NISTIR 8145

A Comparative Analysis of NIST Charpy Machines and Internal Reference Materials

Enrico Lucon
Ray Santoyo

This publication is available free of charge from:
http://dx.doi.org/10.6028/NIST.IR.8145

National Institute of Standards and Technology
U.S. Department of Commerce

NISTIR 8145

A Comparative Analysis of NIST Charpy Machines and Internal Reference Materials

Enrico Lucon
Ray Santoyo
Applied Chemicals and Materials Division
Material Measurement Laboratory

This publication is available free of charge from:
http://dx.doi.org/10.6028/NIST.IR.8145

August 2016

U.S. Department of Commerce
Penny Pritzker, Secretary

National Institute of Standards and Technology
Willie May, Under Secretary of Commerce for Standards and Technology and Director

Abstract

This report documents the results of a series of comparative analyses which were conducted on the Charpy machines located at NIST in Boulder, Colorado. The analyses were performed both on historical data, collected over a period of more than 20 years, and newly obtained test results from almost 300 instrumented and non-instrumented impact tests on verification specimens of low energy (13 J to 20 J), high energy (88 J to 136 J), and super-high energy (176 J to 244 J). The comparative analyses were completed for the purpose of understanding the limits of calibrating instrumented strikers for use in dynamic test through conventional static calibration procedures.

Test results, obtained both in the past and within the current study, show that one of the so-called "master machines" (used to establish reference values for ASTM verification specimens) delivers consistently lower energy values than the other two master machines in the 15 J - 20 J range. This is suspected to be due to its higher stiffness and hammer design (C-type), which increases its natural frequency of vibration and causes early fracture in a quasi-brittle (low-energy) test.

The extremely large number of instrumented Charpy tests performed within this investigation (more than 200) allowed us to assess the reliability of the conventional static calibration procedure for our instrumented strikers, as compared to different approaches for the adjustment/correction of measured force values. Even though these approaches were found beneficial in improving the between-machine consistency, statistically significant differences remained nonetheless. It appears that, in order to reliably calibrate an instrumented striker, forces must be applied at loading rates that are equivalent to those encountered during actual tests (dynamic calibration of instrumented strikers). This should become a clear objective of the NIST Charpy program going forward.

Keywords

Calibration of instrumented strikers; Charpy absorbed energy; Charpy maximum force; Charpy machines; dynamic striker calibration; instrumented Charpy tests; loading rate; static striker calibration.

Table of Contents

Abstract i
Keywords i
Table of Contents ii
1. Introduction 1
2. The NIST Charpy machines 2
3. The NIST instrumented strikers 3
3.1 Instrumented striker for the TK machine (S-79-R) 4
3.2 Instrumented striker for the TO2 and TO3 machines (TO-JS1) 4
3.3 Instrumented striker for the MPM machine (MPM-JS2) 5
4. Analysis of historical data for the master machines 6
4.1 Low-energy specimens 6
4.2 High-energy specimens 10
4.3 Super-high-energy specimens 14
4.4 Summary of findings 17
5. Qualification of NIST Internal Reference Materials to assess the performance of Charpy machines and instrumented strikers 18
5.1 Selection of IRMs 18
5.1.1 Low-energy IRM: LL-137 19
5.1.2 High-energy IRM: HH-107 19
5.1.3 Super-high-energy IRM: SH-38 19
5.2 Characterization of the selected IRMs 19
5.2.1 Characterization of LL-137 21
5.2.2 Characterization of HH-107 22
5.2.3 Characterization of SH-38 23
5.2.4 Reference values and expanded uncertainties 24
6. Example of use of Internal Reference Materials: TO3 (machine and instrumented striker) 26
7. Statistical analyses of IRM test results: a comparative study of NIST Charpy machines 27
7.1 Comparisons among multiple machines: ANOVA analyses 27
7.1.1 Absorbed energy 27
7.1.2 Maximum force 28
7.2 One-on-one machine comparisons: *t*-tests 29
7.2.1 Absorbed energy 29

7.2.2 Maximum force 30
8. Discussion: calibration of an instrumented striker 32
8.1 Comparison between KV and W_t 36
8.2 Effect of dynamic force adjustment 37
8.3 Modified dynamic force adjustment 39
8.4 Force correction based on machine compliance 40
8.5 Overview of force correction results 41
9. Conclusions 43
Acknowledgements 44
References 44

1. Introduction

The Charpy impact test was introduced at the turn of the 20th century through the fundamental works of Russell [1] and Charpy [2], but it was only during and after World War II that this test was extensively used to explain the large number of ship failures which occurred during the war. In 1948, the National Bureau of Standards (which became NIST 40 years later, in 1988) published a report [3] that highlighted the importance of Charpy testing for establishing the fracture behavior of several fractured plates removed from some of the ships that had exhibited structural failures.

In the late 40s and early 50s, very significant scatter was observed in Charpy test results, the origin of which was not clear. At the time, it was common opinion that this scatter was intrinsic to the test itself, rather than a consequence of poor maintenance of the test machines and poor machining quality of the specimens. In a paper published in 1961, Fahey [4] listed the most important contributions to the obtainment of erroneous impact values and consequently high data scatter, as follows:

- improper installation of the machine;
- incorrect dimensions of the anvil supports and striking edge;
- excessive friction in moving parts;
- looseness of mating parts;
- insufficient clearance between specimen ends and side supports;
- poorly machined test specimens;
- improper cooling and testing techniques.

An earlier fundamental study published by Driscoll in 1955 [5] had shown that the scatter of Charpy results from industrial machines could be dramatically reduced after proper maintenance was conducted. Based on original NIST data, it could be shown that a scatter level of 2 % could be achieved when all variables were kept under control. This demonstrated that the Charpy impact test was not inherently scattered, and could be successfully used as an acceptance test. In other words, Driscoll's study showed that not all machines in service at the time exhibited satisfactory performance, but also that most machines (around 90 %) could perform acceptably provided tests were conducted carefully and the machines were in good working condition.

The same study also proposed verification limits corresponding to the larger of 1.4 J (1 ft-lb) and 5 %. These were adopted by ASTM E23 in 1964, when the standard was revised to include for the first time indirect verification testing, *i.e.*, verifying the machine by the use of reference specimens with certified absorbed energy. The requirement to periodically verify a Charpy machine by the use of verification specimens is believed to have significantly improved the performance of machines throughout the world. Indirect verification testing is today required by all international Charpy test standards.

In the United States, the Army Materials and Mechanics Research Center (AMMRC) produced and distributed standardized reference specimens for the indirect verification of Charpy machines, until NIST took over the program in 1989. The three Charpy machines owned by the Army (referred to as the "master machines") were transferred to NIST in Boulder, Colorado, and have been used since to establish reference values of absorbed energy for verification specimens of three energy levels: low energy (13 J to 20 J), high energy (88 J to 136 J), and super-high energy (176 J to 244 J). In accordance with ASTM E23, a Charpy machine is certified for acceptance

testing if its results (average of 5 tests per energy level) agree within the larger of 1.4 J or 5 % with the reference values established on the NIST master machines.

The establishment of this program has succeeded in consolidating a system whereby Charpy test results in the US and the rest world are traceable to NIST master machines. The absolute accuracy of such results, and their SI[1]-traceability, depends on the accuracy of the master machines. Verifying the agreement between the master machines is of paramount importance, in order to maintain a consistent Charpy verification system and also as a quality check of the program itself. This is one of the goals of the investigation presented in this report.

2. The NIST Charpy machines

The characteristics of the five Charpy machines located at NIST in Boulder are listed in Table 1. Note that the Charpy lab in Boulder is also equipped with a small-scale Charpy machine used to test miniaturized Charpy specimens [6]. This machine is not covered in this report.

Table 1 – Charpy machines at NIST in Boulder, Colorado.

Manufacturer[2]	Model	Machine ID	Pendulum type	Capacity (J)	Hammer weight (N)	Impact velocity (m/s)	Hammer length (m)	Falling height (m)	Falling angle (°)
Tokyo Koki Seizosho	1C-36	TK	C	359.5	295.3	4.89	0.90	1.22	110.7
Tinius-Olsen	74	TO2	U	358.6	267.6	5.12	0.90	1.34	119.2
Satec	SI-1K3	SI3	U	409.1	296.6	5.20	0.80	1.38	136.3
Tinius-Olsen	84	TO3	U	406.6	266.1	5.12	0.90	1.53	134.1
MPM Technologies	N/A	MPM	Z	953.6	626.5	5.47	0.91	1.52	131.9
Satec	SI-1K	SI2	U	324.4	237.5	5.18	0.80	1.37	135.0

In reference to the information in Table 1:

(a) The first three machines listed (TK, TO2, SI3) are defined as the master machines.

(b) The last machine listed (SI2) was one of the master machines until 2003, when it was replaced by the higher-capacity SI3. SI2 is no longer used, and is not covered in this report (although some historical data from this machine are covered in Section 4).

(c) The pendulum type (4^{th} column) refers to the shape of the hammer end which supports the striker. The most common designs are C-type and U-type (see Figure 1, (a) and (b) respectively).

(d) The capacity of each machine (5^{th} column) corresponds to the available potential energy for a full swing.

With the exception of SI3 and SI2, all Charpy machines are equipped with instrumented strikers, even though when establishing reference values of absorbed energy for verification

[1] International System of measurement units.

[2] Trade names and manufacturers are mentioned in this report only to accurately describe NIST activities. Such inclusion neither constitutes not implies endorsement by NIST or by the U.S. government.

specimens on the master machines, only non-instrumented strikers are used. The TO2 and TO3 machines use the same instrumented striker.

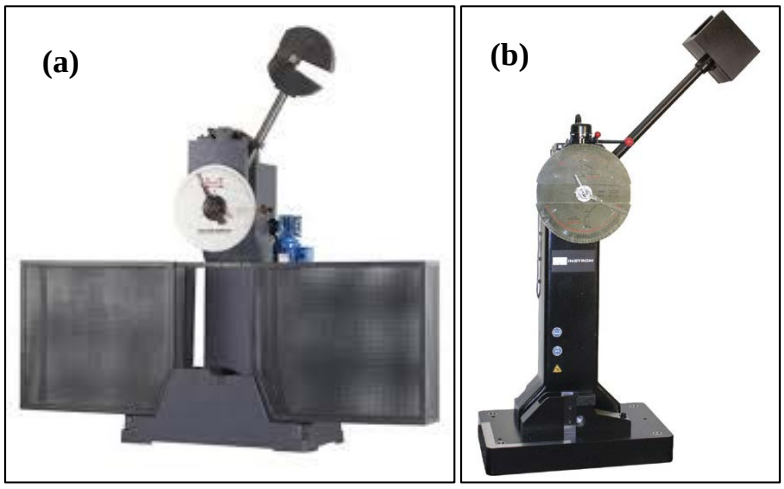

Figure 1 – Examples of C-type (a) and U-type (b) pendulums.

