korea. August 26, 2018 - August 30, 2018.

Stutzman, P.E.

Reference Materials and Proficiency Testing for Validating X-Ray Powder Diffraction Measurements of Cement Composition

1. Introduction

Estimates on the phase content of cement for process control and specification compliance using the Bogue calculations have been used for the past 75 years. The Boque method is an indirect approach based upon a bulk chemical analysis and assumptions of the chemistry of the four principal phases. Bias inherent in the calculations resulting from assumptions of composition and from analytical uncertainties have limited their use in accurately characterizing mineralogical compositions [1,2]. X-ray powder diffraction analysis has become more popular within the industry, providing a direct analytical method for a more complete and potentially more accurate estimate of phase compositions of clinker and cement.

Qualitative and quantitative analysis by X-ray powder diffraction is a complicated analytical procedure, particularly for labs not well acquainted with the analysis procedures, cements, or both. The need for a standardized analytical procedure and some means to validate individual laboratory performance was recognized by interlaboratory studies [3,4,5,6] and became a goal of the ASTM C01 Committee on Cement.

In 1978, ASTM C1.23 Subcommittee on compositional analysis initiated a task group (TG) on X-ray powder diffraction (XRD) to develop a test method to improve the precision and bias [3,4,5,6]. Initial work used sets of calibration mixtures, calibration curves, and compounded test mixtures using an internal standard-based analysis [7,8,9]. Peak intensity measurements remained a source of inconsistency in test results between laboratories [10] leading some to conclude that XRD results were not much better than the Bogue estimates [3,6].

The application of the Rietveld method to XRD analysis of cements [11,12] has dispensed the need to develop complicated calibration curves for each phase, using crystal structure models and the entire diffraction pattern to estimate individual phase intensities. While this has made the measurements more consistent within and between laboratories [13], the inherent data normalization in the Rietveld method can pose a hidden problem. This constraint will obscure bias, if present, and apportion bias across all phases relative to their concentration. So, while this approach has successfully addressed the challenges of reference standards and phase intensity measurement, it remains possible to get a suitable-appearing refinement with a highly biased result. Standard test methods and test validation remain necessary to provide assurance of a suitable analysis.

This paper will present a discussion on the development of a standard test procedure using X-ray powder diffraction for cements, the development and role of certified reference clinkers for phase analysis and, proficiency testing for validating laboratory performance that may encompass a wider range of materials.

Stutzman, Paul. "Reference Materials and Proficiency Testing for Validating X-Ray Powder Diffraction Measurements of Cement Composition." Paper presented at 20th International Conference of Building Materials, Weimar, Germany. September 12, 2018 - September 14, 2018.

2. Standard Reference Material (SRM) Clinkers

The initial interlaboratory studies by the ASTM XRD TG involved sets of compounded mixtures of lab-synthesized phases to serve as calibration mixtures and a second set to serve as validation mixtures. A longer-term goal of the ASTM TG was to develop a well-characterized set of industrial clinkers to facilitate the test development. These reference materials (RM) were first made available in the mid-1980's [14] with reference values based on light microscopy point count analysis. Subsequently, an additional data set by X-ray powder diffraction analysis provided the second independent analytical method to establish certified reference values and uncertainties [15].

Three SRM clinkers, SRM 2686, 2687, and 2688 were developed as certified reference materials. They exhibit a range of textures and phase abundance and are intended to be used in developing and testing quantitative methods for cement analysis, primarily microscopy and X-ray powder diffraction. SRM2686a (the successor to SRM 2686) has an intermediate grain size, heterogeneous phase distribution and periclase disseminated throughout. SRM2687a has a fine-grained, moderately heterogeneous texture. SRM2688 is coarse-grained with uniformly-distributed silicates and matrix (Figure 1).

3. Test Method Development

The initial set of interlaboratory trials for quantitative analysis of the interstitial phases (aluminates, ferrite, and periclase) [7,8] used compounded test mixtures and later the Reference Clinkers to assess precision and bias. Using selective extractions to concentrate the interstitial phases, precisions of 1.1 % within-lab 1 σ (repeatability) and 1.7 % between-lab 1 σ (reproducibility) were calculated on a whole clinker basis. The mean values from the participants closely mirrored the known values, leading to a conclusion of no apparent bias.

The efforts of the task group culminated in the first standard test method for X-ray powder diffraction analysis of cements, ASTM C 1365, Standard Test Method for Determination of the Proportion of Phases in Portland Cement and Portland-Cement Clinker Using X-ray Powder Diffraction Analysis [16,17]. ASTM C1365 is a performance-based standard. While most standard test methods use an explicit protocol, this test method allows flexibility in lab procedure but requires that a set of performance criteria must be demonstrated via analysis of certified reference materials.

The test method was initially based upon the internal standard method using calibration curves but the performance-based approach easily accommodated growing popularity of the Rietveld method. The adoption of the Rietveld method however has presented one new shortcoming. Inherent in the calculations is a data normalization that may obscure bias. In the traditional calibration curve approach that assesses each phase individually, having a data set sum to less than or greater than 100 % indicates some problem with the data or the calibrations that needs additional investigation. While the Rietveld approach has improved consistency, differences between laboratory procedures may result in significant disagreement, so the need to demonstrate appropriate test performance remains with the SRM clinkers and proficiency testing using cements providing a means to qualify laboratory performance.

4. Measurement Uncertainty in X-ray Powder Diffraction Measurements of Cements

XRD measurements of hydraulic cements, like any measurement process, are subject to a random and a lab-specific systematic bias (error) [18]. Four factors contribute to testing variability: 1) the operator, 2) the equipment, 3) instrument calibration, and 4) the testing environment [19]. For example, a new operator may not have as much experience in phase identification and quantitative analysis, one instrument may be lower-powered than others, counting statistics (count time, detector design), specimen grinding and mounting procedures, structure models, and analysis codes. While not an exhaustive list, these variables are relatively constant within a lab, imparting some measurement variability between specimen replicates. However, they will generally vary between labs, resulting in larger differences in between lab test results.

Precision is defined within ASTM as "the closeness of agreement among test results obtained under prescribed conditions" [19]. Precision is expressed as repeatability (within-laboratory, designated s_r) which excludes the four factors and reproducibility (between-laboratory, designated s_R), which includes the four factors. Both measures are expressed as standard deviations of replicate analyses. From the s_r and s_R , the maximum difference (d2s) between two test results may be calculated by multiplying the respective standard deviation by 2.77 to obtain the 95 % repeatability (r) and reproducibility (R) limits [19].

These measures form the basis for qualification in ASTM C1365 using a certified reference material portland cement clinker. The qualification provides both the lab and the user of the test results with a degree of confidence that the test can be successfully accomplished. This is particularly important now, where as mentioned earlier, the normalization in the Rietveld calculation may obscure bias.

An alternate validation approach, proficiency testing, provides a means to test method and lab performance using a wider variety of materials. For cements analyses, laboratories provide a single analysis on each of two cements using a consensus for phase abundance and comparison to the ASTM C1365 precision qualification criteria. Both of these approaches will be illustrated.

In the first example, NIST SRM clinker 2688 has been analyzed by 11 individual laboratories using X-ray powder diffraction. Plots of their results are shown in Figure 2 with the three replicate determinations for each phase. The blue box here encompasses the $\pm 1\sigma$ reproducibility limits for information while the green box represents the qualification criteria, $\pm d2s$, the reproducibility limits. The displacement between individual replicate determinations for each lab represents the repeatability and the displacement of the triplicate clusters between labs represents reproducibility. While not illustrated in the plots to maintain simplicity, the d2s limits for repeatability are coincidentally approximately the width of the blue box. Most labs exhibit a similar repeatability, which generally falls within method limits.



Figure 1. Standard Reference Material (SRM) Clinkers 2686a (top), 2687a (middle) and 2688 exhibit a range of textures and compositions and are used for developing and testing methods for quantitative phase analysis.

Greater differences are seen between lab determinations and, in some cases, one may speculate as to the sources of lab bias. In the cases of labs 6 and 11, for example, there appears to be a correlation between the alite and belite test results. These results may reflect preferred orientation of the silicates and the difficulty in estimating silicate scale factors after orientation corrections have been applied.

So, for a lab to qualify using this SRM, their triplicate results (repeatability) should not be displaced by more than the width of the blue box and their test result must remain within the green box. While both labs 6 and 11 fall within the d2s bounds for alite, they fail due to the bias in belite determination. In some cases where lab repeatability is relatively large, bias may be reduced by establishing a test result as a mean of replicate measurements. This can be imagined in the case where lab 1 has its first aluminate determination on the limit and the third exceeding the limit. By re-defining a test result as the mean of n=3 replicates, they would qualify. By doing so, labs 1, 3, and 5 would be able to qualify their ferrite determinations. Alternatively, they could assess their test procedures to determine and resolve the source of the bias.

While the SRM clinkers have been very useful both in test method development and for validating lab performance, there is a need for reference materials that cover the more complicated phase assemblage of cements. While certified reference cements for phase abundance could be developed, proficiency testing may be a more practical approach.

4. Application of Proficiency Testing to Assess Laboratory Bias

Proficiency testing is a development based upon interlaboratory programs to evaluate test methods and laboratories using paired samples for chemical analyses of rocks and cements by W.J. Youden [20,21,22]. This approach has been used to assess test precision and to evaluate laboratory performance.

Samples of two different, but similar, cements are distributed to a number of laboratories. Each laboratory reports a single test result for each cement phase identified for each of the two cements. A consensus value for each analyte of each cement is taken as the median value of the data. The median is used because of its insensitivity to outliers. Consensus values often mirror the reference values when a large number of labs provide estimates. In the case of alite for SRM 2688 discussed earlier, the certified value of 66.1 % is closely approached at 65.1 % using the median of the eleven labs.

In the following examples, instead of estimating test precision from the data, we apply existing ASTM performance criteria with the Youden graphical display. This allows a visual assessment of the individual and collective test results to assess compliance of a lab, and to provide some insight on systematic and random error in individual test results.


Figure 2. SRM 2688 measurement results from 11 laboratories for alite, belite, aluminate, and ferrite with SRM certificate value indicated by a red line and reproducibility qualification limit boxes for the 95 % limits in green and 1σ in blue.

Stutzman, Paul. "Reference Materials and Proficiency Testing for Validating X-Ray Powder Diffraction Measurements of Cement Composition." Paper presented at 20th International Conference of Building Materials, Weimar, Germany. September 12, 2018 - September 14, 2018.

In the following cement example, a test result is defined as the mean of triplicate determinations. Key assumptions are that repeatability precision from lab to lab is similar and each lab's systematic bias (error) affects both of their test results similarly. A Youden scatter plot is prepared where each point represents results of an analyte from the pair of cements. A consensus value is established by taking the median of the data, as the median is less sensitive to outlying results, and plotted as intersecting lines at right angles to each other.

If the test method is not biased, the points will form a roughly circular pattern centered on the intersection of the median lines. Systematic bias is seen as points falling along the 45 degree diagonal away from the median intersection. Random bias is seen in the degree to which points fall off the 45 degree diagonal line, as points falling in the upper-left and lower-right quadrants have opposite signs to their individual estimates. ASTM E2489 provides a means to calculate test precision using paired sample or single sample data [22].

Using the ASTM C1365 qualification criteria and the Youden plot format, we get Figures 3 through 5. These plots have been augmented with graphical features to denote test method qualification criteria. First, the median lines and 45 degree line segment lengths represent the \pm d2s, or 95 % limits for reproducibility. The circles represent \pm 1 σ and \pm 2 σ from the median consensus values for user information since to qualify, lab test results must be within the \pm d2s limits. An alternate approach to qualify labs was proposed by Youden [21], with lab performance based upon their distance from the consensus as \pm 1 σ , \pm 1.5 σ , \pm 2.5 σ , >2.5 σ being assigned values of 4 through 0, respectively, along with a sign of + or – if their bias was high or low. For multi-analyte materials like cements, the results are averaged to establish their overall performance score.

5. Observations on the Youden Plots

As an anonymous test program, each lab is aware of their identifying code but not that of any other lab. Plotting of the collective results allows one to quickly compare their reported results against the others and the consensus. The location of their x-y point for each phase reveals both systematic and random error of their testing procedures. Recall that the current reproducibility d2s limits are indicated by the lengths of the median and 45 degree segments. The inner circle $(\pm 1\sigma)$ encloses the best-performing labs and the outer circle the well-performing labs $(\pm 2\sigma)$. Labs 3 and 9 exhibited the greatest difficulties in their analyses whereas labs 2 and 22 generally performed within the $\pm 2\sigma$ level.

For the alite plot in Figure 3, lab 9 (dark triangle) falls outside the d2s limits, but roughly along the 45 degree line indicating a strong positive bias. Labs 7 (tan triangle) and 10 (blue diamond) exhibit a moderate negative and moderate positive bias, respectively however, they remain within the d2s limits. In contrast, lab 6 (brown +) is located in the lower left quadrant, reflecting a random error. Lab 6 appears to have had difficulty with Cement A, with a low estimate of alite for Cement B, close to the consensus. The apparent correlation between alite and belite estimates are reflected in the Belite Youden plot where lab 7 is positively biases and lab 10 negatively biased. In addition, labs 4 and 12 show a large systematic bias while lab 20 shows a large random bias. Given alite has a propensity to preferentially orient and the diffraction patterns of alite and belite are quite similar, some labs may be having difficulties with

these effects on the determination of appropriate scale factors that are part of the quantitative estimates.

Laboratory 9 also had significant negative systematic bias for aluminate and ferrite while labs 3 and 11 show significant positive aluminate bias and significant negative ferrite bias for lab 3 and a larger random bias with a low ferrite estimate for cement A. Laboratory 3 also exhibits a large positive bias for periclase as seen in Figure 4.

The calcite plot in Figure 5 shows the elongation of data points reflecting a greater propensity for systematic bias in its phase estimates. This may stem from the propensity for preferred orientation for calcite or from its strong overlap with the alite diffraction pattern. Labs 9 and 3 have systematically high estimates while labs 10 and 12 are systematically low. Lab 5 estimated the calcite content of Cement A close to the consensus but either forgot to include calcite for cement B or had a large random bias component in its test procedure.

In some cases a cement phase is present in only one of the two cements. In these cases, a jittered dot diagram may be used to depict the results along with the consensus value (Figure 6). Labs 9 and 3 performed quite well in their estimation of gypsum, falling well within the d2s values. Labs 12, 10, and 17 significantly underestimate gypsum relative to the consensus while lab 5 significantly overestimates gypsum concentration. Gypsum is another phase that is very prone to preferred orientation, which is likely to affect its phase estimates. These outlying results suggest that these lab's specimen preparation methods and any use of orientation corrections merit investigation.

The ASTM C1365 qualification performance criteria are based on a more exhaustive inter-laboratory study involving eleven competent labs and four cements compounded from the SRM clinkers. Performing within the d2s limits or better ensures that a lab is performing at the level comparable to those in the original study. These data should be revised over time as labs become better acquainted with quantitative X-ray diffraction analysis of cements and as instrumentation and processing software improve. Given estimates of repeatability and reproducibility standard deviations can be made from the paired sample and the single sample study, these performance criteria can evolve [22]. The current qualification criteria are on an individual phase basis. The scoring system proposed in [21] would be useful in routine proficiency testing where a running average of all test results would allow labs the ability to examine their test procedures without being excluded for one bad test result.

Zero-value data create some problems for proficiency testing if repeatability and reproducibility values are being estimated. In this example labs did not designate if a zero value was the result of an actual zero determination or that they did not include that phase in their analysis. The test procedure needs evaluation criteria to establish the presence or absence of a phase so it is clear that the value is a zero value rather than a phase that was not identified and therefore left out of the analysis. Selective extraction for the non-silicate phases, an initial qualitative analysis, or requiring a visual check of the refined phase patterns against the cement pattern are all useful methods to clearly establish the presence or absence of a phase. Finally, the effects of preferred orientation on phase estimates appears to be necessary, and recommendations on specimen preparation and the application or orientation corrections should be established.



Figure 3. Youden plots for alite and belite provide graphical feedback on lab performance for both systematic and random bias.

SP-636



Figure 4. Youden plots for aluminate and ferrite provide graphical feedback on lab performance for both systematic and random bias.



Figure 5. Youden plots for alite and belite provide graphical feedback on lab performance for both systematic and random bias.



Figure 6. A jittered dot diagram with lab numbers used in place of dots provides a rapid means to illustrate individual lab performance against the consensus for a phase that occurs in only one of the two cements.

6. Conclusions

Validation of quantitative X-ray powder diffraction measurements traditionally involved compounding test mixtures of known phase compositions. Validation is particularly important with the Rietveld method as the normalization will obscure bias. ASTM C1365, Standard Test Method for Determination of the Proportion of Phases in Portland Cement and Portland-Cement Clinker Using X-Ray Powder Diffraction, employs a laboratory qualification using replicate determinations of certified reference materials, such as the NIST SRM clinkers. The performance metrics for repeatability and reproducibility precision and for bias in ASTM C1365 are based on multiple interlaboratory studies and standardized procedures for estimating precision. An alternate validation approach can be found in a modified routine proficiency testing, where a group of laboratories provide a single analysis on each of two cements using the ASTM C1365 gualification criteria. Quantitative feedback provided in graphical format allow participants to assess results for random and systematic error. Proficiency testing of this sort provides participating laboratories with a quantitative assessment of their performance relative to peers using a wider range of materials encompassing the broad spectrum of modern hydraulic cement production.

References

¹ H.F.W. Taylor, *Cement Chemistry*, 2nd ed., Thomas Telford, London, 1997.

² P. Stutzman, A. Heckert, A. Tebbe and S. Leigh, Uncertainty in Bogue-calculated phase composition of hydraulic cements, Cem. Conc. Res., 61 (2014) 40-48.

³ A.E. Moore, Comparison of the Results Obtained for the Compound Composition of Portland Cements by X-Ray Diffraction, Microscopy, and Wet Chemical Methods, Analysis of Calcareous Materials, S.C.I. Monograph No. 18, pp. 372-390, 1964. ⁴ A.E. Moore, *Mineralogical Analysis of Portland Cement by X-Ray Diffraction*, Proc. Soc. Chem. Ind., 18-20, 1970.

⁵ L.P. Aldridge, International Cement Analysis Study. Part 3. Quantitative X-ray Diffraction Analysis of Cements, Report No. C.D. 2201, Department of Scientific and Industrial Research, New Zealand, 1975, 20 pp.

[&]quot;Reference Materials and Proficiency Testing for Validating X-Ray Powder Diffraction Measurements of Cement Composition." Paper presented at 20th International Conference of Building Materials, Weimar, Germany. September 12, 2018 - September 14, 2018

¹² J.C. Taylor and L.P. Aldridge, Full-profile Rietveld quantitative XRD analysis of Portland cement: Standard XRD profiles for the major phase tricalcium silicate (C3S: 3CaOSiO₂), Powder Diffraction 8 (3), September, 1993. pp. 138-144.

¹³ P. Stutzman and S. Leigh, *Phase Analysis of Hydraulic Cements by X-Ray Powder* Diffraction: Precision. Bias. and Qualification. Journal of ASTM International. Vol. 4. No. 5, 11 pp., 2007

¹⁴ H.M. Kanare. Production of Portland Cement Clinker Phase Abundance Standard Reference Materials, Final Report, Construction Technology Laboratories, December, 1987, 32 p.

¹⁵ P. Stutzman and S. Leigh, *Phase Composition of the NIST Reference Clinkers by* Optical Microscopy and X-ray Powder Diffraction, NIST Technical Note 1441, 50 pp., 2002

¹⁶ ASTM C 1365-06, Standard Test Method for Determination of the Proportion of Phases in Portland Cement and Portland –Cement Clinker Using X-Ray Powder Diffraction Analysis, ASTM Book of Standards, Vol. 4.01, ASTM International, West Conshohocken, PA.

¹⁷ P. Stutzman and S. Leigh, Phase Analysis of Hydraulic Cements by X-Ray Powder Diffraction: Precision, Bias, and Qualification, Journal of ASTM International, Vol. 4, No. 5, paper ID JAI101085, 11 pp., 2007

¹⁸ W.J. Youden, Statistical Techniques for Collaborative Tests, Association of Official Analytical Chemists, Inc., 1973 64 pp.

¹⁹ ASTM E 691, Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method. in Annual Book of ASTM Standards. Vol. 4.01, 2001. ASTM International, West Conshohocken, PA

²⁰ Youden, Evaluation of Chemical Analyses on Two Rocks, Technometrics, Vol. 1, No. 4., November, 1959, pp. 409-417.

²¹ W.J. Youden, Statistical Aspects of the Cement Testing Program, Proc. ASTM, Vol. 59, 1959, pp. 1120-1128.

²² ASTM E2489, Standard Practice for Statistical Analysis of One-Sample and Two-Sample Interlaboratory Testing Programs, Annual Book of ASTM Standards, ASTM International, West Conshohocken, PA, 2016

> Paul Stutzman paul.stutzman@nist.gov Engineering Laboratory National Institute of Standards and Technology

Author:

Stutzman, Paul. "Reference Materials and Proficiency Testing for Validating X-Ray Powder Diffraction Measurements of Cement Composition." Paper presented at 20th International Conference of Building Materials, Weimar, Germany. September 12, 2018 - September 14, 2018

⁶ L.P. Aldridge, International Cement Analysis Study. Part 4: Comparison of Results, Report No. C.D. 2267. Department of Scientific and Industrial Research. New Zealand. 1978, 18 pp.

⁷ Struble and H. Kanare, *Cooperative Calibration and Analysis of Cement Clinker* Phases, NIST IR 89-4164, 34 pp., 1989

⁸ L.J. Struble, Cooperative Calibration and Analysis of Cement Clinker Phases, Report 2, Civil Engineering Studies, Structural Research Series No. 556, University of Illinois, 21 pp., 1990

⁹ L.J. Struble. Quantitative Phase Analysis of Clinker Using X-Ray Diffraction. Cement. Concrete, and Aggregates, Winter, 1991, pp. 97-102.

¹⁰ G. Yamaguchi and S. Takagi, *The Analysis of Portland Cement Clinker*, Principal paper I-3, Fifth International Symposium on the Chemistry of Cement, Tokyo, pp. 181-216.

¹¹ H. Möller, Standardless quantitative phase analysis of Portland cement clinkers, World Cement, September, 1995, pp. 75-84.

MEASUREMENT OF THERMAL PROCESSING VARIABILITY **IN POWDER BED FUSION**

Jarred C. Heigel¹, Eric Whitenton¹ ¹Engineering Laboratory National Institute of Standards and Technology² Gaithersburg, MD, USA

INTRODUCTION

Laser powder bed fusion (LPBF) is an additive manufacturing (AM) technology used to manufacture high-value metal parts. The layerby-layer nature of the process allows complex geometries and internal features, such as conformal cooling channels, to be realized. For the technology to reach its full potential, it must be capable of manufacturing precision parts. However, the precision of parts fabricated using LPBF is currently orders of magnitude worse than what can be achieved using machining [1]. This is due, in part, to the distortion that arises from thermally induced residual stresses [2]. While distortion can be compensated for and/or corrected through post-processing, this is costly and not always possible for internal features.

An additional challenge that LPBF and other metal AM technologies face is that in addition to creating the part geometry, the material is also created during the process. The melting of the metal feedstock powder using a laser and the rapid cooling of the solidified material, as well as the cyclic re-heating and cooling from adjacent tracks and subsequent layers, has a significant impact on microstructure and material properties [3]. Unfortunately, the material of a LPBF manufactured part is inhomogeneous due to the inconsistency of the thermal processing throughout the part. The imprecision arises from a variety of factors, including the scan strategy, processing conditions [4], and geometric effects, such as the reduced ability to conduct heat away from the melt pool when creating overhangs.

the impact Considering on distortion. microstructure, and material performance, the thermal history of the processed material must be understood before precision geometries or materials can be manufactured using LPBF technologies. The objective of this work is to present preliminary results from in situ thermographic measurements acquired during the build of a complex part and to assess the variability of the thermal processing in response to scan strategy and geometry effects.

EXPERIMENT SETUP

FIGURE 1 shows the experiment setup. A commercial powder bed fusion system is modified to allow a high-speed short-wave infrared (SWIR) camera to observe a small region of the build plane. The camera acquires images at a rate of 1800 frames per second and is sensitive to radiant temperatures from 550 °C to 1050 °C. Details on this setup can be found in [5].



FIGURE 1 - The powder bed fusion system with the infrared camera mounted to it.

2018 ASPE and euspen Summer Topical Meeting Volume 69

² Official contribution of the National Institute of Standards and Technology (NIST); not subject to copyright in the USA. The full descriptions of the procedures used in this paper require the identification of certain commercial products. The inclusion of such information should in no way be construed as indicating that such products are endorsed by NIST or are recommended by NIST or that they are necessarily the best materials, instruments, software, or suppliers for the purposes described.

Heigel, Jarred; Whitenton, Eric.

FIGURE 2 shows the part used in this study, manufactured of nickel-based super alloy 625 (IN 625). FIGURE 3 presents a schematic of the part. The part is designed to be small enough to fit entirely in the field of view of the camera, which is approximately 12 mm x 6 mm. It has a hole on one side and a 45° overhang on the other. The part is manufactured on a 100 mm x 50 mm IN 625 substrate that is 12.7 mm thick. The substrate is bolted on the LPBF build platform.



FIGURE 2 - The part used in this study is small enough to fit in the SWIR camera's field of view.



FIGURE 3 – Part schematic. Dimensions in mm.

The part is manufactured in 249 layers. Each layer is 20 µm thick and consists of three steps: a pre-contour scan of the perimeter of the part, a raster scan to fill in the cross section of the part (called "skin" by the manufacturer), and a postcontour scan of the perimeter. This study focuses on the thermal history generated by the raster scan. The raster scan is performed with a programmed laser power of 195 W and a programmed scan speed of 800 mm/s. The hatch spacing (distance between adjacent scan tracks) is 0.1 mm. The raster scans are contained within "stripes" that are 5 mm wide. These stripes are pre-defined within the build area and rotate 67° between layers. Multiple stripes are used in each layer to scan the entire cross section of the part.

There are four layers of interest due to the part geometry, as shown in FIGURE 4. Layer 25 is the final layer of the 9 mm long 5 mm wide rectangular base. It provides insight into the thermal history unaffected by the hole or overhang features and establishes a baseline to which the other layers can be compared. Layer 125 occurs at the midpoint of the hole where the width of the wall beside it is at a minimum (0.5 mm thick). Layer 226 is performed immediately after the hole is completed and therefore part of the laver is created over powder and not solidified material. Finally, Layer 249 is the final layer, the effects of which are evident in the top of the part (FIGURE 2).



FIGURE 4 - Illustration of the 4 layers of interest and the relative position of the IR camera.

RESULTS AND DISCUSSION

An example thermal video frame acquired during Layer 25 is show in FIGURE 5. This is one of the 1266 frames acquired during the layer. In this frame (# 617), the laser is scanning through the middle of the part in a 5 mm-wide stripe. The stripe boundary is illustrated with dashed green lines. Ejected material (spatter) is imaged in addition to the melt pool.

FIGURE 6A illustrates the radiant temperature history of a single pixel during the processing of

AMERICAN SOCIETY FOR PRECISION ENGINEERING

243

[&]quot;Measurement of Thermal Processing Variability in Powder Bed Fusion." Paper presented at 2018 ASPE and euspen Summer Topical Meeting - Advancing Precision in Additive Manufacturing, Berkeley, CA, United States. July 22, 2018 - July 25, 2018.



FIGURE 5 - A thermal video frame acquired during Layer 25. The part perimeter is illustrated with the dashed lines. The blue "x" in the center shows the location of the data in FIGURE 6.

Layer 25. This pixel correlates to the location indicated by a blue "x" in FIGURE 5. As expected, the measured radiant temperature fluctuates as the material is re-heated, and (in some cases) remelted by successive tracks. The oscillation continues as the material cools. Spatter flying between the surface and the camera causes extreme spikes.

FIGURE 6B demonstrates the calculation of the cooling rate for each pixel in a layer. For this study, cooling rate is defined as the rate the material cools from the true temperature of solidification of 1290 °C to 1000 °C. Since the SWIR camera measures radiant temperature and not true temperature, the equivalent radiant temperatures must be determined.

The radiant temperature equivalent to the solidification temperature was identified in prior work [5]. Using the same experiment setup, the solidification plateau was identified and used to measure melt pool length and to identify the solidification radiant temperature when creating laser tracks on bare substrates [5]. A subsequent study demonstrated that the presence of powder does not significantly affect the radiant temperature [7]. Therefore, radiant temperature of solidification determined from bare plate scans (942 °C) is assumed for the multi-layer build.