3. The NIST instrumented strikers

An instrumented Charpy striker is equipped with strain-gages that measure the elastic deformations occurring during the impact test, as a result of the interaction between striker and specimen. By establishing the calibration factor (or curve) between strain-gage output and applied force, one is able to obtain a full force vs. time (and force vs. displacement) record for every instrumented Charpy test conducted.

Some of the instrumented strikers used at NIST were commercially purchased, and were supplied with factory calibration curves/factors. Other strikers were custom-made. All the strikers were also calibrated (statically) at NIST.

Typically, a static calibration curve is obtained by applying force to the striker by means of a universal testing machine and recording the corresponding output of the striker, after going through any amplification stage. This is also the recommended approach mentioned in the principal test standards for instrumented impact testing (ASTM E2298 and ISO 14556).

Applied force values (measured by a calibrated load cell) are then fitted as a function of striker output with a linear function of the type:

$$F = A \cdot V + B \tag{1}$$

where F is force (in kN) and V is striker output (in V). Since the intercept B can be eliminated by simply zeroing the signal when the striker is unloaded, the only coefficient that is actually needed for signal conversion is the slope A (conversion factor). If the coefficient of determination R^2 of eq. (1) is less than 0.999, a second-order polynomial of the form

$$F = A \cdot V^2 + B \cdot V + C \tag{2}$$

should be used. The intercept C in eq. (2) can be neglected in the conversion for the reason explained above.

3.1 Instrumented striker for the TK machine (S-79-R)

The original instrumented striker for the TK machine (S-79) was accidentally damaged during its static calibration, and new strain-gages had to be installed. The calibration curve obtained for the re-gaged striker (S-79-R) is shown in Figure 2.

The conversion coefficient for this striker is 10.474.

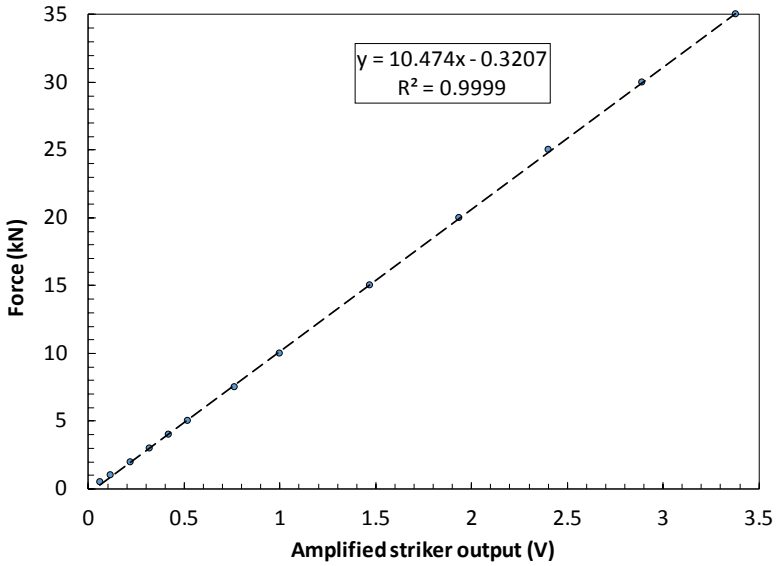

Figure 2 – Static calibration curve for the TK instrumented striker (S-79-R).

3.2 Instrumented striker for the TO2 and TO3 machines (TO-JS1)

The instrumented striker with identification TO-JS1 can be used on both the TO2 and TO3 machines. The static calibration curve obtained is shown in Figure 3.

The conversion coefficient for this striker is 7.478.

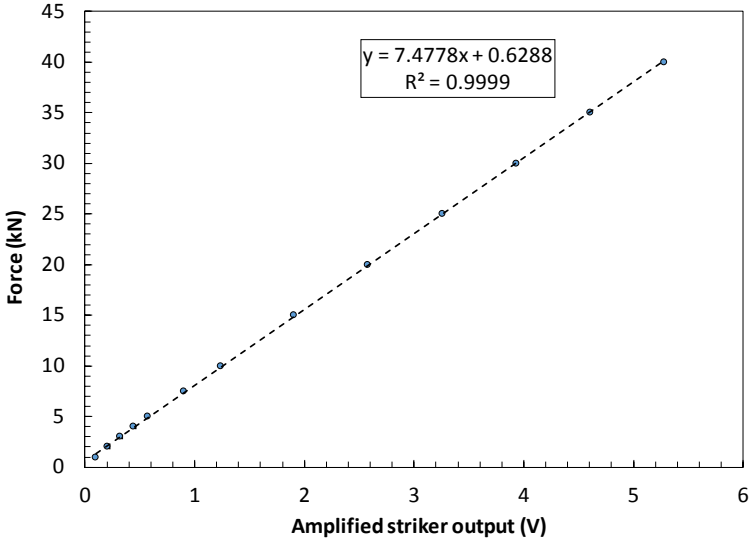

Figure 3 – Static calibration curve for the TO2/TO3 instrumented striker (TO-JS1).

3.3 Instrumented striker for the MPM machine (MPM-JS2)

The first instrumented striker which had been manufactured for the MPM machine (MPM-JS1) was replaced with a nominally identical striker (MPM-JS2) in 2015, due to excessive wear. The static calibration curve obtained for the MPM-JS2 striker is shown in Figure 3.

The conversion coefficient for this striker is 7.439. A significant deviation from linearity is visible in Figure 3, particularly in the low force range ($F \leq 10$ kN).

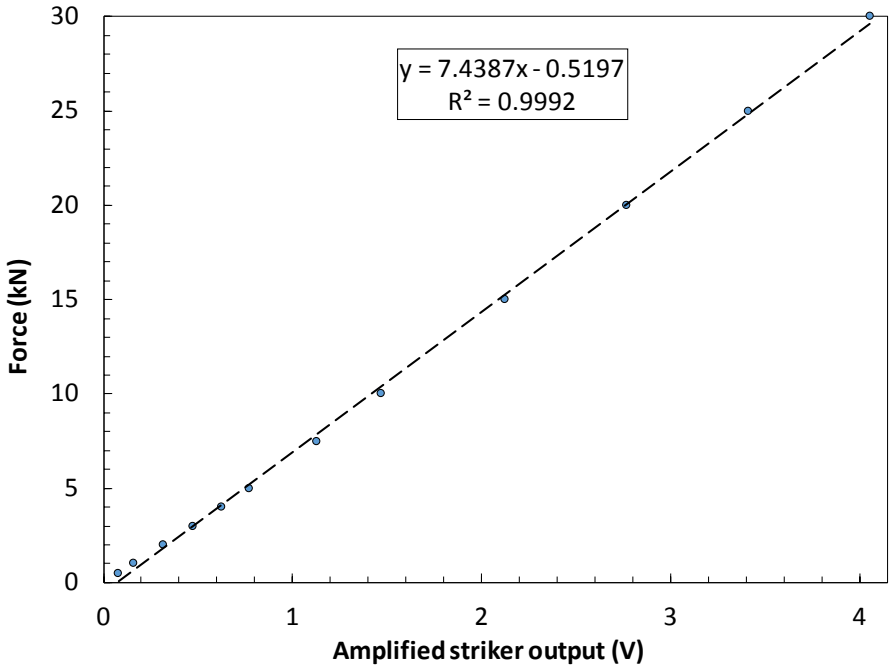

Figure 4 – Static calibration curve for the MPM instrumented striker (MPM-JS2).

4. Analysis of historical data for the master machines

Test results on verification lots for the master machines were collected from May 1995 to February 2016 and statistically analyzed in order to investigate possible trends and statistical differences between individual machines. Collected data consist of average values and standard deviations of absorbed energies calculated over 25 specimens tested on each machine for the qualification of low-energy, high-energy, and super-high-energy verification lots. The results analyzed are those obtained from the so-called "pilot" lots[3].

For each energy level, the analyses were split into two time frames: up to 2003 and after 2003, when SI2 was replaced by SI3. We statistically analyzed the differences between machines in terms of both mean values and coefficients of variation (*CV* = standard deviation/mean, in %). The coefficient of variation is a unitless measure of variability about a mean, and can be used in place of standard deviation to comparing data sets with different means or units.

For both mean values and coefficients of variation, the statistical tool used was the analysis of variance (ANOVA), developed by statistician and evolutionary biologist Ronald Fisher [7]. In its simplest form, ANOVA provides a statistical test of whether or not the means of several groups are equal, and therefore generalizes Student's *t*-test to more than two groups. The statistical significance of the differences between machines is provided by the value of the statistic *F*: if the calculated value of *F* is larger than the critical value F_{crit}, the machines are statistically different. Otherwise, the machines can be considered statistically equivalent. All the analyses were conducted at a significance level α = 5 % (corresponding to a confidence level of 95 %). The analyses were conducted under the (reasonable) assumption that the distributions of absorbed energy results were normal.

4.1 Low-energy specimens

The average values and coefficients of variation for TK, TO2, and SI2 (1995-2003) at the low-energy level are plotted in Figure 5 and Figure 6. The same data for TK, TO2, and SI3 (2003-2016) are illustrated in Figure 7 and Figure 8. In all the figures, the distribution along the X-axis is chronological (*i.e.*, based on the time each lot was qualified).

The results of the ANOVA analyses on average values and coefficients of variation are presented in Table 2 and Table 3 (TK, TO2, and SI2) and Table 4 and Table 5 (TK, TO2, and SI3).

Table 2 - ANOVA results for the mean values of TK, TO2, and SI2 at the low-energy level.

Machine	$\overline{KV}$ (J)	*F*	F_{crit}	Outcome
TK	15.24	26.709	3.085	The machines are statistically different
TO2	16.84			
SI2	16.66			

[3] A "pilot" lot is composed of 75 specimens from a given verification lot, which are tested on the master machines (25 per machine) to preliminarily assess the quality of the material. If the results are found to be acceptable, additional specimens are manufactured and 75 more specimens are tested ("production" lot). The reference value for the verification specimens is established based on the results of both the pilot and production lots, or (if pilot and production lots are statistically different) based on the results of the production lot alone.

Table 3 - ANOVA results for the coefficients of variation of TK, TO2, and SI2 at the low-energy level.