The radiant and true temperatures of solidification can be used to calculate the effective emissivity using the Sakuma-Hatori equations, which relate the camera signal to a calibrated temperature [8]:

$$F(T) = S = \frac{C}{exp\left(\frac{c_2}{AT+B}\right) - 1}$$
(1)



FIGURE 6 - Radiant temperature history at the center of the part in Layer 25.

$$F^{-1}(S) = T_{radiant} = \frac{c_2}{A \ln(C/S^{+1})} - \frac{B}{A}$$
(2)

where S is the camera signal, T is temperature in K, c_2 is the second radiation constant (14388 µm·K) and A, B, and C are constants established through a black body calibration. The black body calibration and the resulting constants for this work are presented in [5]. True temperature and radiant temperature are related:

$$F(T_{true}) \varepsilon = F(T_{radiant})$$
(3)

where ε is the effective emissivity of the surface. The effective emissivity of the recently solidified material is calculated using Equation 3 and the true and radiant temperatures of the solidification. This calculation yields $\varepsilon = 0.168$. Assuming this

244

2018 ASPE and euspen Summer Topical Meeting Volume 69

Heigel, Jarred: Whitenton, Eric

emissivity applies to the material at 1000 °C, the equivalent radiant temperature is 758 °C. Cooling rate is then calculated using:

$$Cooling Rate = \frac{\Delta T_{true}}{t_2 - t_1}$$
(4)

where ΔT_{true} is the temperature range of interest (in this case, 290 °C), t_7 is the time at that occurrence the pixel drops below the solidification temperature, and t_2 is first time after t_7 that the pixel equals the lower temperature range.

FIGURE 7 presents a map of the cooling rate calculated from the SWIR measurements of Layer 25. This figure shows that cooling rate varies significantly due to both scan strategy and layer shape. The typical behavior within a stripe unaffected by the layer shape is evident in the center region of the layer (outlined with white dashes). The cooling rate is lower in the center of the stripe (approximately 25 000 °C/s) and higher near the edges (more than 50 000 °C/s). This difference is attributed to the material in the middle of the stripe retaining heat longer than at the edges of the strip, decreasing its ability to evacuate the heat input from the laser and thus decreasing the rate at which the material cools. While this trend is expected to be consistent for different stripe widths (a variable that most LPBF users can control), the magnitudes may differ.

Although a typical cooling rate distribution within a stripe can be achieved, the layer shape can significantly alter the scan strategy and ultimately the cooling rates. For instance, the shape of Layer 25 prevents most of the area from being scanned with the programmed stripe width. The stripe width is effectively decreased to scan these narrower areas, and the laser more frequently rescans near a recently solidified region, thus



FIGURE 7 – Map of the cooling rate in Layer 25.

keeping the material hotter and reducing the material's ability to evacuate heat input from the laser. Additionally, the timing of the laser seems to be affected when the geometry decreases the stripe width [6], slightly increasing the time between the end of one scan track and the beginning of the next. These two factors explain the lower cooling rates in the 1st and 3rd stripes that fill the upper right and lower left corners of the layer, and the beginning and end of the 2nd stripe across the center of the layer.

Histograms are used to assess the cooling rate as a percentage of area within a layer or region of interest. FIGURE 8 compares the typical stripe (region in FIGURE 7). Laver 25. and all lavers from 1 through 25 (considered a baseline for analysis of the rest of the layers). Three observations are made from these histograms. First, within a typical stripe (in this build), two distinct cooling rates occur, where the edges of the stripe cool more approximately 3 times faster than the center of the stripe. Second, the effective decreasing of the stripe width to complete a layer increases the occurrence of lower cooling rate (comparing Layer 25 to the typical stripe). Third, Layer 25 is representative of the baseline established using all preceding layers, (similarity between the two histograms).

FIGURE 9 shows the cooling rate calculations and analysis of Layer 125, which demonstrates the impact of the narrow features on either side of the hole. The analysis reveals that the cooling rate in the narrow features increases by a factor of 2 to approximately 50 000 °C/s, whereas the cooling rate in the rest of the layer remains comparable to the baseline. These results are



FIGURE 8 - Histogram showing cooling rate as a function of area within the layer.

AMERICAN SOCIETY FOR PRECISION ENGINEERING

245



FIGURE 9 - Cooling rate measured during Layer 125. A) Measured cooling rate at each pixel in the layer, B) comparison with Layers 1 through 25.

counter-intuitive for two reasons. First, the narrow scans are performed on material that is surrounded by a greater percentage of powder, which is insulative, compared to the bulk. The higher thermal resistance would theoretically lead to slower cooling rates. Second, the stripe width in these features is much narrower than typical. It has been demonstrated earlier that this decreases cooling rate. However, the increased cooling rate in these narrow features may result from the ratio between laser-on and laser-off times decreasing. The laser shutter is open for a very short duration to scan the 0.5 mm wide features, while the time between successive laser tracks is not significantly impacted. Furthermore, the time to scan each of the narrow features is very quick, potentially impacting of material preheating which can decrease conductivity. Further experimental and modeling efforts are required confirm these hypotheses.

FIGURE 10 presents the cooling rate in Layer 226 and compares the results to the baseline. This layer occurs immediately after the completion of the hole feature. FIGURE 10A shows that while a portion of the stripe above the

hole experiences cooling rates slower than expected in a typical stripe, a small portion directly above the top of the hole experiences a higher cooling rate. The lower cooling rate is caused by the geometry, since there is less solid mass below the layer to conduct heat away, while the small region with a higher cooling rate is a result of the scan strategy, as explained below.

The thermal video of Layer 226 shows that most of the stripe is executed as expected, but the laser skips over a small area directly above the hole. After all the stripes are completed, the laser then scans the area that was skipped over. This scan after the completion of all the stripes is referred to by the manufacturer as a "down-skin" and is used for material with powder directly underneath the layer. The "down-skin" is performed with process parameters that produce a shallower melt pool, minimizing the chance of over-melting. Although these "down-skin" scans are performed with little or no solid material underneath, the cooling rate is faster for two reasons. First, the altered processing parameters used for these scans results in lower energy and less heating of the part. Second, the surrounding material to the sides of this regions is better able



FIGURE 10 - Cooling rate measured during Layer 226. A) Measured cooling rate at each pixel in the layer, B) comparison with Layers 1 through 25.

246

2018 ASPE and euspen Summer Topical Meeting Volume 69

to conduct heat away because it has had a chance to cool down from the prior stripe scans.

FIGURE 11 presents the cooling rates from the final layer, Layer 249. In general, the cooling rates are similar to the Layer 25, though the material at the overhang experiences a slightly slower cooling rate. However, the reduction in cooling rate at the overhang is not as drastic as above the hole (in Layer 226, FIGURE 10), likely because the overhang is 45°, whereas above the hole a much more extreme overhang is created.



FIGURE 11 - Cooling rate measured during Layer 249. A) Measured cooling rate at each pixel in the layer, B) comparison with Layers 1 through 25.

SUMMARY AND FUTURE WORK

The in situ measurements in this study reveal that the thermal processing in LPBF is extremely imprecise. Both geometry and scan strategy are shown to increase and/or decrease the cooling rates by several factors, and the interaction between the two is complex. Although improvements can be made to improve the thermal processing precision, research must first be conducted to determine the level of precision required to achieve consistent microstructure, material properties, and stresses in the part.

REFERENCES

- [1] F.-W. Goudsmit, G. Oosterhuis, E. Sluiter, and T. Peijnenburg, (2014) Typical requirements for precision mechanical parts and opportunities for additive manufacturing, *Proceedings of the ASPE Topical Meeting: Dimensional Accuracy and Surface Finish in Additive Manufacturing*, Berkeley, CA, pp. 51–55.
- [2] P. Mercelis and J.-P. Kruth, (2006) Residual stresses in selective laser sintering and selective laser melting, *Rapid Prototyping Journal*, 12(5), pp. 54–265.
- [3] T. Keller, G. Lindwall, S. Ghosh, L. Ma, BM. Lane, F. Zhang, U.R. Kattner, E.A. Lass, J.C. Heigel, Y. Idell, M.E. Williams, A.J. Allen, J.E. Guyer, L.E. Levine, (2018) Application of finite element, phase-field, and CALPHADbased methods to additive manufacturing of Ni-based superalloys, *Acta Materialia*, 139, pp. 244-253.
- [4] Y.M. Arısoy, L.E. Criales, T. Özel, B. Lane, S. Moylan, & A. Donmez, (2017). Influence of scan strategy and process parameters on microstructure and its optimization in additively manufactured nickel alloy 625 via laser powder bed fusion. *The International Journal of Advanced Manufacturing Technology*, 90(5-8), pp. 1393-1417.
- [5] J.C. Heigel and B.M. Lane, (2018) Measurement of the melt pool length during single scan tracks in a commercial laser powder bed fusion process, *Journal of Manufacturing Science and Engineering*, 140(5).
- [6] Additive Manufacturing Benchmark Test Series (AM-Bench): Class 1 benchmark test description. https://www.nist.gov/ambench/ amb2018-01-description
- [7] J.C. Heigel, and B.M. Lane, (2017) The effect of powder on cooling rate and melt pool length measurements using in situ thermographic techniques, *Proceedings of the 28th Annual International Solid Freeform Fabrication Symposium – An Additive Manufacturing Conference*, pp. 1340-1348.
- [8] B. Lane and E.P. Whitenton, (2015) Calibration and Measurement Procedures for a High Magnification Thermal Camera, NIST Interagency/Internal Report (NISTIR) – 8098.

AMERICAN SOCIETY FOR PRECISION ENGINEERING

Heigel, Jarred; Whitenton, Eric.

COMPLEMENTARY USE OF OPTICAL METROLOGY AND X-RAY COMPUTED TOMOGRAPHY FOR SURFACE FINISH AND DEFECT DETECTION IN LASER POWDER BED FUSION ADDITIVE MANUFACTURING

Jason C. Fox¹, Felix Kim¹, Zachary Reese^{1,2}, Christopher Evans² ¹Engineering Laboratory National Institute of Standards and Technology^{*} Gaithersburg, Maryland, USA ²Center for Precision Metrology University of North Carolina at Charlotte Charlotte, North Carolina, USA

INTRODUCTION

The development of additive manufacturing (AM) has allowed for production of high-value and complex parts that reduce time-to-market and cost to manufacture [1]. A key benefit of the AM process is the capability to create lattice structures and highly complex, topology optimized parts that reduce weight. Despite this advantage, a limiting factor affecting widespread adoption of AM is the sometimes less desirable as-built surface topography of finished parts compared to those made using traditional manufacturing techniques. As part complexity increases, the ability to alter the surfaces through a secondary operation (i.e., machining or polishing), either in situ or ex situ, decreases [2]. Thus, improvements in the as-built surface texture and understanding its relationship to part quality has been cited as a key need [1,3].

In addition to a stronger understanding of the asbuilt surface texture of AM parts, development of non-destructive evaluation (NDE) techniques such as x-ray computed tomography (XCT) for surface texture measurement is needed. Traditional surface metrology equipment often requires line of sight or adequate clearance to contact the surface with a stylus, which can be difficult to achieve given the highly complex AM geometries [4]. Townsend et al. presented a methodology to extract areal surface texture from XCT data and compared it with a focus variation (FV) microscopy measurements [5]. Thompson et al. also showed a comparison and alignment of confocal microscopy, coherent scanning

interferometry, and FV to XCT data. XCT data in this work was first rotated to align the surface normal to the z-axis via principal component analysis and then translated via commercially available software. The work also suggested the integration of results to improve understanding of complex surface topography measurement [6,7].

The purpose of this research is to investigate the complementary use of optical metrology and XCT for inspection of parts built through laser powder bed fusion (L-PBF) AM, with the primary focus being the benefit that the volumetric XCT data can provide when aligned with the data from conventional surface metrology equipment. Samples made from 17-4 stainless steel (17-4SS) and nickel super alloy 625 (IN625) are measured using a laser confocal (LC) microscope and XCT system. Defects, which include reentrant features on the surface and near surface pores, are identified and located from the XCT data using methodologies described by Fox et al. [8]. The data from the XCT and laser confocal (LC) systems is then aligned using an iterative closest point (ICP) algorithm. This alignment allows for identification of locations in the LC data close to near-surface defects and provides more context for numerical analysis.

METHODOLOGY

Optical Measurements

Surface height data was acquired with a Zeiss LSM 800 LC system. All measurements were performed with a 10x objective and 0.5x tube lens or a 20x objective lens. These settings create

AMERICAN SOCIETY FOR PRECISION ENGINEERING

195

Official contribution of the National Institute of Standards and Technology (NIST); not subject to copyright in the USA. The full descriptions of the procedures used in this paper require the identification of certain commercial products. The inclusion of such information should in no way be construed as indicating that such products are endorsed by NIST or are recommended by NIST or that they are necessarily the best materials, instruments, software, or suppliers for the purposes described.

lateral resolutions of 1.25 µm and 0.31 µm, respectively. Measurements were processed with Digital Surf's ConfoMap software to remove outliers and level the surfaces. No other filtering was applied to the data.

XCT Measurements

XCT scans were acquired with North Star Imaging (NSI) CXMM 50 metrological CT system. The system is equipped with a 225 kV source including a tungsten target, a rotary stage, and a 127 μ m/pixel flat panel detector (FIGURE 1). The voxel size of the XCT data was 10.9 μ m. A typical filtered backprojection algorithm supplied by the vendor was used to reconstruct the dataset [9]. XCT acquisition parameters are shown in TABLE 1, and the methodology for selection of those parameters are detailed in Fox et al [8].



FIGURE 1. (a) XCT system setup, and (b) an example image of the sample.

Parameter	Value			
Voltage	160 kV			
Current	80 µA			
Exposure time	2 s			
Filter material/thickness	Cu/4 mm			
Number of projection	1000			
Source-to-detector distance (SDD)	492.76 mm			
Source-to-object distance (SOD)	42.29 mm			
Voxel size	10.9 μm x 10.9 μm x 10.9 μm			

TABLE 1. XCT acquisition parameters

Experiment Samples

Parts for the analysis were built on the EOS M270 system at the National Institute of Standards and Technology (NIST). Three parts were built for this analysis: two using the commercially available EOS StainlessSteel GP1 (corresponds to US classification 17-4SS [10]) and one EOS NickelAlloy IN625 (corresponds to classification unified numbering system UNS N06625 [11]). It should be noted that the material used for the build was powder reclaimed from prior builds

using an 80 µm sieve. It is assumed that the condition of the powder can have a large effect on the surface quality of parts being built and analysis of the powder is currently underway.

For the two parts built using the 17-4SS, the first was built entirely with the default parameters defined by the manufacturer, and will be referred to as "sample 60." The second was built with a laser power of 195 W and laser scan speed of 700 mm/s for the contour passes only (outermost portion of the part) with the rest of the parameters set to the default defined by the manufacturer. This will be referred to as "sample 195". This high power contour setting was chosen because prior work by the authors had shown that these settings create a greater variety of surface features [4,12]. The part built using the IN625 was built with the default parameters defined by the manufacturer and will be referred to as "sample 625." Samples built for the analysis are shown in FIGURE 2.



FIGURE 2. Sample 60 (left), sample 195 (center), and sample 625 (right). Scale is in cm.

The samples were built with the same orientation in the build chamber relative to the recoater blade. Surfaces of the samples will be referred to as follows: the top surface is parallel with the build plane and contains the part labels "60," "195," and "625," seen in FIGURE 2. Side surfaces will be referred to using primary intercardinal directions when looking down at the label on the top surfaces. For example, in FIGURE 2, the southeast (SE) and southwest (SW) surfaces of each part can be seen and the northeast (NE) and northwest (NW) surfaces cannot be seen.

RESULTS

Analysis of Optical Metrology Data

Height measurements were taken from the surfaces of the samples using a laser confocal microscope. The sample is aligned in the instrument such that the positive y-axis of the microscope is parallel to the build direction, the intersection of the top surface and left edge of the

2018 ASPE and euspen Summer Topical Meeting Volume 69

sample is at (0,0). Each surface was scanned over a 5 x 5 stitch using a 10x objective and 0.5x tube lens to create a 5.89 mm x 5.89 mm view of the surface with 1.25 µm point spacing. Additionally, each surface was scanned over a 3 x 3 stitch using a 20x objective and 1.0x tube lens to create an 895 µm x 895 µm view of the surface with 0.31 µm point spacing. The scan center was positioned at (2.5,-3) mm.

Analysis of XCT Data

The reconstructed XCT data were imported into VGStudioMAX 3.0 software and the 3-2-1 registration technique available in the software was used to align the orthogonal faces to the same coordinate system and minimize part-topart variation during the analysis [13]. Then, a stack of images was exported such that the cross-sectional image shows both SW and NE edges of the sample. In addition to the 10.9 µm per voxel resolution, a sub-voxel interpolation to 2.5 µm per voxel was performed through commercially available software and was used to create images for analysis at a higher resolution [14]. From these images, a 2D array of surface heights can be created using the methods described in Fox et al. [8].

Data Alignment

To align the data from various sources, height maps were converted to point clouds and registered using an ICP algorithm, available in the commercially available MATLAB software package. Alignment was performed in multiple steps. For alignment of the 20x LC data and 10x LC data, an initial crop of the 10x LC data is performed to reduce the computational burden of the alignment and limit the search area to the general region where the 20x LC measurement was performed (2.5 mm to the right and 3 mm down from the upper left corner of the surface). From this point, all alignments were completed in two steps. In the first step, the lowest 70 % of points were ignored and the two surfaces are roughly aligned using the ICP algorithm on the highest 30 % of points on the surface. In the second step, the translation and rotation determined in the first step is used as the starting point for the ICP algorithm on the full set of surface points. The alignment was performed using these two steps because even a minor error in selection of a fiducial (on the order of tens of micrometers) can lead to errors in alignment on the order of millimeters in the x and y directions with the ICP algorithm. Aligned sections from the various data sources can be seen in FIGURE 3.



FIGURE 3. Aligned data maps from various sources.

AMERICAN SOCIETY FOR PRECISION ENGINEERING

197

Qualitatively, this method provides good alignment between the various sources. To analyze the quality of the alignment signed distance maps quantitatively, and distributions can be seen in FIGURE 4 through FIGURE 6. Additionally, the signed distances are used to determine the root mean square error (RMSE) between two aligned surfaces. It should be noted that while sub-voxel interpolation of the XCT data to 2.5 µm per voxel is available in the commercial software, it did not exhibit a significant difference in the identification of the surface from the original 10.9 µm per voxel data. This can be seen when comparing FIGURE 5 and FIGURE 6.

Identification of Defects

Images of the surface were taken via scanning electron microscope (SEM) for qualitative comparison. FIGURE 7 shows an example image from the SW surface of sample 60. The image shows a 1 mm x 1 mm area of the surface that encompasses the area previously scanned with the laser confocal microscope using the 20x objective.



FIGURE 4. Signed distances distribution and map for alignment of the 10x LC and 20x LC data.

198



FIGURE 5. Signed distances distribution and map for alignment of the 20x LC and XCT data.



FIGURE 6. Signed distances distribution and map for alignment of the 20x LC and Sub-Voxel XCT data.

2018 ASPE and euspen Summer Topical Meeting Volume 69



FIGURE 7. SEM image of the SW surface of sample 60.

While the current resolution of the XCT data is limited, it has the distinct advantage of being able to detect pores and undercuts on and near the surface. The method described in Fox et al. [8] was used to determine undercuts on the surface and sub-surface pores from the surface to approximately 275 µm below the surface. The resultant pore and undercut locations are presented in FIGURE 8.



2 mm

FIGURE 8. Location of near surface pores and undercuts on the SW surface of sample 60 identified from XCT data.

From the identification of pores and undercuts in FIGURE 8, surface data near these defects can be extracted from the previously aligned data sources. Defect locations are overlaid onto the 20x LC data in FIGURE 9. A closer view of the surface data near the rightmost defect location in FIGURE 9 (highlighted with a red box) is shown in FIGURE 10. While this methodology allows us to align and identify regions where subsurface defects exist, it is still inconclusive whether trends in the surface data from the LC system can identify these defects. As such, further analysis of the surface data is ongoing.



FIGURE 9. Surface data for the SW surface of sample 60, taken via 20x objective on the laser confocal system, with defect areas (pore and undercut locations) indicated by plus signs.



FIGURE 10. Surface data taken via 20x objective on the laser confocal microscope for the rightmost defect area (highlighted with a red box) in FIGURE 9.

CONCLUSIONS

In this work, IN625 and 17-4SS samples were built on a commercially available L-PBF system and analyzed via laser confocal microscope and XCT. The data were aligned using an ICP algorithm for complementary analysis. Pores and undercuts were identified using XCT data, which allowed for more stringent analysis of the surface near these features. However, at 10.9 µm per voxel the current resolution of XCT prevents strong quantitative analysis of the surface data from XCT or relevant comparison to the laser confocal microscope surface data. While quantitative analysis has been performed on XCT data with larger voxel sizes, it has utilized sub-voxel

AMERICAN SOCIETY FOR PRECISION ENGINEERING

interpolation. The current analysis has shown that sub-voxel interpolation provided little change in the determined surface, and greater benefit may be derived from proprietary surface determination algorithms in the commercially available software. This was not included in the current analysis.

Future work will include the classification of defects found in the data (i.e., subsurface pores vs undercuts on the surface), and identification of the depth of pores. Additionally, advanced pore detection techniques and image analysis, such as those presented in Kim et al. [15], and sectioning of the samples for comparison to the XCT and laser confocal microscope data are in progress. It is anticipated that this type of analysis will create avenues to get richer information from the surface data and add a potential methodology for feature based metrology.

ACKNOWLEDGEMENTS

This work was performed in part in the NIST Center for Nanoscale Science and Technology (CNST) NanoFab Facility. The authors would like to thank Maxwell Praniewicz for his suggestions to improve the alignment methodology, and North Star Imaging for providing the XCT system through a Cooperative Research and Development Agreement with NIST and allowing the use of the data acquired.

REFERENCES

- [1] Measurement Science Roadmap for Metal-Additive Manufacturing. Based Gaithersburg, MD: NIST, US Department of Commerce; 2012.
- [2] Pyka G, Burakowski A, Kerckhofs G, Moesen M, Van Bael S, Schrooten J, et al. Surface modification of Ti6Al4V open porous structures produced additive by Engineering manufacturing. Advanced 2012;14:363-70. Materials doi:10.1002/adem.201100344.
- [3] America Makes & ANSI Additive Manufacturing Standardization Collaborative Standardization Roadmap for Additive Manufacturing, Public Draft v1.0. 2017.
- [4] Fox JC, Moylan S, Lane BM. Preliminary Study Toward Surface Texture as a Process Signature in Laser Powder Bed Fusion Additive Manufacturing. Proceedings of the 2016 ASPE Summer Topical Meeting: Dimensional Accuracy and Surface Finish in Additive Manufacturing, Raleigh, NC: 2016.
- [5] Townsend A, Pagani L, Scott P, Blunt L. Areal surface texture data extraction from X-

ray computed tomography reconstructions of manufactured metal additively parts. Precision Engineering 2017;48:254-64. doi:10.1016/j.precisioneng.2016.12.008.

- [6] Thompson A, Körner L, Senin N, Lawes S, Maskery I, Leach R. Measurement of internal surfaces of additively manufactured parts by computed X-ray tomography. 7th Conference Industrial on Computed Tomography, Leuven, Belgium: 2017.
- [7] Thompson A, Senin N, Giusca C, Leach R. Topography of selectively laser melted surfaces: A comparison of different measurement methods. CIRP Annals -Manufacturing Technology 2017. doi:10.1016/j.cirp.2017.04.075.
- [8] Fox JC, Kim FH, Reese ZC, Evans C. Investigation of complementary use of optical metrology and x-ray computed tomography for surface finish in laser powder bed fusion additive manufacturing, Leuven, Belgium: 2017, p. 132-6.
- L.A. Feldkamp, L.C. Davis, J.W. Kress, [9] Practical cone-beam algorithm, J. Opt. Soc. Am. A 1(6) 612-619. 1984.
- [10] Material data sheet EOS StainlessSteel GP1 for EOSINT M270 2009. https://scrivitopublic-cdn.s3-eu-west-1.amazonaws.com/eos/public/5f84f5d2c88a c900/05fb1582834a38c85ef6dd859733a230 /EOS_StainlessSteel-GP1_en.pdf (accessed December 7, 2015).
- [11] Material Data Sheet EOS NickelAlloy IN625 2011. http://ip-saas-eoscms.s3.amazonaws.com/public/d1327facdc a0e32a/373a60ec4f5c891b7dbcdf572e37d3 b0/EOS_NickelAlloy_IN625_en.pdf (accessed June 13, 2017).
- [12] Fox JC, Moylan SP, Lane BM. Effect of Process Parameters on the Surface Roughness of Overhanging Structures in Laser Powder Bed Fusion Additive Procedia Manufacturing. CIRP 2016:45:131-4.

doi:10.1016/j.procir.2016.02.347.

- [13] Volume Graphics GmbH, VGStudioMax 3.0.
- [14] Avizo 9.4, Thermo Fisher Scientific, Waltham, MA.
- [15] Kim FH, Moylan SP, Garboczi EJ, Slotwinski JA. Investigation of pore structure in cobalt chrome additively manufactured parts using X-ray computed tomography and threedimensional image analysis. Additive 2017;17:23-38. Manufacturing doi:10.1016/j.addma.2017.06.011.

2018 ASPE and euspen Summer Topical Meeting Volume 69

EVOLUTION OF COOLING LENGTH IN PARTS CREATED THROUGH LASER POWDER BED FUSION ADDITIVE MANUFACTURING

Zachary C. Reese¹, Jason Fox², John Taylor¹, and Chris Evans¹

¹ Center for Precision Metrology University of North Carolina at Charlotte Charlotte, North Carolina, USA ² The National Institute of Standards and Technology^{*} Gaithersburg, Maryland, USA

ABSTRACT

Additive manufactured (AM) components, specifically those created through laser powder bed fusion (LPBF) methods, exhibit an abundance of surface textures of varving forms and patterns. These topographies have historically been categorized solely using their Ra values, a metric which offers limited information to discern differences among the morphologies of AM surfaces. This diversity is illustrated not only among parts made using different machines or processing parameters, but also between different locations in the build chamber and even on the same part. Current instruments have the potential to acquire three dimensional (3D) maps of the AM surface, enabling a range of field and feature based descriptors that offer potential in characterizing AM surface morphology. This work explores feature-based metrology of AM surfaces in an effort to relate features to part quality, as well as aid the modeling community in better understanding of melt pool geometry. The feature of interest in this work, the chevron pattern seen on top of scan lines, is theorized to be connected to the laser parameters used in the printing of the part. This research explores the chevron pattern, observing changes in this feature over a range of laser power and velocity combinations.

INTRODUCTION

Metal AM is a growing technology which fabricates parts directly from 3D computer aided design (CAD) models. Parts are built up layer-bylayer starting at the substrate. A single layer of

metal powder is deposited over the substrate and the first laver of the part is selectively melted. fusing with the substrate. The build plate is lowered, and the process repeats until the part is completed. The layer-by-layer building of these parts allows for them to be created with complex geometries not attainable through traditional manufacturing methods.

Though AM processes can create geometries that other methods cannot, they are not without their own disadvantages. Parts often take on the order of hours to days being built and suffer from poor "as-printed" surface quality. While finishing processes such as laser ablation or bead blasting can improve surface roughness [1-4], the increasing complexity of these parts can render these processes ineffective. These rough surfaces are a major hurdle in the full-scale implementation of AM [5], as the understanding of their morphology is still in its adolescence. Often, these complex surfaces were being categorized by their Ra values alone; however, it has been shown that these classical parameters prove to be similar for visibly different surfaces [6-9]. In previous works, the authors have shown that based on the areal average roughness (Sa), as well as other parameters outlined in International Organization for Standards (ISO) 27178-2 [14], that it is impossible to discern between two qualitatively different surfaces (Figure 1) [6]. Additionally, two works by Fox et al. launched an investigation into downward facing, as-printed surfaces. These works

AMERICAN SOCIETY FOR PRECISION ENGINEERING

Official contribution of the National Institute of Standards and Technology (NIST); not subject to copyright in the USA. The full descriptions of the procedures used in this paper require the identification of certain commercial products. The inclusion of such information should in no way be construed as indicating that such products are endorsed by NIST or are recommended by NIST or that they are necessarily the best materials, instruments, software, or suppliers for the purposes described.



FIGURE 1. Comparison of two different metal AM surfaces based on ISO parameters

conclude that the use of Ra alone is insufficient in discerning the difference between surfaces predominantly covered by powdered metal particles and those dominated by scan tracks [7,8]. This has driven a push toward featurebased investigation for surface characterization. Rather, many researchers are beginning to rely on surface features to characterize metal AM parts [6-9].

With the previous standard method returning very little in the way of new surface information and a push toward feature-based investigation for surface characterization, the authors have identified the prominent chevron pattern that is often seen on upward facing surfaces of AM parts as a feature of interest (Figure 2). It is theorized that the chevron pattern may be linked to the dimensions of the melt pool.



Figure 2. Chevron pattern on upward facing surface of AM part

Preliminary work in numerical simulation has been performed in an effort to suitably model the melt pool dimensions for given build parameters. Gockel and Beuth have shown that, through finite element simulation, one can predict microstructure and grain morphology in Ti-6Al-4V. When operating in certain regions of the laser power-velocity (PV) process map, one can control the melt pool geometry and affect the cooling rate of single scan passes [10]. Other modeling attempts have explored this idea, for various nickel alloys rather than titanium. Keller et al. simulated the laser melt pool in nickel alloy 625 using finite element analysis. Simulations matched the surface temperatures captured insitu with thermographic measurements in order to approximate thermal conditions beneath the surface of the scan line. Understanding these thermal conditions will allow one to predict and eventually control microstructure during builds [11]. This work was further expanded on by Ghosh et al. extending the same numerical simulations to nickel alloy 718 [12].