Machine	$\overline{CV}$	F	F_{crit}	Outcome
TK	4.92 %	2.400	3.085	The coefficients of variation of the machines are not statistically different
TO2	4.82 %			
SI2	4.25 %			

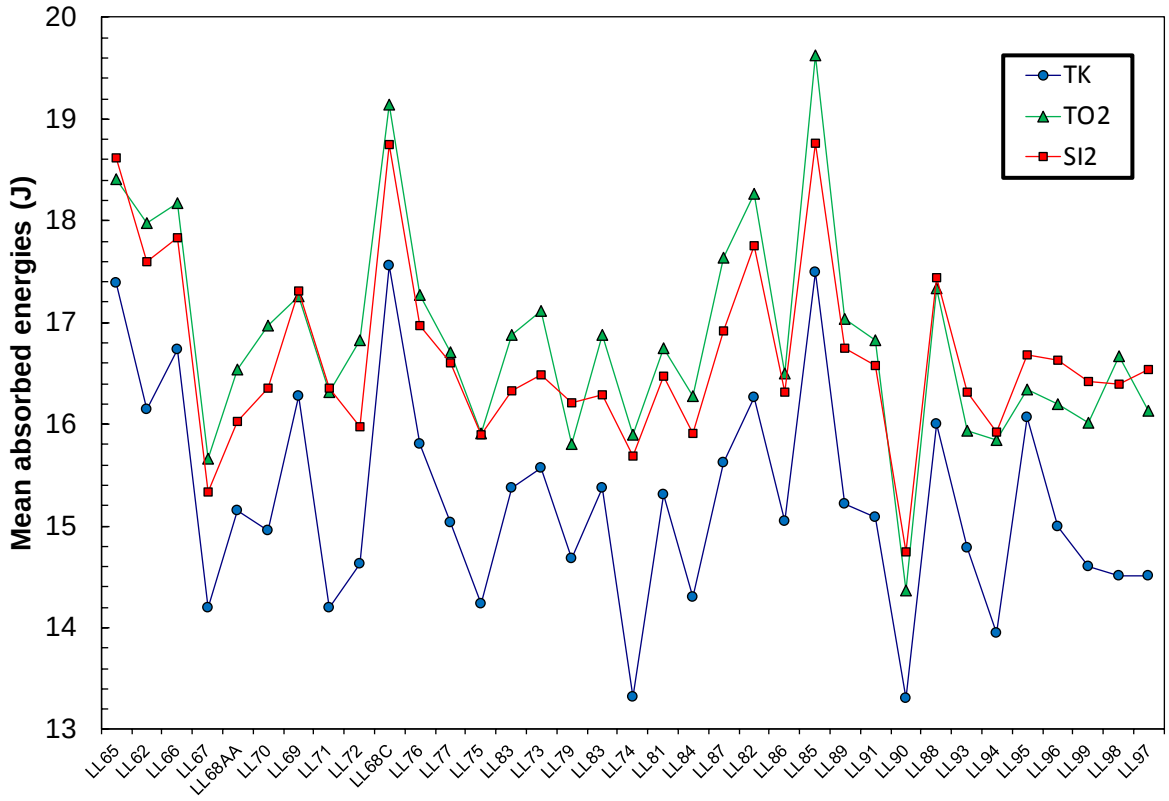

Figure 5 – Average absorbed energy values at the low-energy level for TK, TO2, and SI2.

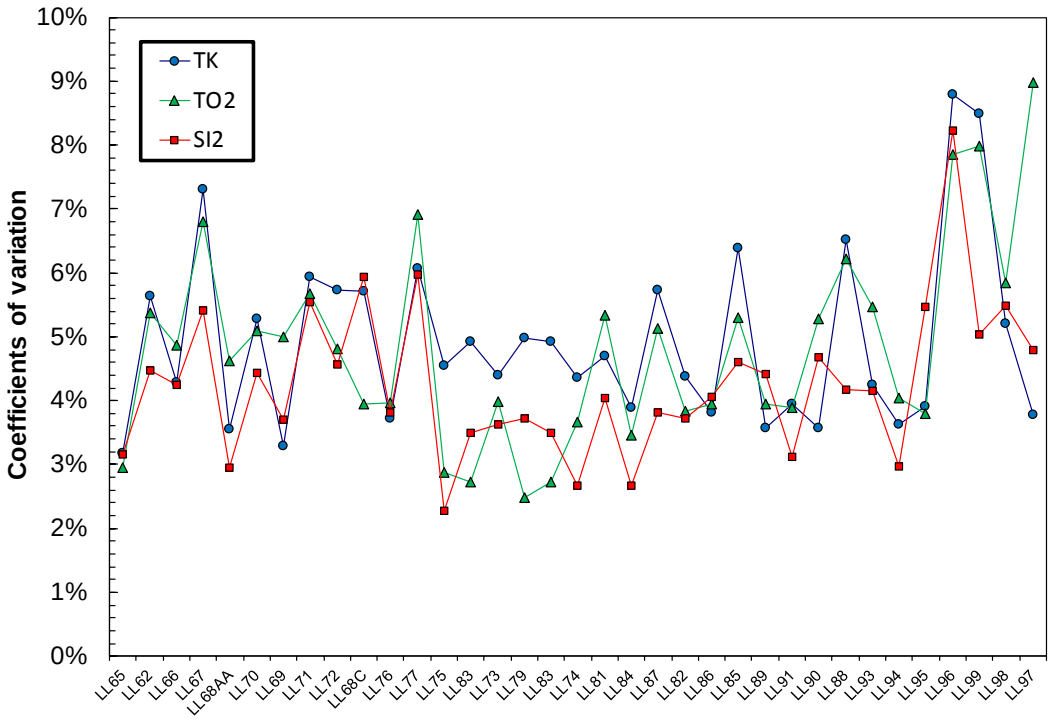

Figure 6 – Coefficients of variation at the low-energy level for TK, TO2, and SI2.

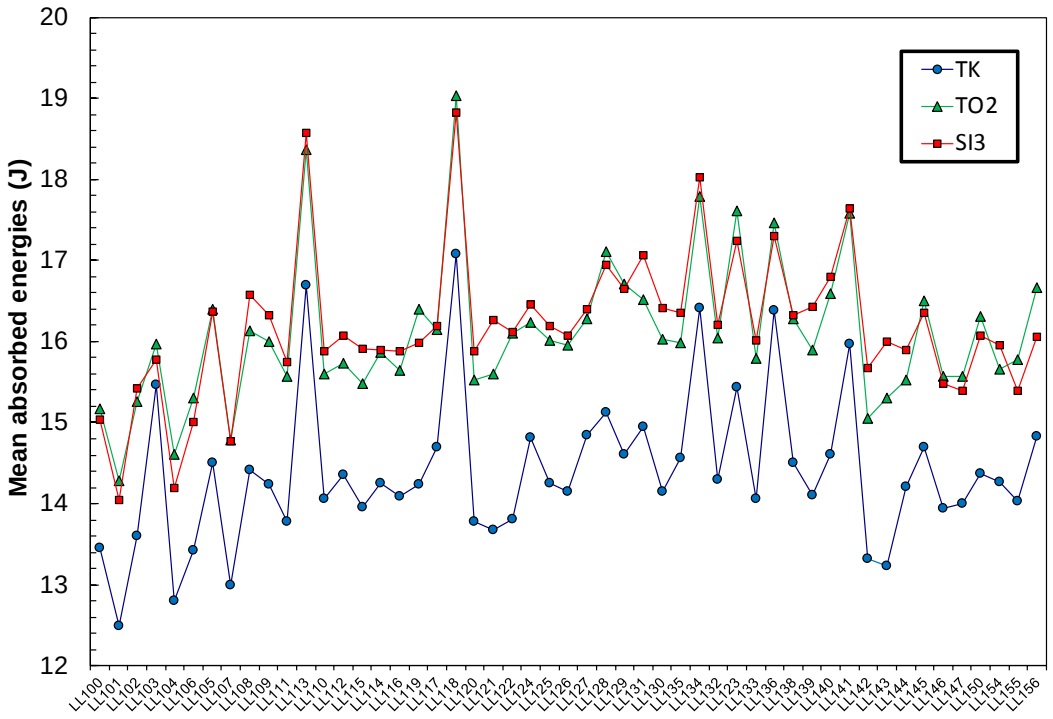

Figure 7 - Average absorbed energy values at the low-energy level for TK, TO2, and SI3.

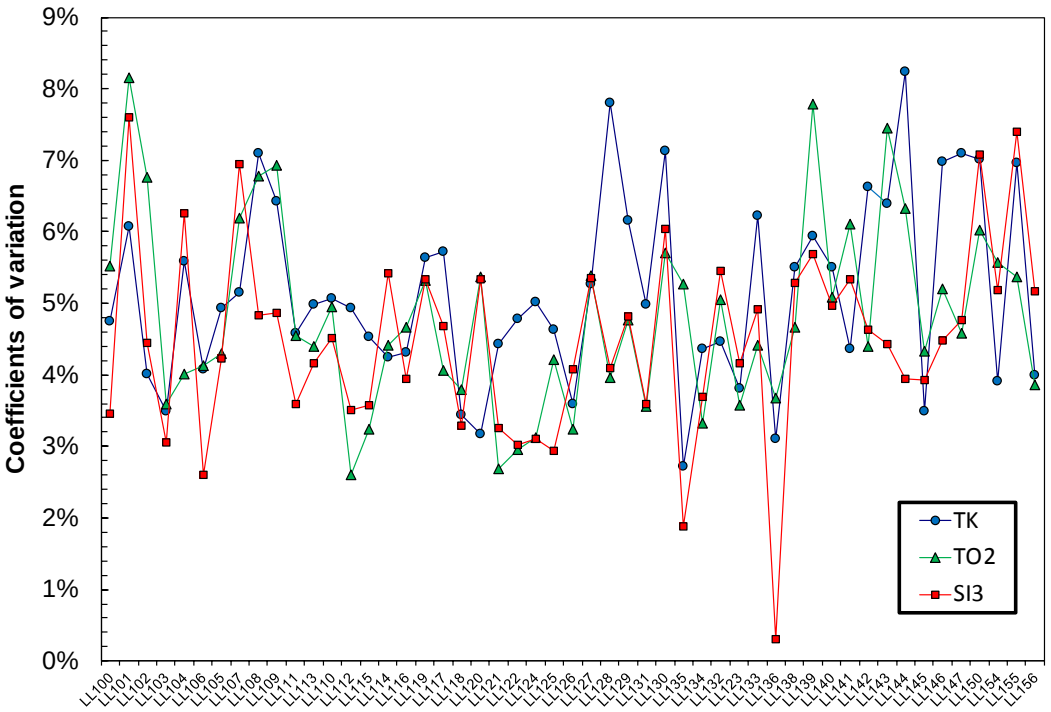

Figure 8 – Coefficients of variation at the low-energy level for TK, TO2, and SI3.