EXPERIMENTAL SETUP

The first step in this feature-based metrological approach was to look at the scan lines that appear on the top of the build without outside influence from the rest of the surface. To make this case as simple as possible, a number of single scans without powder were printed to build knowledge about the scan lines themselves. Single scans were printed into a nickel alloy 625 substrate in an EOS M270 Direct Metal Laser Sintering (DMLS) machine. These scan lines have a chevron pattern that occurs on the top of each line and is qualitatively different between scans of different PV combinations (Figure 3). Given this variation in the pattern, the angle

184

between the legs of the chevrons was chosen as a metric to quantify these variations.

100 W 400 mm/s



195 W 800 mm/s

FIGURE 3. Variation in chevron pattern between different laser power and velocity combinations

If one knows the angle of this pattern and the width of the scan track, it is trivial to back out the length of the melt pool from its widest (and deepest) point to the tail. The length of the solidified melt pool from the tip of the chevron to the point where the legs reach the width of the scan line will be referred to as the "cooling length" (Figure 4). The importance of cooling length is described in another work by Gockel et al. who show that the cooling length is correlated with the microstructure of that track after cooling [12].

This information can potentially supplement the in-situ thermal data for validating models of melt pool dimension.



FIGURE 4. Distinction between melt pool length and cooling length

All builds for this paper were conducted in an EOS M270 DMLS machine. Six sets of 24 single pass scan tracks were printed over the course of two days to investigate the stability of this feature run to run as well as day to day. Each scan line is 10 mm long and offset by 2.5 mm from the

previous line. This ensures that the melt pool has sufficient time to reach steady state conditions in each line and is not influenced by the residual heat of the previous tracks. Power and velocity (PV) combinations are shown in Figure 5. Single scan lines ensure the chevron structure occurs in the middle of the track with little variation in the angle between the legs of each chevron. Printing with only beam on plate (i.e., no metal powder) mitigates any chance of spatter particles ejecting from the melt pool, denuding effects which can cause irregularity in the scan lines, partially melted powder particles adhering to the surface, and the variability of powder layer thickness.



FIGURE 5. PV combinations for scan lines

Once printed, height maps of the single scans were obtained using a coherence scanning interferometer (CSI) and processed using a commercial software package. Each scan line was stitched with a 20x objective (418 µm² field of view (FOV)) with an overlap of 20 % between sites. After the data is leveled and the height data from the base plate removed, it can be seen in Figure 6A that the chevron pattern is barely visible in the unfiltered measurement. Even the larger spatial frequency chevrons are roughly an order of magnitude smaller than the scan line itself. Removing a best fit cylinder flattens the scan line, improving the visualization of the chevron pattern. The pattern can be further highlighted by applying appropriate Fourier filters. Figure 6B shows the same scan line with a Fourier band-pass filter at a frequency of 20 per mm to 100 per mm. At this point, the large-scale chevron pattern is fully segmented from the underlying surface. However, adjusting the filter cutoffs to 100 per mm to 300 per mm can further bring out the finer scale features of this pattern as shown in Figure 6C.

AMERICAN SOCIETY FOR PRECISION ENGINEERING



FIGURE 6. A) Unfiltered scanline. B) Scan line with cylinder removed and a Fourier bandpass filter from 20 per mm to 100 per mm. C) Filter cutoffs adjusted to 100 per mm to 300 per mm.

With chevrons isolated from the surface, areal autocorrelation was used to evaluate the angle of this heavily repeated pattern. The chevrons are measured where the scan track had reached steady state and are highly symmetric (Figure 6B). This allows for the left leaning leas of the chevron pattern (Figure 6C) to be analyzed. simplifying the analysis. ISO 25178 [14] defines three autocorrelation parameters: the fastest decay autocorrelation length (Sal), the texture aspect ratio (Str), and the texture direction (Std). These parameters provide information not only on how directional the surface being analyzed is, but also the direction of the fastest and slowest decay of the autocorrelation length. The angle at which the major radius of the center lobe occurs gives the parameter Std which coincides with the angle of the chevron pattern shown in Figure 7. Cooling length is calculated by taking half the width of the scan line divided by the tangent of the found chevron angle from the autocorrelation function.



Chevron height map FIGURE 7. (Left). Autocorrelation of surface and Std designated as θ (Right)

RESULTS AND DISCUSSION

From this measurement method, it is clear that the chevron pattern itself shows some variation in angle from the center of the scan line to the edges. It is likely that this is due to the legs of this pattern being slightly rounded. While this leads to some variation in angle in each scan line, this method has still proven viable in telling the lines differences between scan and approximating the cooling length of the melt pool ex situ. Figures 8 and 9 show the results from six runs of 24 scan lines and their comparison with modeling predictions for the scan track width and cooling length. Track width was measured ex-situ through microscope images taken over 1 mm of track. Images are taken at, nominally, the center of the track to avoid any edge effects. This measurement method is outlined by work conducted by Fox et. al. [15].

The model chosen for comparison is the Rosenthal solution [16]. The Rosenthal solution provides the analytical explanation to a moving heat source, originally developed to understand welding. Dykhuizen and Dobranich later adapted this model, applying it to laser based additive manufacturing processes [17, 18]. From Figure 8, one can see the experimental results of width measurement with error bars set to 1 standard deviation. Experimental width measurements match not only the predicted pattern, but also the approximate values modeled by the Rosenthal solution.

Cooling length measurements, shown in Figure 9 with error bars set to the largest deviation from the mean, appear to diverge from the model's predictions. This, however, was expected as the Rosenthal solution is a fairly simplified model and does not take into account some of the physical processes occurring during the build. Among those not considered is the creation of voids in the wake of the melt pool due to processing

conditions with high energy density, referred to as keyholing. Excluding the region where keyholing is expected (all power cases for the 100 mm/s and 200 mm/s cases), a linear pattern is apparent. Thought not predicted by the model, some of these linear trends have a slight negative slope. The decrease in cooling length as velocity increases for a given power is feasible as the energy imparted into the surface is decreased with faster scan speeds when all else is held constant. Results of this very slight cooling length decrease as velocity increases coincide closely with results shown by Heigel and Lane's in-situ thermographic measurements of melt pool lengths, however in both cases more results are needed [19]. Model results for cooling length could also be improved by moving toward higher fidelity finite element simulations as well as incorporating more physics into the process [20].





FIGURE 8. Experimental melt pool width comparison with the Rosenthal solution data (error bars set to one standard deviation).







FIGURE 9. Experimental cooling length comparison with the Rosenthal solution data (error bars set to largest deviation seen).

FUTURE WORK

This work will be expanded in two major directions. The scan lines measured in this work will be sectioned and undergo microstructural analysis to relate microstructure to the length-todepth ratio of the melt pool. This method will also be applied to single scan lines with powder as well as multiple adjacent scan tracks with and without powder to understand how this analysis method will progress when applied to building full sized parts.

REFERENCES

[1] Mumtaz KA, Hopkinson N. Selective Laser Melting of thin wall parts using pulse shaping. Journal of Materials Processing Technology 2010:210:279-287.

doi:10.1016/j.jmatprotec.2009.09.011

- [2] Liu X, Chu PK, Ding C. Surface modification of titanium, titanium alloys, and related materials for biomedical applications. Materials Science and Engineering: R: 2004;47:49-121. Reports doi:10.1016/j.mser.2004.11.001.
- [3] Lane BM, Moylan SP, Whitenton EP. PostProcess Machining of Additive

AMERICAN SOCIETY FOR PRECISION ENGINEERING

Manufactured Stainless Steel. Proceedings of the 2015 ASPE Spring Topical Meeting: Achieving Precision Tolerances in Additive

- [4] Yasa E, Kruth J-P, Deckers J. Manufacturing by combining Selective Laser Melting and Selective Laser Erosion/laser re-melting. CIRP Annals Manufacturing Technology2011;60:263-6. doi:10.1016/j.cirp.2011.03.063.
- [5] Measurement Science Roadmap for Metal Based Additive Manufacturing, Gaithersburg, MD: NIST; 2012.
- [6] Z. Reese et.al. Observations on the surface morphology of laser powder bed fusion metal surfaces. Met and Props 2017, June 2017, Gothenburg, Sweden. Unpublished conference paper, 2017.
- [7] Fox, Jason C., Shawn P. Moylan, and Brandon M. Lane. "PRELIMINARY STUDY TOWARD SURFACE TEXTURE AS A PROCESS SIGNATURE LASER IN FUSION ADDITIVE POWDER BED MANUFACTURING." 2016 Summer Topical Meeting: Dimensional Accuracy and Surface Finish in Additive Manufacturing. 2016.
- [8] Fox JC, Moylan SP, Lane BM. Effect of process parameters on the surface roughness of overhanging structures in laser powder bed fusion additive manufacturing. Procedia CIRP, Charlotte, NC: 2016. doi:10.1016/j.procir.2016.02.347
- [9] Senin, Nicola, and Richard K. Leach. "Information-rich surface metrology." Procedia CIRP (2018).
- [10] Gockel, Joy, and Jack Beuth. "Understanding Ti-6AI-4V microstructure control in additive manufacturing via process maps." Solid Freeform Fabrication Proceedings, Austin, TX, Aug (2013): 12-14.
- [11]Keller, Trevor, et al. "Application of finite element, phase-field, and calphad-based methods to additive manufacturing of Nibased superalloys." Acta Materialia (2017).
- [12] Ghosh, Supriyo, et al. "On the primary spacing and microsegregation of cellular dendrites in laser deposited Ni-Nb alloys." Modell. Simul. Mater. Sci Eng (2017).
- [13] Gockel, Joy, Jack Beuth, and Karen Taminger. "Integrated control of solidification microstructure and melt pool dimensions in electron additive beam wire feed manufacturing of Ti-6Al 4V." Additive Manufacturing 1 (2014): 119-126.

- [14] ISO 25178-2 (2012) Surface texture: Areal - Terms, definitions, and surface texture parameters
- [15] Fox JC, Lane BM, Yeung H, "Measurement of process dynamics through coaxially aligned high speed near-infrared imaging in laser powder bed fusion additive manufacturing", SPIE Proc. 10214 Thermosense: Thermal Infrared Applications XXXIX, 1021407 (5 May 2017); doi: 10.1117/12.2263863
- [16] D. Rosenthal: Trans. ASME, 1946, vol. 68, pp. 849–66.
- [17] R. Dykhuizen and D. Dobranich: "Analytical Thermal Models for the LENS Process," Sandia National Laboratories Internal Report, 1998.
- [18] R. Dykhuizen and D. Dobranich: "Cooling Rates in the LENS Process," Sandia National Laboratories Internal Report, 1998.
- [19] Heigel, Jarred C., et al. "Measurement of the melt pool length during single scan tracks in a commercial laser powder bed fusion process." Proc. MSEC, Los Angeles, CA (2017): 575-591.
- [20] Francois, Marianne M., et al. "Modeling of additive manufacturing processes for metals: opportunities." Current Challenges and Opinion in Solid State and Materials Science 21.LA-UR-16-24513 (2017).

188

EFFECT OF SUBSURFACE DEFECTS ON THE SURFACE TOPOGRAPHY OF ADDITIVELY MANUFACTURED COMPONENTS

Zachary C. Reese¹, Jason Fox², Felix Kim², John Taylor¹, and Chris Evans¹

¹Center for Precision Metrology University of North Carolina at Charlotte Charlotte, North Carolina, USA ²The National Institute of Standards and Technology^{*} Gaithersburg, Maryland, USA

ABSTRACT

Additively manufactured (AM) components exhibit an abundance of surface textures and patterns. Past work investigating components created through laser powder bed fusion (LPBF) has shown that these patterns, specifically the chevron resulting from solidification of the melt pool on upward facing surfaces, can be correlated to quality of the final part [1]. Additionally, these patterns as well as the scan tracks have been observed to vary around the region of potential subsurface defects based on flash thermography data. This work explores parts with seeded subsurface defects at various depths from the top surface that are printed in nickel alloy 625 on a commercially available LPBF machine. Surface height maps are obtained using a scanning white light interferometer (SWLI) to determine the defect's impact on the surface texture of subsequent layers. Data obtained from the SWLI are aligned with data from X-ray computed tomography (XCT) to confirm the locations of defects and applicability of this approach as an in-situ detection method is discussed.

INTRODUCTION

Laser powder bed fusion (LPBF) is an additive manufacturing (AM) process that uses a highpower laser to selectively melt layers of metal powder. These layers are built up from an underlying substrate to form a three-dimensional structure. While the manufacturing method allows for almost unlimited design capacity, the process

itself is not without issues [2]. LPBF parts take on the order of hours or days to build, suffer from poor "as printed" surface texture, and potentially harbor unseen internal defects. This, plus the potential for part-to-part variation, inhibits the gualification and certification of AM parts for aerospace applications [2]. The internal defects, such as porosity and lack of fusion, are only found through destructive or nondestructive secondary analysis of the parts which can be costly and time consuming. Additionally, options for defect analysis become slim if one wants to keep the part intact. Flash thermography has shown some promise; however, it is limited on the depth of the defect that can be observed. X-ray computed tomography (XCT) is a viable option, but requires expertise and expensive equipment for analysis. Currently these analysis methods, both destructive and nondestructive, can only be conducted ex-situ when the part is complete, and it is far too late to remedy the defect. In-situ observation and recognition of deviations during the printing process are required if defects are to be fixed before the part is complete. Currently there is no reliable method of defect recognition process. the printing during However, understanding "as-printed" surface features may be the first step toward an in-situ method for defect detection.

Historically, these complex surfaces were being categorized by their average roughness (Ra) values alone: however, it has been shown that

AMERICAN SOCIETY FOR PRECISION ENGINEERING

Official contribution of the National Institute of Standards and Technology (NIST); not subject to copyright in the USA. The full descriptions of the procedures used in this paper require the identification of certain commercial products. The inclusion of such information should in no way be construed as indicating that such products are endorsed by NIST or are recommended by NIST or that they are necessarily the best materials, instruments, software, or suppliers for the purposes described.

these classical parameters prove to be similar for visibly different surfaces [4-6]. This has driven a push away from using classical parameters for surface characterization. Rather. manv researchers are beginning to rely on surface features to characterize metal AM parts [3-6]. In previous works, the authors have shown the connection of the chevron pattern on the top of scan lines, seen in Figure 1, to the laser parameters set for the build. A parallel study utilizing flash thermography on parts created through LBPF showed potential subsurface defects. Scanning electron microscope (SEM) images were taken at the site of the potential defect to observe any difference in surface texture. Figure 2 shows that the scan lines lose all definition in the area above the potential defect. The chevron pattern also seems to almost disappear. Upon investigation of the other potential defect sites, a pattern of flattened scan lines with no distinct boundaries emerged. In order to study this phenomenon under controlled conditions, parts with seeded defects were conceived and printed.

100 W 400 mm/s



195 W 800 mm/s

FIGURE 1. Variation in chevron pattern between different laser power and velocity combinations.



FIGURE 2. SEM image over largest potential defect (bottom). Return of scan lines away from defect (top).