Table 4 - ANOVA results for the mean values of TK, TO2, and SI3 at the low-energy level.

Machine	$\overline{KV}$ (J)	F	F_{crit}	Outcome
TK	14.39	63.475	3.056	The machines are statistically different
TO2	16.09			
SI3	16.19			

Table 5 - ANOVA results for the coefficients of variation of TK, TO2, and SI3 at the low-energy level.

Machine	$\overline{CV}$	F	F_{crit}	Outcome
TK	5.15 %	3.273	3.056	The coefficients of variation of the machines are statistically different
TO2	4.81 %			
SI3	4.48 %			

The examination of Figure 5 and Figure 7 indicates a tendency of TK to provide systematically lower energy values than TO2 and SI2 (or SI3). Based on the results of the ANOVA analyses (Table 2 and Table 4), the difference is statistically significant (conversely, TO2 and SI2/SI3 are not). The average difference between TK and the mean of TO2/SI2 (or TO2/SI3) is $\Delta KV = 1.51$ J ± 0.37 J (or $\Delta KV = 1.75$ J ± 0.31 J). Note that additional ANOVA analyses indicated no difference between TO2 and SI2 ($F = 0.607 < F_{crit} = 3.982$) or between TO2 and SI3 ($F = 0.280 < F_{crit} = 3.936$).

When plotting the difference between TK and mean{TO2-SI2/SI3} as a function of test date (Figure 9), a slightly increasing trend is observed.

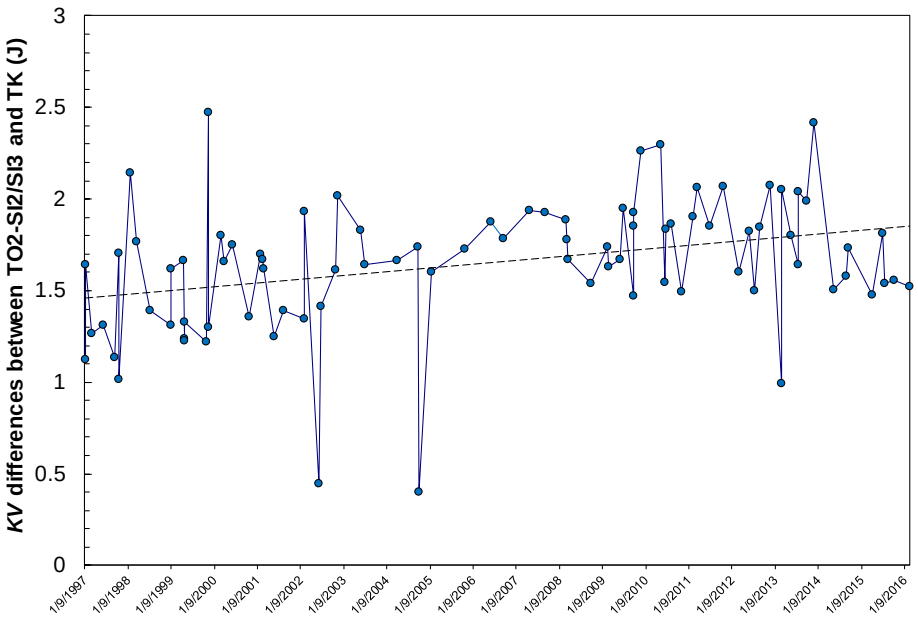

Figure 9 - KV differences between TK and the average of the other master machines.

An investigation by Manahan *et al.* [8], conducted in cooperation with NIST, offers a possible explanation to the systematically lower values yielded by the TK machine.

Calculations performed on a simple two-mass, two-spring model of the striker/specimen assembly in [8] suggest that this difference in absorbed energy can be explained in terms of frequency shift in the applied force, caused by a stiffer design of the striker assembly in the TK machine. Indeed, this machine was found to be stiffer (less compliant) than TO2 or SI3 when the compliance of all the NIST Charpy machines was measured [9]. As shown in [8], a 20 % higher stiffness of the striker assembly results in a 7 % increase in the natural frequency of the striker/specimen assembly, and therefore causes the critical fracture force to be reached at a smaller specimen displacement, which in turn corresponds to less absorbed energy to fracture the specimen. These findings show that the test machine design (*e.g.*, C-type instead of U-type hammer) can interact with low-energy specimens to materially affect the fracture behavior of the material, by promoting earlier fracture and therefore lower *KV* values. This is a major point, as it implies that absorbed energy is not a material "property", but depends on the design of the test machine. As a consequence, a comparison between absorbed energies obtained on the same material is not meaningful, unless the type/design of the Charpy machines used is taken into account.

4.2 High-energy specimens

The average values and coefficients of variation for TK, TO2, and SI2 at the high-energy level are plotted in Figure 10 and Figure 11. The same data for TK, TO2, and SI3 are illustrated in Figure 12 and Figure 13.

The results of the ANOVA analyses on average values and coefficients of variation are presented in Table 6 and Table 7 (TK, TO2, and SI2) and Table 8 and Table 9 (TK, TO2, and SI3).

Table 6 - ANOVA results for the mean values of TK, TO2, and SI2 at the high-energy level.

Machine	$\overline{KV}$ (J)	*F*	F_{crit}	Outcome
TK	96.08	0.197	3.078	The machines are not statistically different
TO2	95.37			
SI2	94.93			

Table 7 - ANOVA results for the coefficients of variation of TK, TO2, and SI2 at the high-energy level.

Machine	$\overline{CV}$	*F*	F_{crit}	Outcome
TK	2.42 %	1.062	3.078	The coefficients of variation of the machines are not statistically different
TO2	2.43 %			
SI2	2.21 %			

Table 8 - ANOVA results for the mean values of TK, TO2, and SI3 at the high-energy level.

Machine	$\overline{KV}$ (J)	*F*	F_{crit}	Outcome
TK	100.18	1.129	3.054	The machines are not statistically different
TO2	97.97			
SI3	97.89			

Table 9 - ANOVA results for the coefficients of variation of TK, TO2, and SI3 at the high-energy level.

Machine	$\overline{CV}$	*F*	F_{crit}	Outcome
TK	3.23 %	2.853	3.054	The coefficients of variation of the machines are not statistically different
TO2	2.98 %			
SI3	2.86 %			

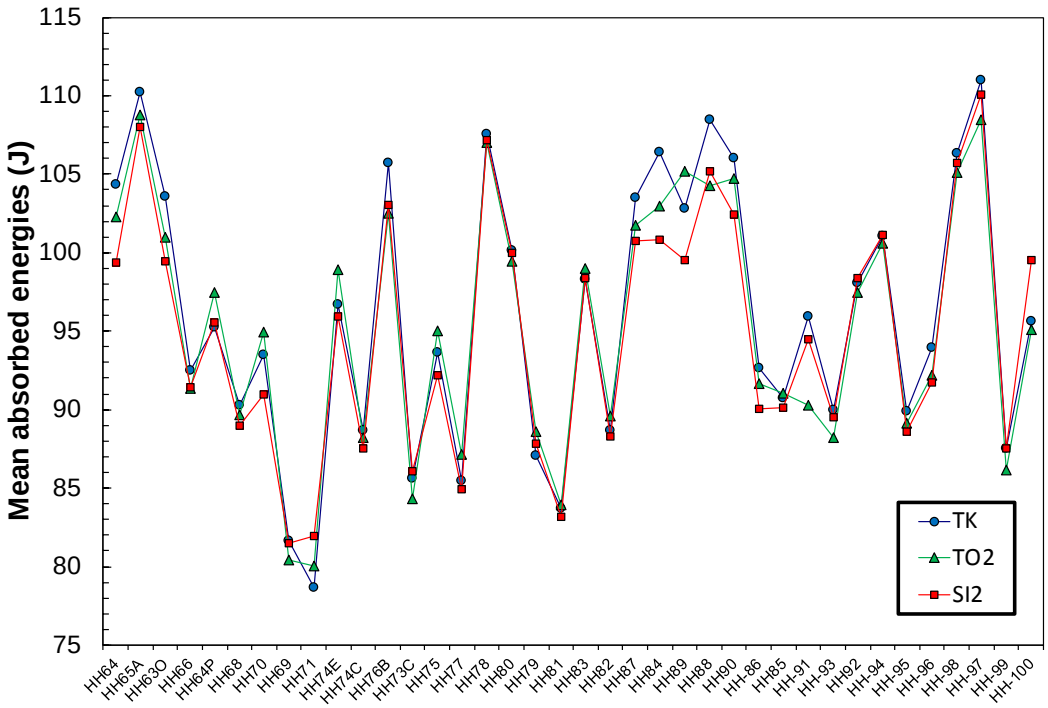

Figure 10 - Average absorbed energy values at the high-energy level for TK, TO2, and SI2.

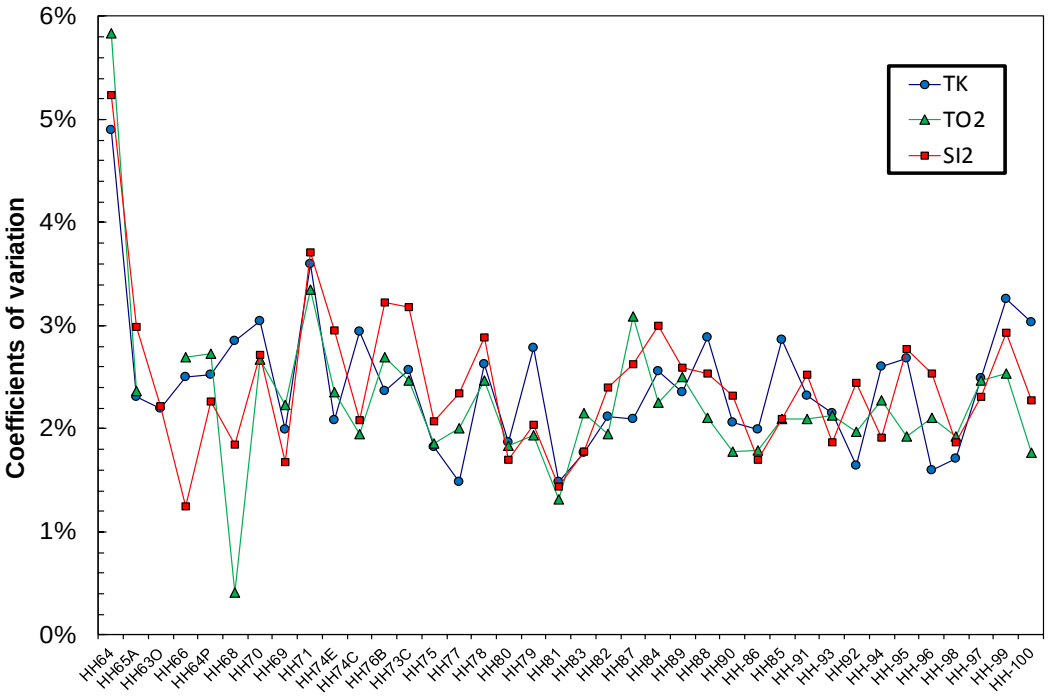

Figure 11 – Coefficients of variation at the high-energy level for TK, TO2, and SI2.

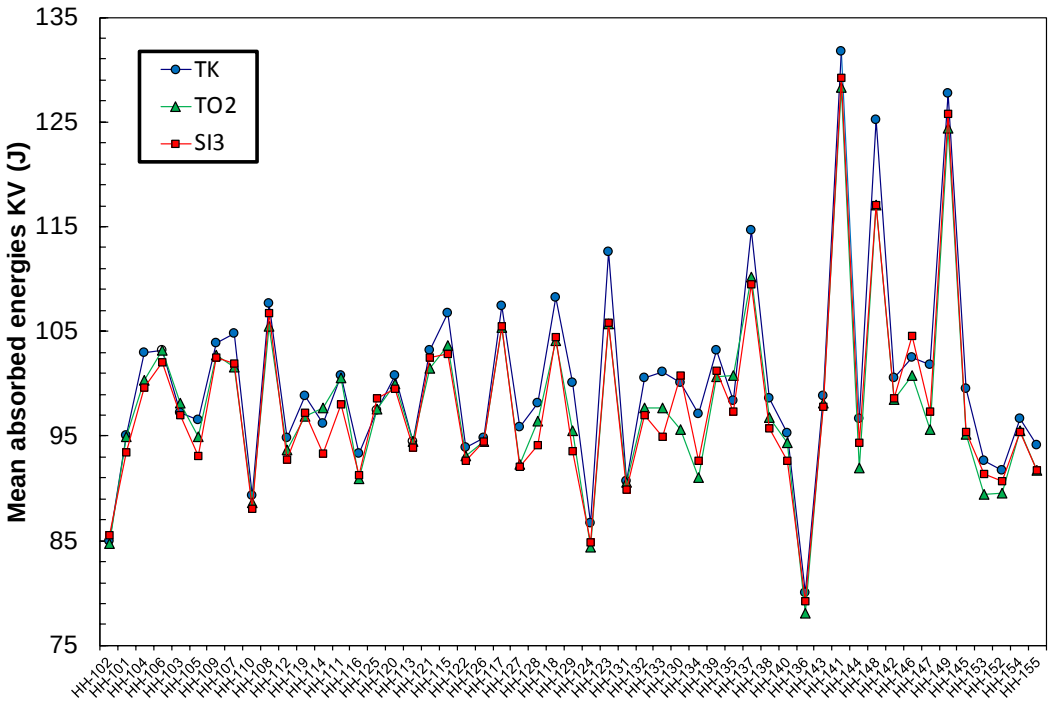

Figure 12 - Average absorbed energy values at the high-energy level for TK, TO2, and SI3.

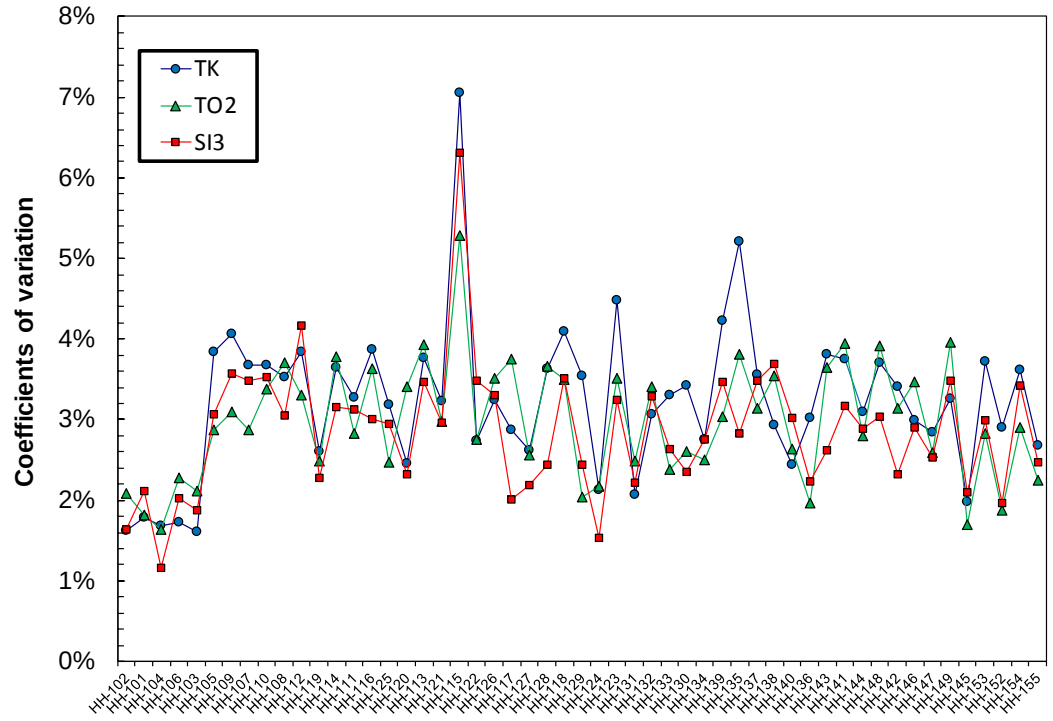

Figure 13 – Coefficients of variation at the high-energy level for TK, TO2, and SI3.

As demonstrated by the ANOVA analyses conducted, the NIST master machines are not statistically different at the high-energy level.

4.3 Super-high-energy specimens

The average values and coefficients of variation for TK, TO2, and SI2 at the super-high-energy level are plotted in Figure 14 and Figure 15. The same data for TK, TO2, and SI3 are illustrated in Figure 16 and Figure 17.

The results of the ANOVA analyses on average values and coefficients of variation are presented in Table 10 and Table 11 (TK, TO2, and SI2) and Table 12 and Table 13 (TK, TO2, and SI3).

Table 10 - ANOVA results for the mean values of TK, TO2, and SI2 at the super-high-energy level.

Machine	$\overline{KV}$ (J)	F	F_{crit}	Outcome
TK	224.77	0.023	3.136	The machines are not statistically different
TO2	224.68			
SI2	225.70			

Table 11 - ANOVA results for the coefficients of variation of TK, TO2, and SI2 at the super-high-energy level.

Machine	$\overline{CV}$	F	F_{crit}	Outcome
TK	2.37 %	1.118	3.136	The coefficients of variation of the machines are not statistically different
TO2	2.19 %			
SI2	2.31 %			

Table 12 - ANOVA results for the mean values of TK, TO2, and SI3 at the super-high-energy level.

Machine	$\overline{KV}$ (J)	F	F_{crit}	Outcome
TK	222.98	1.973	3.555	The machines are not statistically different
TO2	217.83			
SI3	227.60			

Table 13 - ANOVA results for the coefficients of variation of TK, TO2, and SI3 at the super-high-energy level.

Machine	$\overline{CV}$	F	F_{crit}	Outcome
TK	2.97 %	0.267	3.555	The coefficients of variation of the machines are not statistically different
TO2	2.70 %			
SI3	3.01 %			

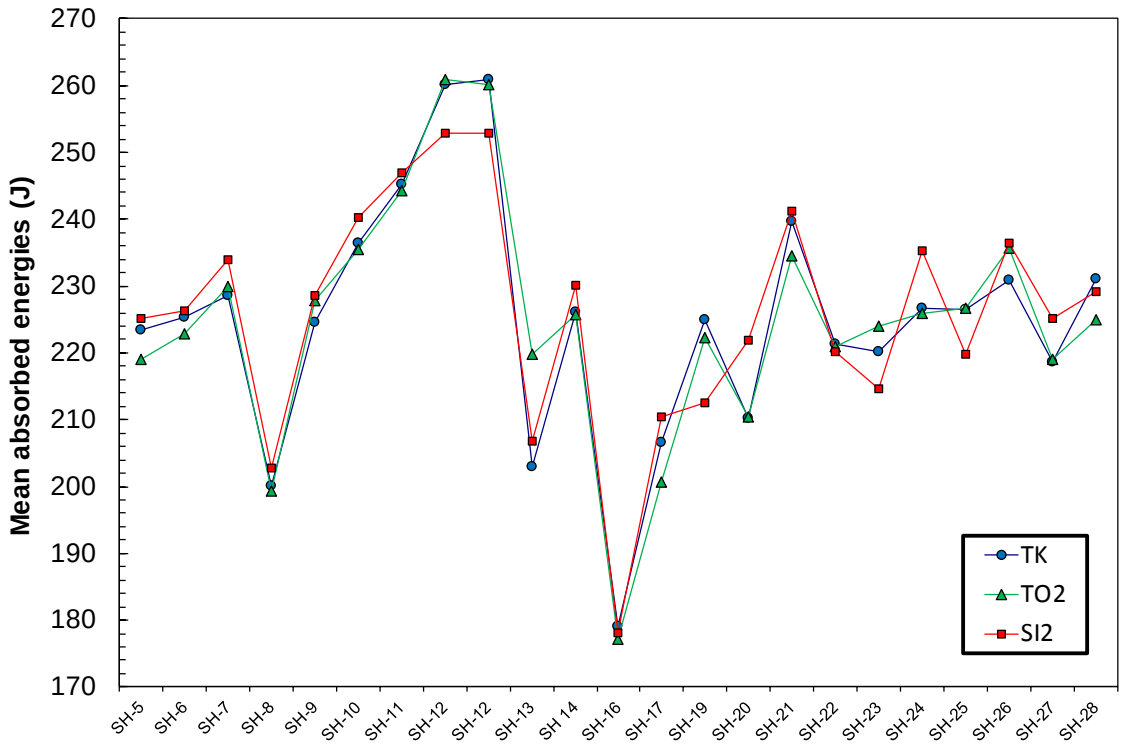

Figure 14 - Average absorbed energy values at the super-high-energy level for TK, TO2, and SI2.