Methodology

The largest hurdle in designing these parts comes in deterministically planning and creating stochastic defects. Defects are not only created in the surface through user error such as poor parameter selection, but also can occur randomly due to the chaotic nature of the printing process. The goal of the defect artifact is to encapsulate as many of these conditions in which defects can occur as possible. Three defect conditions were settled on: open hole, keyholing, and lack of fusion. Open hole defects are cylindrical areas of powder left purposely un-melted (Figure 3A). with Keyhole defects were printed the recommended laser power and half the recommended laser velocity to ensure the deposition was melting in a keyhole mode (Figure 3B). In the lack of fusion defects, the nominal hatch spacing was doubled to give subsequent printed layers issues in adhering to the defect area (Figure 3C). Defects are printed in a known grid pattern into the square artifact base. All are printed as cylinders ranging from 0.5 mm to 2.5 mm in diameter (Figure 4). The first two columns in Figure 4 (left to right) are left open to air (open air) in order to realize a local coordinate system. The rest are printed over with up to eight subsequent layers to better understand how these defects may propagate throughout a build.

The artifact is built like a staircase in that each column after the open-air defects has one more layer than the previous, and the final column is covered with eight layers (320 μ m) (Figure 5). Four artifacts were printed at defect depths from two to ten layers (80 μ m to 400 μ m) deep to understand the size of seeded defect that could reliably be printed. All artifacts were made of nickel alloy 625 in an EOS M290 Direct Metal Laser Sintering (DMLS) machine.



FIGURE 3. Microscope images of open-air defects A) Open hole B) Keyhole C) Lack of fusion.

	0 #800	o	o	o	o	o	0	o	0
	0 ##00	o	o	o	0	o	0	0	o
	0 #200	0	0	0	0	0	0	0	0
	O ##0O	0	0	0	0	0	0	0	0
	0 #100	0	0	0	0	0	0	0	0
	0 ##0	o	o	o	o	0	o	o	0
	0 #80	0	0	ο	0	0	0	0	0
Г	→ *** →	0	ο	0	0	0	0	0	0
		0	0	0	0	0	0	0	0
0		\bigcirc	0	\bigcirc	\bigcirc	0	\bigcirc	\bigcirc	\bigcirc

FIGURE 4. Grid pattern and size of defects in the artifact



FIGURE 5. Final computer aided design model for defect artifact.

Results and Discussion

Once the artifacts were built, they were subject to X-ray computed tomography (XCT) analysis to ensure that the defects that have been printed over still exist within the part. Parts were sent to North Carolina Agricultural and Technical State University (NC A&T) for preliminary XCT analysis. It can be seen in Figure 6 that in the artifact with the deepest seeded defects (ten layers or 400 μ m), all defect types are present within the surface and the largest diameter defects of each type persist throughout the entire build with up to eight layers on top.



Figure 6. Defects propagating through build (0-8 layers from left to right). A) Open hole B) Keyhole C) Lack of Fusion

Areal surface data was obtained through a commercially available coherence scanning interferometer (CSI) and processed using the packaged software. To capture these defects, up to 100 sites of areal data with a field of view (FOV) of 418 µm x 418 µm were stitched together to encompass the area of the defect and outside

AMERICAN SOCIETY FOR PRECISION ENGINEERING



Layers 4-8



FIGURE 7. Surface topography evolution of open hole defect from 0 to 8 layers (2.25 mm x 2.25 mm FOV)

surrounding topography to view any lateral propagation that may take place as the layers are built on top of the defects. Figure 7 shows the evolution of the surface topography as the largest open hole defect goes from open air up to eight layers deep. One can see that a defect of this size takes a full three layers to fill over the defect itself. Additionally, there is a prominent sag in the spot of the defect throughout the build. At the center of this sag, it is apparent that the scan lines are losing their independence as seen in the SEM data from Figure 2. This correlates with the preliminary surface topography results found in the stochastic defect in the real-world part. The flattened scan lines are just half of the correlation between the defect artifacts and the first part analyzed. Applying a Fourier filter to the area map of the eighth layer reveals that the chevron pattern ceases to exist (Figure 8). It is theorized that the disappearance of the pattern has to do with the altered heat transfer properties of building over and around the seeded defect; however, in-depth microstructural analysis is needed to confirm this.



FIGURE 8. Chevron pattern adjacent to seeded defect (left). Lack of chevron pattern on eighth layer above defect

FUTURE WORK

While preliminary analysis is promising, a lot of analysis is yet to be done to prove the correlation between subsurface defects and change in the scan lines and chevron pattern seen on the upward facing surfaces. Higher fidelity XCT of defects removed from the artifact via wire electrical discharge machining (EDM) will be used to estimate the size of the initial voids as well as investigate any micro-scale porosity that propagate throughout the mav part. Microstructural analysis will be completed to understand how these defects affect the expected microstructure. Finally, another series

of builds is being completed to see if the pattern holds for deeper defects (up to 30 layers).

ACKNOWLEDGEMENTS

The authors would like to thank the NC A&T team for the XCT analysis of these parts.

REFERENCES

- [1] Gockel, Joy, and Jack Beuth. "Understanding Ti-6AI-4V microstructure control in additive manufacturing via process maps." Solid Freeform FabricationProceedings, Austin, TX, Aug (2013): 12-14.
- [2] Energetics Inc. for National Institute of Standards and Technology. "Measurement science roadmap for metal-based additive manufacturing" (2013). http://www.nist.gov/el/isd/upload/NISTAdd Mfg_Report_FINAL-2.pdf (accessed December 15, 2014).
- [3] Z. Reese et.al. Observations on the surface morphology of laser powder bed fusion metal surfaces. Met and Props 2017, June 2017, Sweden. Unpublished Gothenburg, conference paper, 2017.
- [4] Fox, Jason C., Shawn P. Moylan, and Brandon M. Lane. "PRELIMINARY STUDY TOWARD SURFACE TEXTURE AS A SIGNATURE IN PROCESS LASER POWDER BED FUSION ADDITIVE MANUFACTURING." 2016 Summer Topical Meeting: Dimensional Accuracy and Surface Finish in Additive Manufacturing. 2016.
- [5] Fox JC, Moylan SP, Lane BM. Effect of process parameters on the surface roughness of overhanging structures in laser powder bed fusion additive manufacturing. Procedia CIRP, Charlotte, NC: 2016. doi:10.1016/j.procir.2016.02.347
- [6] Senin, Nicola, and Richard K. Leach. "Information-rich surface metrology." Procedia CIRP (2018).

AMERICAN SOCIETY FOR PRECISION ENGINEERING

Examining Workcell Kinematic Chains to Identify Sources of Positioning Degradation

Alexander Klinger¹, Brian A. Weiss²

^{1,2}National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, 20899, USA alexander.klinger@nist.gov brian.weiss@nist.gov

ABSTRACT

Automated industrial workcells are becoming increasingly complex and varied due to greater accessibility of advanced robotic and sensing technologies. Degradation monitoring and diagnostics must advance to reduce the impact of increased system complexity on troubleshooting faults and failures and to optimize system operations. A new methodology is being developed for the design and implementation of monitoring kinematic chains commonly found in robot workcells. This method will enable the identification of degraded components which contribute to relative positioning accuracy error between moving objects, tools, devices, and other components. The proposed methodology is being developed and tested on a six degree of freedom industrial robot arm workcell use case developed at the National Institute of Standards and Technology (NIST). Industrial robot users and integrators can use this method to examine the kinematic chains within their workcells and design a key position monitoring implementation. With the added key position monitoring, degradations can be identified at a designed resolution allowing for enhanced maintenance planning and production control. The methodology will be extended to other manufacturing workcells in the future.

1. INTRODUCTION

Robotics, along with supporting sensing and automation technologies, are actively used for a variety of tasks throughout the manufacturing industry (Ahmad & Plapper, 2016; Chen, Fuhlbrigge, & Li, 2008; Shi, Wang, & Li, 2007). As robotic and sensor technologies have evolved, the diversity and complexity of use cases have grown. There is a need to better understand the health, and consequently, the wear of these automation systems, as the manufacturing community increases their reliance on advanced technologies. Systems can be maintained more efficiently,

leading to a more robust process, if/when advanced health information is generated.

A critical element of industrial robot workcells is reliable (inclusive of repeatability) positioning. In many robot workcells, a six degree of freedom (6DOF) industrial robot arm is used as a macro positioner for an end effector (i.e., tooling mounted to the tool flange of the robot). The tool may be a gripper in a material handling application or it could be a welding gun, paint applicator, or tooling that is specific to other operations (Agheli, Qu, & Nestinger, 2014; Chen et al., 2008; Weiss & Klinger, 2017). A workcell may be configured such that end effectors are changed during operations to enable a single robot to serve as the macro positioner for multiple tasks. Reliability of positioning accuracy is required when using industrial robots.

Researchers from the National Institute of Standards and Technology (NIST) are focused on developing the necessary measurement science to verify and validate monitoring, diagnostic, and prognostic technologies to promote enhanced maintenance and control strategies within manufacturing operations (Pellegrino, Justiniano, Raghunathan, & Weiss, 2016; Weiss et al., 2015). Part of this effort is focused on measuring technologies supporting 6DOF robot workcells (Klinger & Weiss, 2018; Weiss & Klinger, 2017). This article presents NIST's development of a new methodology for the design and implementation of monitoring kinematic chains commonly found in robotic workcells. This method will enable the identification of degraded components that contribute to relative positioning accuracy errors between moving objects, tools, devices and other components. The proposed methodology is being developed and tested on a 6DOF industrial robot arm workcell use case developed at NIST. This methodology will offer a low-cost, minimally invasive means for a manufacturer to quickly assess the health of their robot workcell and ascertain where in the kinematic chain positional errors are originating. Ultimately, this research will be used to promote industry-driven standards to verify and validate health monitoring of workcell technologies (Vogl, Weiss, & Donmez, 2014a, 2014b; Weiss, Alonzo, & Weinman, 2017).

Klinger, Alexander; Weiss, Brian. "Examining Workcell Kinematic Chains to Identify Sources of Positioning Degradation." Paper presented at Annual Conference of the Prognostics and Health Management Society 2018, Philadelphia, PA, United States. September 24,

2018 - September 27, 2018.

Alexander Klinger et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The remainder of this paper is organized as follows. Section 2 presents background on the need for identifying points/areas of degradation within the kinematic chain along with some prior work. Section 3 provides a detailed discussion of the proposed kinematic chain analysis methodology. Section 4 discusses the kinematic chain inspection process to identify degradations. Section 5 highlights future work. Finally, Section 6 concludes the paper and acknowledges future work.

2. BACKGROUND

The Prognostics and Health Management (PHM) community is focused on advancing monitoring, diagnostic, and prognostic capabilities to maintain and/or improve asset availability, product/process quality, and productivity of a range of operations, including those within manufacturing. PHM has been actively applied to manufacturing factories with varying success (Helu & Weiss, 2016; Jin, Siegel, et al., 2016; Jin, Weiss, Siegel, & Lee, 2016). This has included the development and implementation of numerous strategies to minimize reactive maintenance and optimize the balance between preventive and predictive maintenance (Lee, Lapira, Bagheri, & Kao, 2013; Lee, Ni, Djurdjanovic, Qiu, & Liao, 2006; Peng, Dong, & Zuo, 2010; Vogl, Weiss, & Helu, 2016).

Most manufacturing robot workcells can be considered complex systems of systems. The overall physical system can be broken down into sub-systems, components, subcomponents, etc. Decomposing the physical elements of a robot workcell into a hierarchy of elements offers a means of defining boundaries that can drive maintenance activities. This physical hierarchy can be connected with functional and informational hierarchies to provide a means of understanding complex relationships and identifying metrics and measures of workcell health (Sharp & Weiss, 2018; Weiss & Qiao, 2017; Weiss, Sharp, & Klinger, 2018). The more complex a system, the more critical it becomes to understand its inherent relationships to see how the state of mechanical degradation of physical elements impacts process performance.

Research has been performed in monitoring the health, including positioning, of a robot, itself, separate from the workcell. Vision and laser tracker systems are two types of technologies that have been paired with PHM algorithms to enhance health awareness (Qiao & Weiss, 2017; Spiewak, Zaiss, & Ludwick, 2013; Wan, Song, Xu, Liu, & Chen, 2018). There is a need to expand the health monitoring capabilities beyond the robot, itself, and to consider the entire robot workcell. Understanding and monitoring the positioning health of all elements of kinematic chains is a necessary part of a comprehensive PHM system for robot workcells. Prior research has been performed in monitoring the health of robot workcells, yet research has not been found that focuses specifically on monitoring kinematic chain health. Existing robot monitoring approaches present unique

solutions that are either too high-level and lacking specific guidance (Lopez-Mellado & Alami, 1990) or rely upon potentially costly technology (Jokinen & Lastra, 2016; Novak-Marcincin, Torok, Janak, & Novakova-Marcincinova, 2014). NIST personnel believe that monitoring and analyzing the kinematic chain for positional degradation can be done in a cost-effective and minimallyintrusive manner to increase benefit for the robot workcell owners and users.

3. KINEMATIC CHAIN ANALYSIS

Kinematic chains can be documented at various levels of complexity for the same mechanism. At the most detailed level, a kinematic chain contains all bodies, including components and sub-components, fasteners, and other parts. The documentation of some kinematic chains may only include links between actuated joints or express entire actuated assemblies as single links.

It is beneficial to observe multiple points along a kinematic chain in a serial manipulator to monitor positioning accuracy and identify the source of positioning errors. By inspecting repeatability at multiple points along a chain, the source of repeatability degradation can be narrowed down to the component(s) located between adjacent inspection points. To facilitate this, a way of representing a kinematic chain is needed that provides sufficient detail to choose relevant inspection points to meet monitoring objectives. To develop this representation, a suggested starting point is to document the chain at the component level while identifying the relationships between components through their interfaces.

The kinematic chain is made of links which can be connected to other links or a reference frame. Each component or element of the system is considered a link in the kinematic chain. Each link has a left-hand side (LHS) and a right-hand side (RHS) which correspond to that component's interfaces with other links in the chain. It is important to note that LHS and RHS elements are physical pieces of the link, itself. For example, if a robot is a link, the LHS would likely be the robot's base since that would be physically connected to a mounting surface and the RHS would likely be the robot's tool flange since this is where an end effector would be joined. At the LHS and RHS boundaries of each link, permanent connections are represented by brackets '[,]' and intermittent connections with parenthesis '(,)'. For example, a gripper, which is permanently fixed to a robot tool flange, will have a '[' on its LHS corresponding to the permanent connection to a robot and a ')' on its RHS corresponding to an intermittent connection to a part as shown in Figure 1.

When individual links are joined to form a kinematic chain, the physical interfaces, represented by the LHS and RHS of adjacent links and the type of connection (permanent or intermittent), are easily identifiable as seen in Figure 2. A reference frame must exist at a boundary at either the far left

Klinger, Alexander; Weiss, Brian. "Examining Workcell Kinematic Chains to Identify Sources of Positioning Degradation." Paper presented at Annual Conference of the Prognostics and Health Management Society 2018, Philadelphia, PA, United States. September 24,

ANNUAL CONFERENCE OF THE PROGNOSTICS AND HEALTH MANAGEMENT SOCIETY 2018





or far right (or both) of the chain to provide a basis of measurement. Using this convention also allows for the kinematic chain to be represented in text. A single link can be represented as [LHS, Link, RHS] or more simply [Link]. For example, the chain in Figure 2(a) can be represented as 'Reference] [LHS 1, Link 1, RHS 1] [LHS 2, Link 2, RHS 2)' or in a simpler form 'Reference] [Link 1] [Link 2).'

A variety of components are typically chosen by a machine integrator for specific tasks during industrial robot workcell design and integration. This variety can pose a challenge in expressing the range of possible configurations. At





Figure 2. Kinematic chain representation of two links with key features identified. (a) Generic representation, (b) robot and gripper example.

minimum, every robotic workcell includes an industrial robot arm and an end effector. These components will serve as the starting point of the kinematic chain as shown in Figure 2(b). Not all kinematic chains remain intact throughout the entirety of a manufacturing process. In use cases where there are multiple, swappable end effectors, multiple chains will exist throughout the overall process. Likewise, an end effector may perform two or more functions. Each function will generate a unique kinematic chain which includes the interactions that are specific to those functions. Additional chains will need to be documented and analyzed in these scenarios.

Beyond the robot arm and end effector, additional workcell components are considered part of the kinematic chain if they impact the positional accuracy of any component or action within the process. These components include elements that are not actuated or actively controlled yet are critical in the relative positioning of parts and equipment during operation. For example, a hard stop may be used to orient a part. If the hard stop is not properly positioned, the part will not be oriented properly when aligned with the stop. Therefore, the hard stop should be included in the kinematic chain. Additional components can also include parts being worked on and the fixturing that is holding them.

3.1. NIST Use Case Kinematic Chains

A use case and testbed with a material handling robot and a path-following robot were developed and built at NIST to serve as a platform to support robotic workcell PHM research (Klinger & Weiss, 2018; Weiss & Klinger, 2017). The use case begins with the material handling robot, with a gripper end effector, picking parts from an input, and placing them on a work fixture (Figure 3(a)). The path following robot, with a pen holding end effector, then draws on the part (Figure 3(b)). Once the drawing operation is complete, the part is then picked by the material handling robot and placed in an output location (Figure 3(c)).



Figure 3. NIST use case: (a) Material handling robot moves a part to a work fixture. (b) Drawing robot draws on the part. (c) Material handling robot moves the completed part to the output

Klinger, Alexander; Weiss, Brian. "Examining Workcell Kinematic Chains to Identify Sources of Positioning Degradation." Paper presented at Annual Conference of the Prognostics and Health Management Society 2018, Philadelphia, PA, United States. September 24,

2018 - September 27, 2018.

(a) Use Case Kinematic Chain – Graphic Representation



World Frame][Robot Base, Robot, Tool Flange][Gripper Mounting, Gripper, Gripper Fingers)... ...(Part Geometry, Part, Part Geometry)(Fixture Geometry, Fixture, Fixture Mounting][World Frame

(c) Use Case Kinematic Chain - Short Text Representation

World Frame][Robot][Gripper)(Part)(Fixture][World Frame

Figure 4. The NIST use case material handling robot kinematic chain: (a) graphical representation, (b) long-form text representation, (c) short-form text representation.

As seen in Figure 4, for this use case, the world frame is the reference frame and is set by the table which the robot and fixtures are rigidly mounted on. The robot is the first link, followed by the gripper, then the part, and finally the fixture. The critical relationships in this kinematic configuration include the robot base to a table - a permanent connection shown in Figure 5(a), the robot tool flange to the gripper mounting - a permanent connection shown in Figure 5(b), the gripper fingers to the part - an intermittent connection shown in Figure 5(c), the part to the fixture - another intermittent connection, and the fixture to the table - a permanent connection highlighted in Figure 5(d).

When the robot is in position to pick or place a part, there is a constrained relationship between the part and the gripper, as well as the part and the fixture. The robot must position the gripper in a designed relationship to the fixture (and part) during part pick and placement because of these physical constraints. The interfaces of the intermittent connection between the gripper and part are shown as the RHS of the gripper (gripper fingers) and the LHS of the part (part geometry) respectively. Likewise, interfaces of the intermittent connection between the part and the fixture are shown as the geometries of the part and the fixture.

4. INSPECTING THE KINEMATIC CHAIN TO IDENTIFY DEGRADATION

Identifying and tracking degradation of workcell components is possible by measuring positioning repeatability of key points along kinematic chains. By inspecting the positioning repeatability of individual components relative to a reference frame, components which have degraded in their ability to position repeatedly beyond a design limit can be identified. The proposed inspection method can treat kinematic chains as open loop chains due to the nature of serial manipulators. There may be workcell configurations where the kinematic chain is represented as a closed loop, as in the NIST use case example, but by inspecting individual points along the chain



Figure 5. NIST use case material handling robot relationships: (a) robot base to the table, (b) robot tool flange to gripper mounting, (c) gripper fingers to part when on fixture, (d) fixture to table.

Klinger, Alexander; Weiss, Brian. "Examining Workcell Kinematic Chains to Identify Sources of Positioning Degradation." Paper presented at Annual Conference of the Prognostics and Health Management Society 2018, Philadelphia, PA, United States. September 24,

2018 - September 27, 2018.

this does not pose an issue. In cases where the chain is a closed loop, which is typically due to multiple components interfacing directly with the same reference frame as the robot, inspections may need to be carried out working from both the LHS and RHS reference frames. This will be discussed further and through example with the NIST use case.

Inspection of components starts at a reference frame. The measurement sensor(s) is in a known location to the reference frame and is assumed to be positioned and operating properly. Working left-to-right in the kinematic chain, the sensor(s) is fixed in the left most reference frame. Inspection begins at the RHS of the first component (link) with the testing of the position of the RHS of the component at a taught position. This assesses if the first component is maintaining sufficient repeatability to succeed. If the test is successful, then all parts of the first link are being positioned nominally. This is represented in Figure 6(a). After the first component is tested to be within specification the process continues, from left to right, by inspecting each component to confirm that each link is achieving its nominal test positions. After the RHS of the second link is tested successfully, the interface between the first and second link can be declared nominal and all parts of the second link are positioned nominally. This is represented in Figure 6(b).



Figure 6. Annotated generic link inspection method.

The decision of whether a component is being positioned nominally and a test is successful will depend on the designed allowance in the system. Due to tolerance stacking, it is not unusual to expect the allowance of components further from the reference in the kinematic chain to be higher and must be considered when choosing the sensor(s) used for inspection. It can also be beneficial to choose inspection points other than the RHS of links in the kinematic chain depending on the workcell. To help choose these points, an analysis of modes of positioning degradation is recommended.

4.1. Identifying Positioning Degradation Modes Along the Kinematic Chain

Likely sources of positional relationship issues can be identified once the kinematic chain and the interactions between components are mapped. The type and cause of positioning error will vary based upon the components within the workcell. Degradation modes can be found by assessing each link and the interactions between the links.

A link can be actuated, like a robot, or rigid, like a part. These two types of links can degrade in many ways which can be classified as either rigid body deformations, or degraded actuation and positioning. With an actuated link, such as a robot arm, if any joint in the robot is not positioning itself correctly, the tool flange and end effector will not be positioned correctly. These positioning errors can present themselves instantaneously after crashes or over time after wear. Likewise, deformation of rigid bodies can occur instantaneously or over time. For example, gripper fingers may bend when colliding with a foreign object in the workcell as compared to the fingers fatiguing and bending over time after repeated use. Each link is susceptible to degradation and will influence how that component and the rest of the workcell performs.

In addition to the positioning error modes from the links, themselves, the interactions between the links must be studied to assess the impact of link degradation on the relationships. Many different interactions can exist within the workcell. Couplings between components are either permanent or intermittent. Generally, the interface between two components is critical when the components' interaction is intermittent. This type of interaction requires the components to be in their designed positions and may involve a specific actuation for coupling. This means that if a component is positioned improperly, the geometry is inaccurate or unexpected, or a component cannot actuate as intended to successfully execute the interaction, the process can be considered degraded and may not perform nominally. For example, when a part is being grasped by a gripper, the part must have the expected geometry and be properly positioned for the gripper which must be in the expected position and actuate as designed to successfully grasp the part.

4.2. Identifying Positioning Degradation Modes in the NIST Use Case

A variety of positioning degradation modes are identified within the NIST use case. Degradation modes are identified for the components of the kinematic chain as well as the interactions between components. Starting at the world frame and examining the kinematic chain of Figure 4, moving left to right, modes are identified. Table 1 summarizes the

Klinger, Alexander; Weiss, Brian. "Examining Workcell Kinematic Chains to Identify Sources of Positioning Degradation." Paper presented at Annual Conference of the Prognostics and Health Management Society 2018, Philadelphia, PA, United States. September 24, 2018 - September 27, 2018.
identified degradation modes. The modes internal to components, including robot wear, gripper wear, bad parts, and fixture wear are expressed as between the LHS and RHS of their respective components. The connection is assumed as the only failure mode between two links that are permanently fixed to one another and other modes are considered as part of the components. For the intermittent interactions between the part and the gripper, and the part and the fixture, any of the components can contribute to positioning error leading to multiple modes listed.

Table 1. Use case relationships table	
---------------------------------------	--

Kinematic Chain Section	Degradation Mode(s)
World Frame][Robot Base	Loose Connection
[Robot Base, ROBOT, Robot Tool Flange]	Robot Wear
Robot Tool Flange][Gripper Mounting	Loose Connection
[Gripper Mounting, GRIPPER, Gripper Fingers]	Gripper Wear
Gripper Fingers)(Part Geometry	Gripper Wear, Bad Part*
[PART]	Bad Part*
Part Geometry)(Fixture Geometry	Bad Part*, Bad Fixture**
[Fixture Geometry, Fixture, Fixture Mounting]	Bad Fixture**
Fixture Mounting][World Frame	Loose Connection

*From either part inaccuracy or damage

**From either fixture positioning inaccuracy or fixture wear/damage

4.3. Selection of Key Measurement Points to Identify Points / Areas of Degradation and Inspection Method

Discrete points along the kinematic chain can be selected for inspection to detect positioning deviations from nominal or verify nominal positioning. These key measurement points need to be tested in one or more degrees of freedom to determine if they can be positioned within designed limits. Each point should be chosen to provide information on the source of a positioning error when measured in a specific order. As discussed in Section 3, a good starting point in a LHS-referenced kinematic chain is the RHS of each link in the chain. Inspecting the RHS of each link will provide insight into the repeatability of each component. However, if a component such as an end effector is actuated, it may be beneficial to inspect both the RHS and another point earlier in the kinematic chain. Inspecting before an actuation point of a particular component, the source of repeatability degradation can be isolated to within the component or to the connection between the component and the preceding component. Additionally, an actuated component can be inspected in multiple configurations or states to provide further isolation of degradation.

4.4. Selection of Key Measurement Points in the NIST Use Case

Applying this methodology to the NIST use case, the key points to be selected should be the robot tool flange, gripper fingers, and part geometry because they are the RHS of the components of the chain. Because the fixture link is directly connected to the world frame on its RHS, the LHS of the fixture link, fixture geometry, should be selected as a key point. Examining the gripper further, it is actuated and can have multiple states (i.e., open, gripping a part, closed) throughout its operations and can be inspected in more places than only the RHS. By selecting the gripper body as a key point before the finger actuation in the chain, the gripper mounting can be tested independently of the finger actuation. Likewise, testing the fingers in both an open and closed state, the operation and positioning of the gripper fingers can be tested independently of the gripper mounting. To test the part geometry, the system must grasp an ideal part so the part position can be tested in-situ. To test the fixture geometry, a sensor can be mounted to or embedded in the fixture. These points are represented in Figure 7.



Figure 7. Use case key measurement point selection graphic.

Appropriate tolerancing or analysis thresholding can then be chosen for the pass / fail classification of each of the key points selected. Given the tolerance and monitoring requirements of the use case, appropriate sensors can be chosen. The key points can be tested in a sequence following the kinematic chain, left-to-right, once the sensors are integrated. The feedback from testing allows a series of binary decisions to be made to identify degraded components. Decision trees and testing criteria will be dependent on the use case. The decision tree and testing flow for the NIST use case can be seen in Figure 8. This series of tests can be performed at any interval and records kept, aiding in the tracking of degradations and predicting future degradation. If the tests are done in the prescribed order, according to Figure 8, a failed test will automatically indicate where, in the kinematic chain, a degradation is originating, providing a target component to repair or replace. After a failure of a test point, all points tested further along the kinematic chain will be unreliable without a full understanding of the cause of the failed test.

Klinger, Alexander; Weiss, Brian. "Examining Workcell Kinematic Chains to Identify Sources of Positioning Degradation." Paper presented at Annual Conference of the Prognostics and Health Management Society 2018, Philadelphia, PA, United States. September 24,



Figure 8. Binary decision tree to identify degraded component.

5. FUTURE WORK

Measuring or testing the selected key points along the kinematic chain can be done in many ways. Regardless of the specific details, there will be a sensor and a target, which the sensor measures. Measurements can be made through physical contact, visual imaging, or another type of measurement technology. The key characteristic of the sensor used for the kinematic inspection method outlined is that the sensor outputs information on the position of a specific point or object. Future work includes the assessment of measurement methods and sensors for examining robot workcell kinematic chains. This work will involve understanding the implications of using discrete and analog sensing methods, data collection strategies, and prognostic implications. This work will be done in parallel with the implementation of the discussed method on the NIST use case with commercially available and/or purpose-built sensors.

6. CONCLUSION

This paper presents a method to document kinematic chains within a robot workcell enabling the selection of key measurement points for testing to gain intelligence on workcell health. Testing of key measurement points along the kinematic chain can provide intelligence on the degradation of positioning of components. By choosing the measurement points with the help of a degradation mode analysis, a test or series of tests can be designed to identify the degraded component(s) which are potentially degrading the workcell's process. Initial efforts in this research are very promising where next steps are currently in process. This includes the development of a novel sensor to pair with the kinematic chain methodology to meet the effort's goals of providing a low-cost, minimally invasive means to identify degradations within the kinematic chain. Preliminary data sets will be captured upon integration of the novel sensor with the kinematic chain testing methodology. These initial data sets will offer critical insight as to the practical feasibility of kinematic chain methodology and the deployment/usage of the novel sensor.