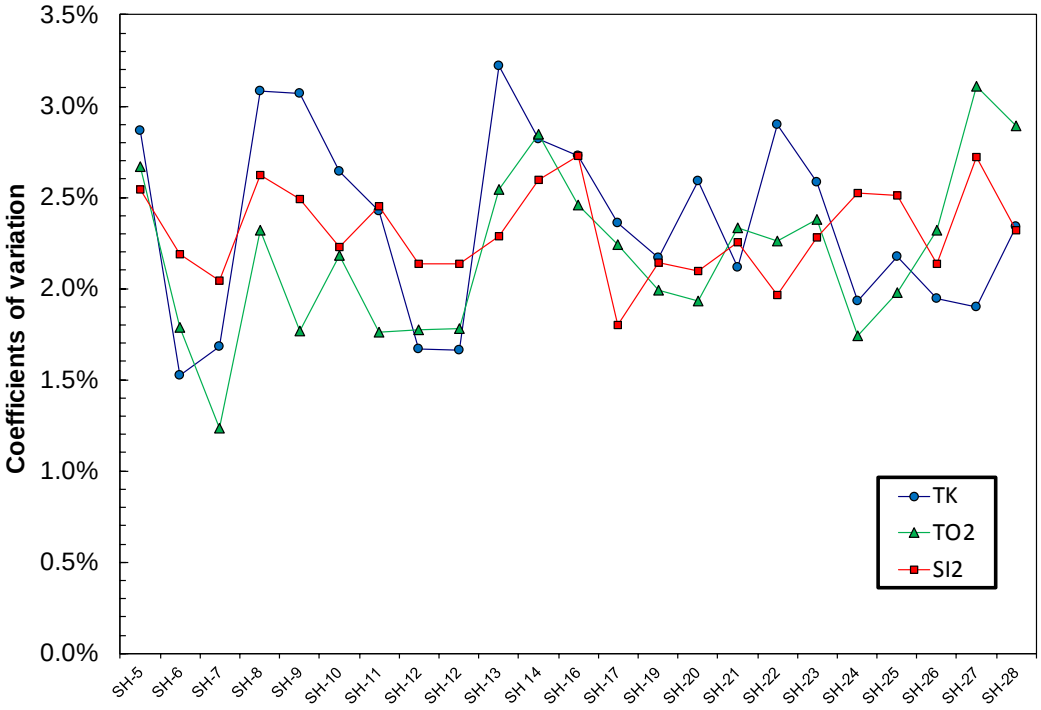

Figure 15 – Coefficients of variation at the super-high-energy level for TK, TO2, and SI2.

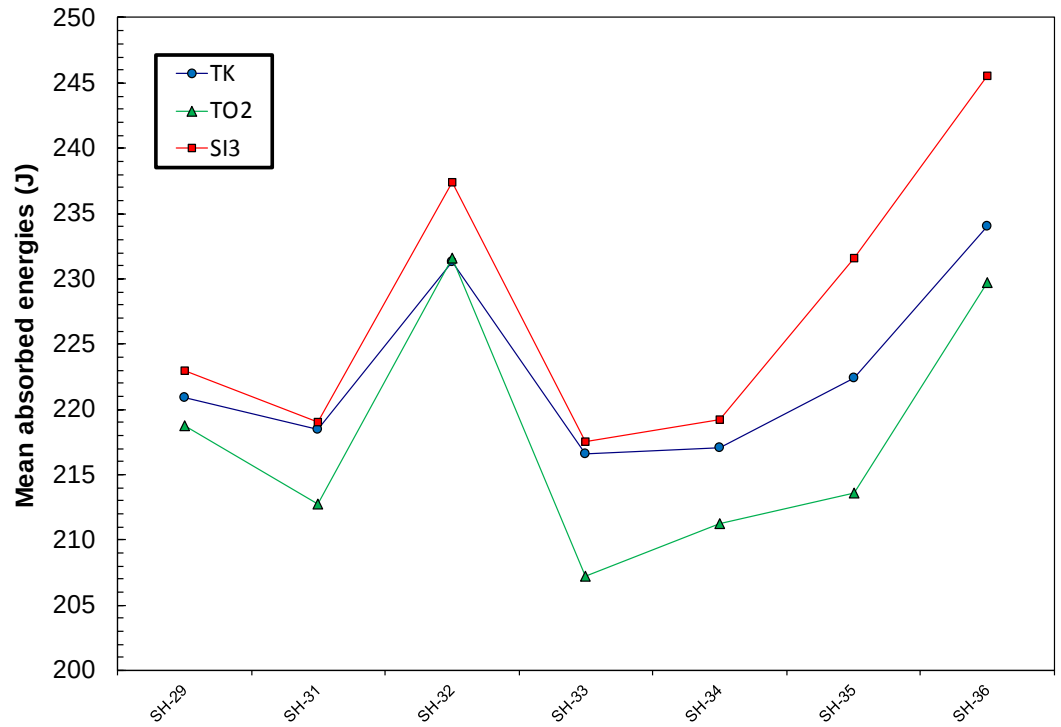

Figure 16 - Average absorbed energy values at the super-high-energy level for TK, TO2, and SI3.

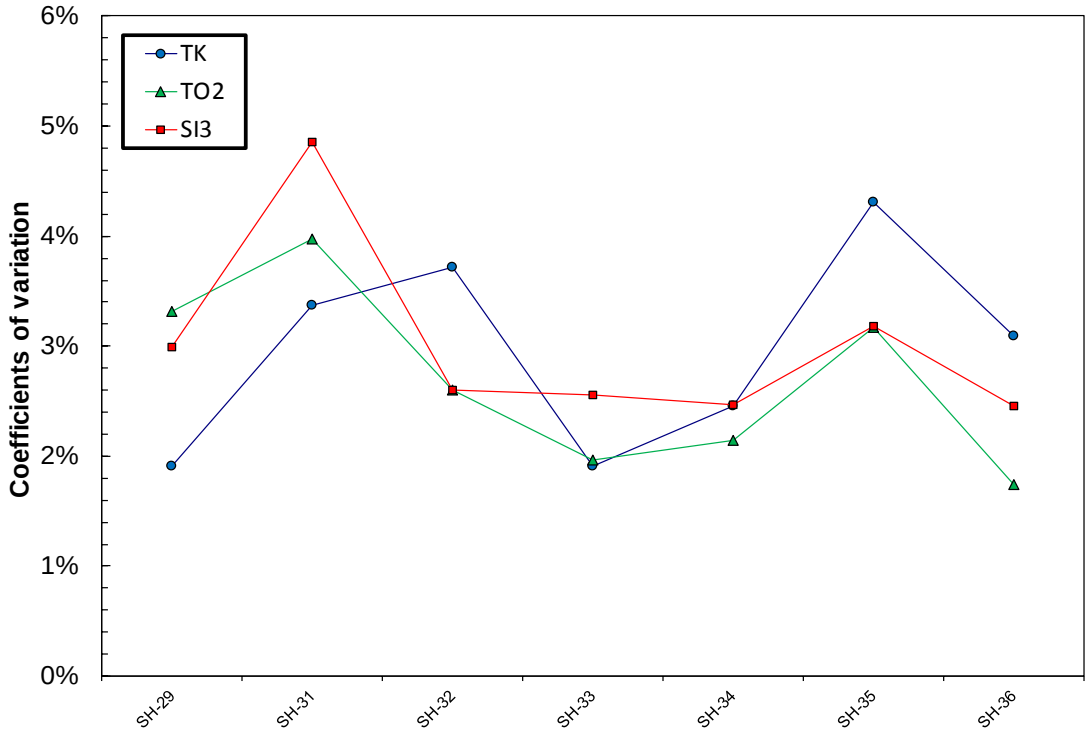

Figure 17 – Coefficients of variation at the super-high-energy level for TK, TO2, and SI3.

As shown by the ANOVA analyses conducted, the NIST master machines are not statistically different at the super-high-energy level.

4.4 Summary of findings

The main findings that emerged from our analysis of NIST Charpy historical data can be summarized as follows.

1. There were no statistically significant differences (at a 95 % confidence level) between the NIST master machines at the high- and super-high-energy levels.
2. At the low-energy level, a statistically significant difference was detected between the TK machine (C-type pendulum) and the other two machines (TO2 and SI3, both U-type pendulums), with TK yielding lower absorbed energy values.
3. Previously conducted measurements of machine compliance [8], coupled with an earlier investigation into the effect of machine design on Charpy test results for brittle materials [9], suggest that the lower absorbed energy yielded by TK is due to its higher stiffness, as compared with the other master machines.

5. Qualification of NIST Internal Reference Materials to assess the performance of Charpy machines and instrumented strikers

As part of the NIST Quality Assurance Program, we selected three reference materials (one per energy level) and fully characterize them for use as internal reference standards. They were qualified at room temperature (21 °C ± 1 °C) by performing instrumented impact tests on all the Charpy machines at NIST Boulder (except SI2).

These NIST Internal Reference Materials (IRMs) can be used to verify the performance of any of our machines in case of repair, suspected damage, major interventions, etc. In particular, the maximum force at room temperature was accurately characterized, so that new or refurbished/re-gaged instrumented strikers can be qualified.

5.1 Selection of IRMs

The most important criterion for selecting IRMs was the availability of a sufficient number of specimens to last for several years. It was also desirable to use the same materials that NIST distributes as verification specimens at the different energy levels (low-energy and high-energy level – 4340 steel with different heat treatments; super-high-energy level – T200 maraging steel). The material quality (homogeneity) of the verification lots selected was of secondary importance.

The main criterion imposed by NIST to assess the homogeneity of a verification lot is the lot sample size.

The sample size n of a verification lot is the minimum number of specimens from a given production lot that should be tested in a verification test [10], and is defined as:

$$n = \left(\frac{3 s_p}{E} \right)^2 \tag{2}$$

where: s_p is the pooled standard deviation of the lot (defined below) observed in qualification testing, and E is the greater of 1.4 J or 5 % of the mean energy. The pooled standard deviation is given by:

$$s_p = \sqrt{\frac{\sum_{1}^{P} {}_i s_i^2}{P}} \tag{3}$$

with: s_i = standard deviation of the test results obtained from the i-th machine, and
P = number of machines used for the qualification of the lot (P = 3 if only the master machines are used; P = 5 if MPM and TO3 are also used).

According to the requirements set by the Charpy program [10], a verification lot is acceptable if both the pilot and production lots yield a sample size $n \leq 5$. This allows NIST to sell sets of verification specimens that consist of 5 samples.

In our case, however, even if the sample size for a specific IRM is greater than 5, we can test a set of more than 5 specimens for the qualification of a specific machine and/or instrumented striker.

5.1.1 Low-energy IRM: LL-137

LL-137 was a limited-size qualification lot provided in 2015 by one of our suppliers in order to demonstrate the quality of the production. In spite of the limited number of samples available (275 after preliminary testing to verify the quality of the lot), the excellent homogeneity of impact properties which emerged from our characterization tests (see 5.2.1) allow the use of smaller verification sets (3 specimens), and therefore this lot should last for many years to come.

5.1.2 High-energy IRM: HH-107

HH-107 was a high-energy verification lot received at NIST in 2007. Although the test results of the pilot lot were satisfactory ($n < 5$), the production lot results were found not acceptable ($n > 5$). Although two more production lots were sent to NIST by the supplier, their results remained unacceptable and therefore this lot was not approved for customer distribution and kept in storage.

5.1.3 Super-high-energy IRM: SH-38

SH-38 is one of the last super-high-energy lots received by NIST in 2008, before super-high-energy level went out-of-stock. Despite testing one pilot lot and two production lots, the unsatisfactory test results ($n < 5$) forced NIST to reject SH-38 for customer distribution and keep it in storage.

5.2 Characterization of the selected IRMs

Two parameters were selected for characterizing the selected IRMs: absorbed energy, as measured by the machine encoder (*KV*, in J) and maximum force from the instrumented force/displacement record (F_m, in kN). Based on the results obtained from the master machines for *KV* and from all the NIST machines[4] for F_m, reference values and associated expanded uncertainties were calculated by means of the procedure outlined below [11].

The reference value for absorbed energy ($\hat{KV}$) or maximum force ($\hat{F}_m$) is simply calculated as the unweighted average of the machines used for certification:

$$\hat{X} = \frac{\sum_{i=1}^{P} \overline{X_i}}{P} \quad , \tag{4}$$

where X is the relevant test parameter, $\overline{X_i}$ is the average for the specific machine i, and P is the number of machines considered ($P = 3$ for *KV*[5] and $P = 4$ for F_m[6]).

[4] Except SI3, which did not have an instrumented striker at the time of testing.
[5] The master machines are 3.
[6] The machines equipped with instrumented strikers are 4.

The combined standard uncertainty (u_c) of the reference value is obtained by combining the within-machine standard uncertainty (u_w), the standard uncertainty due to machine bias (u_b), and the standard uncertainty of specimen homogeneity (u_h):

$$u_c = \sqrt{u_w^2 + u_b^2 + u_h^2} \quad , \tag{5}$$

with a number of effective degrees of freedom (ν_{eff}) given by the Welch-Satterthwaite formula [11].

The within-machine standard uncertainty is based on the "pooled" standard deviation (S_p):

$$S_p = \sqrt{\frac{\sum_{i=1}^{P} (n_i - 1)\sigma_i^2}{\sum_{i=1}^{P} n_i - P}} \quad , \tag{6}$$

with n_i = number of specimens tested on machine i and σ_i = standard deviation of test results from machine i. The within-machine standard uncertainty (u_w) is given by:

$$u_w = \frac{S_p}{\sqrt{N}} \quad , \tag{7}$$

where N is the total number of specimens tested.

The standard uncertainty due to bias between the machines (u_b) is obtained as:

$$u_b = \frac{\left| \overline{X}_{\max} - \overline{X}_{\min} \right|}{2\sqrt{3}} \quad , \tag{8}$$

where $\overline{X}_{\max}$ and $\overline{X}_{\min}$ are the largest and the smallest average values from individual machines, respectively.

Finally, the standard uncertainty due to specimen inhomogeneity (u_h) can be thought of as a correction for specimen inhomogeneity and is based on test results for 25 verification specimens broken on a single machine and the results for 15 production lot specimens tested on the same master machine. The mathematical derivation and the analytical expression of this uncertainty component are rather complex, and for the sake of conciseness the reader is referred to specific NIST publications such as [10,11].

Once the combined standard uncertainty is calculated, the expanded uncertainty (U), which corresponds to a 95 % confidence interval on the true reference value, is obtained by multiplying u_c by the t-value corresponding to a cumulative probability of 95 % and an effective number of degrees of freedom equal to ν_{eff}.

Additional statistics that are provided by the standard NIST analysis of test results from one or more Charpy machines are:

- Standard error SE, calculated as the ratio between the standard deviation (σ) and the square root of the number of tests (N).
- Range, calculated as the difference between the largest value and the smallest value from the tests performed.

- Coefficient of variation *CV* (%), calculated as the ratio between the standard deviation and the average value.
- Sample size *n*, calculated by means of eq. (2).

5.2.1 Characterization of LL-137

Twenty low-energy specimens from lot LL-137 were tested at room temperature on each of the three master machines (SI3, TO2, TK). Additional specimens were tested on the other two machines (15 samples on MPM and 13 samples on TO3). The results obtained are summarized in Table 14 (absorbed energy) and Table 15 (maximum force).

The sample sizes calculated from the test results are provided in Table 16.

Table 14 - Absorbed energy test results for the low-energy lot LL-137.

Machine	*N*	$\overline{KV}$ (J)	σ (J)	*SE* (J)	Range (J)	*CV* (%)
SI3	20	17.69	0.275	0.062	1.00	1.6
TO2	20	18.45	0.364	0.081	1.56	2.0
TK	20	16.53	0.224	0.050	0.80	1.4
MPM	15	18.19	0.351	0.091	1.17	1.9
TO3	13	18.36	0.358	0.099	1.26	2.0
Master	**60**	**17.55**	**0.847**	**0.109**	**3.12**	**4.8**
All	**88**	**17.78**	**0.799**	**0.085**	**3.12**	**4.5**

Table 15 – Maximum force test results for the low-energy lot LL-137.

Machine	*N*	$\overline{F_m}$ (kN)	σ (kN)	*SE* (kN)	Range (kN)	*CV* (%)
TO2	13	32.47	0.475	0.132	1.82	1.5
TK	18	32.17	0.497	0.117	2.06	1.5
MPM	15	31.44	0.277	0.072	1.02	0.9
TO3	12	31.52	1.126	0.325	3.91	3.6
All	**58**	**31.91**	**0.754**	**0.099**	**4.25**	**2.4**

Table 16 - Sample sizes obtained for the low-energy lot LL-137. Sample sizes smaller than 5 (NIST acceptance requirement) are in green, sample sizes larger than 5 are in red.

	Parameter	
Machine	*KV*	F_m
SI3	0.35	
TO2	0.61	1.03
TK	0.23	0.46
MPM	0.57	0.35
TO3	0.59	5.83
Master	0.40	
All	0.45	1.87

The results obtained demonstrate the excellent quality of lot LL-137. All sample sizes are smaller than 2 with the exception of the maximum force on TO3, and most of them are less than 1. The recommended size for an internal verification set is therefore 3 specimens[7].

5.2.2 Characterization of HH-107

Sixty-six high-energy specimens from lot HH-107 were tested at room temperature on the three master machines (25 on SI3 and TO2 each, 16 on TK). Additional specimens were tested on the other two machines (26 samples on MPM and 15 samples on TO3). The results obtained are summarized in Table 17 (absorbed energy) and Table 18 (maximum force).

The sample sizes calculated from the test results are provided in Table 19.

Table 17 - Absorbed energy test results for the high-energy lot HH-107.

Machine	*N*	$\overline{KV}$ (J)	σ (J)	*SE* (J)	Range (J)	*CV* (%)
SI3	25	108.19	5.222	1.044	16.95	4.8
TO2	25	109.03	5.490	1.098	22.91	5.0
TK	16	111.04	4.842	1.211	15.59	4.4
MPM	26	119.16	6.345	1.244	25.61	5.3
TO3	15	109.14	4.752	1.227	16.16	4.4
Master	**66**	**109.23**	**5.258**	**0.647**	**22.91**	**4.8**
All	**107**	**111.63**	**6.913**	**0.668**	**33.92**	**6.2**

Table 18 – Maximum force test results for the high-energy lot HH-107.

Machine	*N*	$\overline{F_m}$ (kN)	σ (kN)	*SE* (kN)	Range (kN)	*CV* (%)
TO2	25	26.84	0.311	0.062	1.12	1.2
TK	16	24.95	0.386	0.015	1.16	1.5
MPM	25	26.01	0.419	0.084	1.48	1.6
TO3	14	29.06	0.722	0.193	2.28	2.5
All	**80**	**26.59**	**1.396**	**0.156**	**5.84**	**5.2**

Table 19 - Sample sizes obtained for the high-energy lot HH-107. Sample sizes smaller than 5 (NIST acceptance requirement) are in green, sample sizes larger than 5 are in red.

	Parameter	
Machine	*KV*	F_m
SI3	8.39	
TO2	9.13	0.49
TK	6.85	0.86
MPM	10.21	0.94
TO3	6.83	2.22
Master	8.19	
All	8.66	1.03

[7] Even though the results documented in Table 16 would justify the use of a set of 2 specimens, the minimum "reasonable" number of samples to be tested for statistical significance is deemed to be 3.

The results obtained confirm that the HH-107 lot is not acceptable for distribution according to NIST standards (all sample sizes based on absorbed energy are larger than 5), however it can be used as one of NIST IRMs provided sets of 9 specimens are used for the internal verification of a Charpy machine.

Note that, even though the homogeneity of the lot based on *KV* is rather poor, all the sample sizes calculated from maximum force data are smaller than 3. Therefore, should the verification be restricted only to the instrumented striker, a set of 3 specimens would be adequate.

5.2.3 Characterization of SH-38

Seventy-two super-high-energy specimens from lot SH-38 were tested at room temperature on the three master machines (25 on SI3 and TO2 each, 22 on TK). Additional specimens were tested on the other two machines (15 samples on MPM and 17 samples on TO3). The results obtained are summarized in Table 20 (absorbed energy) and Table 21 (maximum force).

The sample sizes calculated from the test results are provided in Table 22.

Table 20 - Absorbed energy test results for the super-high-energy lot SH-38.

Machine	*N*	$\overline{KV}$ (J)	σ (J)	*SE* (J)	Range (J)	*CV* (%)
SI3	25	171.17	10.331	2.066	37.73	6.0
TO2	25	175.64	10.467	2.093	39.02	6.0
TK	22	175.52	7.272	1.550	27.57	4.1
MPM	15	178.78	8.376	2.163	24.00	4.7
TO3	17	166.99	6.376	1.547	20.68	3.8
Master	**72**	**174.05**	**9.655**	**1.138**	**40.29**	**5.5**
All	**104**	**173.58**	**9.559**	**0.937**	**43.99**	**5.5**

Table 21 – Maximum force test results for the super-high-energy lot SH-38.