NIST DISCLAIMER

The views and opinions expressed herein do not necessarily state or reflect those of NIST. Certain commercial entities, equipment, or materials may be identified in this document to illustrate a point or concept. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

REFERENCES

- Agheli, M., Qu, L., & Nestinger, S. S. (2014). SHeRo: Scalable hexapod robot for maintenance, repair, and operations. Robotics and Computer-Integrated Manufacturing, 30(5), 478-488. doi:10.1016/j.rcim.2014.03.008
- Ahmad, R., & Plapper, P. (2016). Safe and Automated Assembly Process using Vision Assisted Robot Manipulator. Procedia CIRP, 41, 771-776. doi:http://dx.doi.org/10.1016/j.procir.2015.12.129
- Chen, H., Fuhlbrigge, T., & Li, X. (2008). Automated industrial robot path planning for spray painting process: a review. Paper presented at the IEEE International Conference on Automation Science and Engineering, 2008. CASE 2008..
- Helu, M., & Weiss, B. A. (2016). The current state of sensing, health management, and control for small-tomedium-szed manufacturers. Paper presented at the ASME 2016 Manufacturing Science and Engineering Conference, MSEC2016.
- Jin, X., Siegel, D., Weiss, B. A., Gamel, E., Wang, W., Lee, J., & Ni, J. (2016). The present status and future

Klinger, Alexander; Weiss, Brian. "Examining Workcell Kinematic Chains to Identify Sources of Positioning Degradation." Paper presented at Annual Conference of the Prognostics and Health Management Society 2018, Philadelphia, PA, United States. September 24,

growth of maintenance in US manufacturing: results from a pilot survey. Manufacturing Rev., 3, 10.

- Jin, X., Weiss, B. A., Siegel, D., & Lee, J. (2016). Present Status and Future Growth of Advanced Maintenance Technology and Strategy in US Manufacturing. International Journal of Prognostics and Health Management, 7 (Special Issue on Smart Manufacturing PHM)(Sp5), 18.
- Jokinen, J., & Lastra, J. L. M. (2016). Implementation of nonintrusive monitoring and fault diagnosis in industrial robot system. Paper presented at the 2016 IEEE 14th International Conference on Industrial Informatics (INDIN).
- Klinger, A. S., & Weiss, B. A. (2018). Robotic Work Cell Test Bed to Support Measurement Science for PHM. Paper presented at the 2018 ASME Manufacturing Science and Engineering Conference (MSEC), College Station, Texas.
- Lee, J., Lapira, E., Bagheri, B., & Kao, H.-a. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. Manufacturing Letters, 1(1), 38-41.
- Lee, J., Ni, J., Djurdjanovic, D., Qiu, H., & Liao, H. (2006). Intelligent prognostics tools and e-maintenance. Computers in Industry, 57(6), 476-489.
- Lopez-Mellado, E., & Alami, R. (1990). A failure recovery scheme for assembly workcells. Paper presented at the 1990 IEEE International Conference on Robotics and Automation.
- Novak-Marcincin, J., Torok, J., Janak, M., & Novakova-Marcincinova, L. (2014). Interactive Monitoring of Production Process with Use of Augmented Reality Technology. Applied Mechanics and Materials, 616, 19.
- Pellegrino, J., Justiniano, M., Raghunathan, A., & Weiss, B. A. (2016). Measurement Science Roadmap for Prognostics and Health Management for Smart Manufacturing Systems. NIST Advanced Manufacturing Seriess (AMS).
- Peng, Y., Dong, M., & Zuo, M. J. (2010). Current status of machine prognostics in condition-based maintenance: a review. The International Journal of Advanced Manufacturing Technology, 50(1-4), 297-313.
- Qiao, G., & Weiss, B. A. (2017). Accuracy Degradation Analysis for Industrial Robot Systems. Paper presented at the ASME International Manufacturing Science and Engineering Conference, Los Angeles, California.
- Sharp, M., & Weiss, B. A. (2018). Hierarchical modeling of manufacturing work cell to promote а contextualized PHM information across multiple levels. Manufacturing Letters, 15, 46-49.
- Shi, Y. H., Wang, G. R., Li, & G. J. (2007). Adaptive robotic welding system using laser vision sensing for underwater engineering. In 2007 IEEE International

Conference on Control and Automation, Vols 1-7 (pp. 1795-1800).

- Spiewak, S., Zaiss, C., & Ludwick, S. J. (2013, November 15-21, 2013). High Accuracy, Low-Invasive Displacement Sensor (HALIDS). Paper presented at the ASME 2013 International Mechanical Engineering Congress and Exposition, IMECE 2013, San Diego, CA, United states.
- Vogl, G. W., Weiss, B. A., & Donmez, M. A. (2014a). Standards for prognostics and health management (PHM) techniques within manufacturing operations. Paper presented at the Annual Conference of the Prognostics and Health Management Society 2014, Fort Worth, Texas, USA.
- Vogl, G. W., Weiss, B. A., & Donmez, M. A. (2014b). Standards Related to Prognostics and Health Management (PHM) for Manufacturing (NISTIR 8012). Retrieved from Gaithersburg, Maryland, USA: http://dx.doi.org/10.6028/NIST.IR.8012
- Vogl, G. W., Weiss, B. A., & Helu, M. (2016). A review of diagnostic and prognostic capabilities and best practices for manufacturing. Journal of Intelligent Manufacturing. doi:10.1007/s10845-016-1228-8
- Wan, A., Song, L., Xu, J., Liu, S., & Chen, K. (2018). Calibration and compensation of machine tool volumetric error using a laser tracker. International Journal of Machine Tools and Manufacture, 124, 126-133.
- Weiss, B. A., Alonzo, D., & Weinman, S. D. (2017). Summary Report on a Workshop on Advanced Monitoring, Diagnostics, and Prognostics for Manufacturing Operations.
- Weiss, B. A., & Klinger, A. S. (2017). Identification of Industrial Robot Arm Work Cell Use Cases and a Test Bed to Promote Monitoring, Diagnostic, and Prognostic Technologies. Paper presented at the 2017 Annual Conference of the Prognostics and Health Management (PHM) Society, St. Petersburg, FL.
- Weiss, B. A., & Qiao, G. (2017). Hierarchical Decomposition of a Manufacturing Work Cell to Promote Monitoring, Diagnostics, and Prognostics. Paper presented at the ASME 2017 International Manufacturing Science and Engineering Conference (MSEC2017), Los Angeles, California.
- Weiss, B. A., Sharp, M., & Klinger, A. (2018). Developing a hierarchical decomposition methodology to increase manufacturing process and equipment health awareness. Journal of Manufacturing Systems.
- Weiss, B. A., Vogl, G. W., Helu, M., Qiao, G., Pellegrino, J., Justiniano, M., & Raghunathan, A. (2015). Measurement Science for Prognostics and Health Management for Smart Manufacturing Systems: Key Findings from a Roadmapping Workshop. Paper presented at the Annual Conference of the

Klinger, Alexander; Weiss, Brian. "Examining Workcell Kinematic Chains to Identify Sources of Positioning Degradation." Paper presented at Annual Conference of the Prognostics and Health Management Society 2018, Philadelphia, PA, United States. September 24,

Prognostics and Health Management Society 2015, Coronado, CA.

BIOGRAPHIES



Alexander Klinger has a B.S. in Mechanical Engineering (2013) from Lehigh University, Bethlehem, Pennsylvania and a Masters of Engineering in Mechanical Engineering (2017) from the University of Maryland, College Park, Maryland. He is currently a mechanical engineer working on the

Prognostics, Health Management and Control project within the Engineering Laboratory (EL) at the National Institute of Standards and Technology (NIST). Prior to his role at NIST, he was a mechanical engineer at the United States Postal Service (USPS) Engineering Headquarters, working on automation and robotic systems. He has also worked in various engineering roles in the manufacturing and entertainment industry.



Dr. Brian A. Weiss has a B.S. in Mechanical Engineering (2000), Masters in Engineering (2003), and Ph.D. in Mechanical Engineering (2012) from the University of Maryland, College Park, Maryland. He is currently the Project Leader of the Prognostics, Health Management, and Control (PHMC)

project within the Engineering Laboratory (EL) at the National Institute of Standards and Technology (NIST). Prior to his leadership role in the PHMC project, he spent 15 years conducting performance assessments across numerous military and first response technologies including autonomous unmanned ground vehicles; tactical applications operating on Android devices; advanced soldier sensor free-form, two-way, speech-to-speech technologies; translation devices for tactical use; urban search and rescue robots; and bomb disposal robots. He is currently a member of the American Society of Mechanical Engineers (ASME). His efforts have earned him numerous awards including a Department of Commerce Gold Medal (2013), Silver Medal (2011), Bronze Medals (2004 & 2008), and the Jacob Rabinow Applied Research Award (2006).

Klinger, Alexander; Weiss, Brian. "Examining Workcell Kinematic Chains to Identify Sources of Positioning Degradation." Paper presented at Annual Conference of the Prognostics and Health Management Society 2018, Philadelphia, PA, United States. September 24,

Development of Tools to Evaluate the Impact of New Residential Fire Protection Technology

Paul Reneke, Morgan Bruns, Richard Peacock, Thomas Cleary and Stanley Gilbert

Paul.reneke@nist.gov, 301 975-6696

National Institute of Standards and Technology, Gaithersburg, MD 20899

There are a number of fire technologies that could be beneficial in improving fire safety in residential structures including barrier fabrics in residential upholstered furniture, new fire alarm standards, and nano-composite flame retardants to name a few. However, there does not exist an accepted method for measuring absolute or relative improvements in fire safety from new fire protection technologies. The National Institute of Standards and Technology (NIST) has undertaken an effort to create tools and identify gaps in data, tools, and knowledge to help quantify improvements in fire safety. Toward this effort NIST is conducting an analysis of the recent changes in ANSI/UL 217 2015 [1] to estimate the relative impact on occupant safety. This analysis includes definition of residential fire scenarios in structures of various sizes, Monte Carlo sampling of inputs for modeling fire growth and spread, and statistical analysis of the results to quantify the potential impact on fire safety of smoke alarms meeting ANSI/UL 217 2015. We will discuss the preliminary results of this analysis as well as identify gaps in data and knowledge that need to be filled to fully understand the impact on community fire safety.

One of the issues that has existed with detectors is the fact that ion detectors respond significantly faster than photo detectors to flaming smoke but much slower to smoke from smoldering fires. The new ANSI/UL 217 2015 has additional requirements for flaming and smoldering polyurethane foam smokes so that new smoke alarms that pass the standard must perform better than typical photoelectric alarms in flaming fires and better than typical ionization alarms in smoldering smoke. To model the effects of smoke alarm performance, we added two features to the Consolidated Fire and Smoke Transport model known as CFAST[2], the NIST computer fire model used in the analysis. First, we had to allow fires to start off as smoldering fires then switch over to flaming fires, which includes having CFAST produce and track two types of smoke. The second change was adding a new type of smoke detector that can react differently to the two types of smoke. These features were added and are now available in the latest version of CFAST, version 7.3.0.

After a review of the literature on residential fires we found that egress from one- and two-family structures is very different from large and/or commercial buildings. In large commercial buildings much of the egress time is dominated by the flow effects of large crowds and the distances that need to be traveled. The impact of individual behavior may average out. In one- and two-family residences the egress distances are relatively short and the number of people is small, which means that egress is dominated by individual behaviors. However, there isn't a model that attempts to calculate residential egress. The solution is to follow previous analysis, set an expert's best estimate of an appropriate egress time, and see if the alarms give enough time.

One behavior that needs to be considered in order to truly understand residential fires is that people, in their own homes, will often attempt to fight the fires and statistics indicate they are successful more times than not[3]. While issues such as liability may make fire safety officials unwilling to suggest that people should consider fighting a fire they discover, the reality is that people will often do just that. In an attempt to start developing an understanding of this phenomenon we identified a point depending on the fire heat release rate we would estimate the occupants would have a high probability of successfully fighting the fire and after which they would have a low probability of putting the fire out. Since there is very little data about residents fighting fires it is difficult to know how realistic this analysis is, but it is an important part of the residential fire problem and needs to be addressed.

We conducted an analysis of National Fire Incident Reporting System (NFIRS) data to determine what characteristics made fires the deadliest. The results showed that fires in living rooms and bedrooms are among the most dangerous. We also found that fires ignited from smoking were the deadliest fires. Data was available on upholstered chair fires starting from both flaming sources and smoldering sources. Data on mattress fires for both old mattresses and for mattresses meeting the most recent standards were also available. This resulted in four different model fires, each with its own set of distributions so that the detectors would see a range of environments.

To develop floor plans, we started with the American Home Survey (AHS) that the Census Bureau puts out every two years. From the AHS we took the distribution of total floor area and given a total floor area we have a distribution of the number of rooms. The floorplan was generated by assuming any two rooms had a 50/50 probability of being connected. There was an added requirement that the completed graph or layout of the building had to be a planar graph, meaning a graph that can be put in a plane or drawn on a piece of paper without the edges crossing. The planar graph requirement was put in place to keep the floorplans from being connected in such a way that was physically impossible.

Part of doing a Monte Carlo analysis is to perform a large enough number of runs so that the statistics are reproducible. To do this requires automating the process. Using the Python computer language, a purpose specific program was developed to generate, run and process the cases using CFAST. This allowed us to generate data from tens of thousands of cases. After generating a significant amount of data, we looked for methods that would allow us to determine when we had enough data for the specific questions.

After generating the data, the issue is how to analyze and present the data in a statistically rigorous and meaningful way that people using the analysis can understand what it is saying and make decisions the analysis was meant to inform.

Among the findings is that it is difficult to accurately assess actual improvements due to proposed fire safety technology improvements with the limited knowledge on typical room sizes and configuration in a range of residential structures. Additional information on typical floorplans is needed to improve the assessments. A more significant need is to evolve the definition of the Required Safe Egress Time(RSET) side of an ASET/RSET analysis for residential evacuation. The Available Safe Egress Time(ASET) calculations are basically the

same for residential and commercial structures. Residential evacuation is behaviorally more complex than commercial occupancy, where the majority of occupants do not have significant attachments or responsibility to the structure. We propose calculating a <u>Residential Safe Egress</u> <u>Time for residential structures</u>, which takes into account such activities as investigation, firefighting, and assisting others to escape in addition to the travel time included in typical RSET calculations.

- 1. Underwriters Laboratories, I., ANSI/UL 217-2015: Standard for Safety Smoke Alarms. 2015: Northbrook, IL.
- Peacock, R.D., et al., CFAST, Consolidated Model of FIre Growth and Smoke Transport (Version 7), Volume 1: Technical Reference Guide, in Technical Note 1889v1. 2015, National Institute of Standards and Technology: Gaithersburg, MD.
- 3. Greene, M.A. and C. Andres, 2004-2005 National Smaple Survey of Unreported Residential Fires. 2009, U.S. Consumer Product Safety Commission: Washington, DC.