Machine	*N*	$\overline{F_m}$ (kN)	σ (kN)	*SE* (kN)	Range (kN)	*CV* (%)
TO2	24	33.33	0.348	0.071	1.09	1.0
TK	22	30.06	0.264	0.056	0.90	0.9
MPM	15	31.58	0.123	0.032	0.43	0.4
TO3	15	33.98	0.137	0.035	0.43	0.4
All	**76**	**32.17**	**1.590**	**0.182**	**4.53**	**4.9**

Table 22 - Sample sizes obtained for the high-energy lot SH-38. Sample sizes smaller than 5 (NIST acceptance requirement) are in green, sample sizes larger than 5 are in red.

	Parameter	
Machine	*KV*	F_m
SI3	13.11	
TO2	12.79	0.39
TK	6.49	0.28
MPM	7.90	0.05
TO3	5.25	0.06
Master	10.85	
All	9.57	0.23

Based on the results obtained and the calculated sample sizes, the internal verification of a Charpy machine with the super-high-energy lot SH-38 requires a set of 11 specimens. On the other hand, the homogeneity of the lot in terms of maximum force is excellent. Therefore, if only the instrumented striker needs to be verified, the use of a set of 3 specimens is recommended.

5.2.4 Reference values and expanded uncertainties

The reference values and expanded uncertainties (*U*) calculated for the IRMs are listed in Table 23. As noted in the table, statistical calculations for absorbed energy are based on master machines only, while the parameters calculated for maximum force are based on test results from TK, TO2, MPM, and TO3 (indicated in the table as "all" machines). This choice was based on the fact that we wanted to be consistent with the NIST qualification process of the indirect verification lots, whereby the reference energy values are obtained from the master machines only, whereas there are no indications that any of our instrumented striker/machine combinations should be excluded from the establishment of reference maximum forces.

Table 23 - Overview of statistical parameters calculated for the NIST IRMs.

IRM	Parameter	Machines	Reference value	*U*	Minimum value	Maximum value	Sample size	Recomm. set size
LL-137	*KV* (J)	Master	17.55	0.076	16.15	18.95	0.40	3
	F_m (kN)	All	31.90	0.263	30.31	33.50	1.87	3
HH-107	*KV* (J)	Master	109.45	1.292	103.98	114.92	8.19	9
	F_m (kN)	All	26.71	0.163	25.37	28.05	1.03	3
SH-38	*KV* (J)	Master	174.11	2.212	165.40	182.82	10.85	11
	F_m (kN)	All	32.24	0.068	30.63	33.85	0.23	3

Table 23 also provides the ranges of acceptable values for the verification of Charpy machines and instrumented strikers, in terms of "minimum value" and "maximum value". These values are based on the following criteria:

- Absorbed energy: the criteria prescribed in ASTM E23 are used, *i.e.*, the acceptable range corresponds to the reference value ± the larger between 1.4 J or 5 %.
- Maximum force: the acceptable range is defined as the reference value ± 5 %.

Furthermore, sample sizes and recommended set sizes (number of specimens to be tested), already presented in previous subsections, are also provided. Note that, if only the verification of an instrumented striker based on F_m is required, testing of 3 specimens per energy level is sufficient. As already mentioned, 3 is considered the minimum size of a set for the results to be statistically meaningful.

6. Example of use of Internal Reference Materials: TO3 (machine and instrumented striker)

At the time we performed the testing for the characterization of the IRMs, the acquisition of the force signal from most instrumented strikers was based on a signal-based trigger system. Namely, acquisition was triggered by the striker signal rising above a threshold value, which was normally set at 0.5 V. This type of trigger was often responsible for the missed acquisition of an instrumented impact test due to the presence of electric noise on the network[8].

Since then, the signal-based trigger system was replaced by an optical trigger system for all NIST Charpy machines. Signal acquisition is triggered by a copper tab attached to the swinging hammer passing in front of a photocell attached to the machine frame. By accurately positioning the tab, acquisition can be started just before the specimen is actually impacted. This new triggering system increased our rate of successful acquisitions to practically 100 % of the performed tests.

After the installation of the optical trigger system on the TO3 machine, we simulated the verification of the machine and its instrumented striker (TO-JS1) by testing specimens of the three NIST IRMs (LL-137, HH-107, SH-38) and evaluating the results obtained (absorbed energies and maximum forces) with respect to the criteria outlined in 5.2.4. The number of specimens tested for each IRM (3 for LL-137, 9 for HH-107, and 11 for SH-38) was also in accordance with the recommendations of Table 23.

All the results obtained (mean values of *KV* and F_m), shown in Table 24, fall within the acceptable ranges of Table 23. The verification of this machine and instrumented striker would therefore be considered successful.

Table 24 - Test results obtained for the TO3 machine in the simulated verification with IRMs.

Lot	Parameter	Number of tests	Mean value	PASS/FAIL
LL-137	*KV* (J)	3	17.94	PASS
	F_m (kN)		31.86	PASS
HH-107	*KV* (J)	9	109.49	PASS
	F_m (kN)		27.17	PASS
SH-38	*KV* (J)	11	168.62	PASS
	F_m (kN)		33.70	PASS

[8]This explains why the number of F_m values in a specific test series is often lower than the number of *KV* values recorded (see sections 5.2.1, 5.2.2, and 5.2.3).

7. Statistical analyses of IRM test results: a comparative study of NIST Charpy machines

The results of the impact tests (in most cases instrumented) performed for the characterization of the IRM lots (LL-137, HH-107, SH-38) were statistically analyzed as a mean to compare the NIST Charpy machines, in terms of both absorbed energy and maximum force. This new statistical analysis adds to the knowledge base gained from the analysis of historical data presented in Section 4, and provides a baseline for using these lots to evaluate machine performance in the future.

Statistical analyses among all five NIST machines or among the three master machines were performed by means of the one-way analysis of variance (ANOVA, single-factor), which tests the null hypothesis that samples in two or more groups are drawn from populations with the same mean values. It is typically used to compare means of three or more samples through the *F* distributions. If the calculated *F* value is larger than the critical *F* value (F_{crit}), the null hypothesis is rejected and the means are considered statistically different at the selected significance level (95 %).

The magnitude of the difference can be qualitatively assessed by comparing the numerical values of *F* and F_{crit}: if *F* and F_{crit} belong to different decades (*e.g.*, $F = 70$ and $F_{crit} = 3$), the difference can be considered to be significant (or very significant if *F* and F_{crit} differ by more than one decade). If *F* and F_{crit} belong to the same decade (*e.g.*, $F = 7$ and $F_{crit} = 3$), the means are deemed just slightly different.

When comparing only two machines, Student's two-sample *t*-test assuming equal variances was used. This is a common statistical test which tests the null hypothesis that two sets of data are significantly different from each other. This test, which was introduced in 1908 by William Sealy Gosset (whose pen name was "Student") [12], assumes that variables follow a normal distribution. If the calculated *t*-value is larger than the critical *t* value (t_{crit}), the null hypothesis is rejected and the means are considered statistically different. The same qualitative assumptions as for ANOVA (see above) about the significance of the difference, based on the orders of magnitude of *t* and t_{crit}, were also used for the *t*-test.

For both ANOVA and *t*-test, an alternative way to interpret the results is to compare the calculated *P*-value to the significance level of the analysis, $\alpha = 0.05$. The *P*-value is defined as the probability of obtaining a result equal or "more extreme" that what was actually observed, when the null hypothesis is true. Qualitatively, one can say that the smaller is *P*, the more significant is the statistical difference between the two data sets.

7.1 Comparisons among multiple machines: ANOVA analyses

7.1.1 Absorbed energy

The results of the ANOVA analyses for the three master machines (SI3, TO2, TK) are summarized in Table 25.

The results of the statistical analyses on the historical data for the master machines (see 4.1, 4.2, and 4.3) are confirmed: the three NIST master machines can be considered substantially

equivalent, with the exception of the TK machine, which yields lower *KV* values at the low-energy level.

Table 25 - Results of ANOVA analyses for absorbed energy (master machines).

IRM	Machine	Mean (J)	*F*	F_{crit}	*P*	Test result
LL-137	SI3	17.69	217.33	3.16	< 0.001	Machines are very significantly different
	TO2	18.44				
	TK	16.53				
HH-107	SI3	108.19	1.46	3.14	0.2395	Machines are not significantly different
	TO2	109.12				
	TK	111.04				
SH-38	SI3	171.17	1.74	3.13	0.1833	Machines are not significantly different
	TO2	175.64				
	TK	175.52				

Table 26 presents the results of the ANOVA analyses extended to all machines (SI3, TO2, TK, TO3, MPM).

This time, the machines are found to be statistically different at all energy levels, although the difference is most significant at the low-energy level (probably due to the consolidated bias of the TK machine). At the super-high-energy level, the difference is not significant.

Table 26 - Results of ANOVA analyses for absorbed energy (all machines).

IRM	Machine	Mean (J)	*F*	F_{crit}	*P*	Test result
LL-137	SI3	17.69	121.32	2.48	< 0.001	Machines are very significantly different
	TO2	18.44				
	TK	16.53				
	TO3	18.34				
	MPM	18.19				
HH-107	SI3	108.19	17.00	2.46	< 0.001	Machines are significantly different
	TO2	109.12				
	TK	111.04				
	TO3	109.14				
	MPM	119.16				
SH-38	SI3	171.17	4.61	2.46	0.001869	Machines are different
	TO2	175.64				
	TK	175.52				
	TO3	166.99				
	MPM	178.78				

7.1.2 Maximum force

The results of the ANOVA analyses for the four NIST machines equipped with instrumented strikers (TO2, TK, TO3, MPM) are summarized in Table 27.

The four machines were found to be statistically different, with very significant statistical differences at all energy levels.

Generally speaking, the observed differences for maximum forces tend to be more significant than for absorbed energies. This may be well due to the much better homogeneity of the lots in terms of F_m than KV (see Table 23). In most cases, the recommended number of samples in a set is 3, which is much higher than the sample size calculated from test results, unlike in the case of absorbed energy. One can therefore expect between-machine differences to show up more clearly in terms of maximum force than absorbed energy, as KV differences will be somewhat obscured by sample-to-sample inhomogeneity.

Table 27 - Results of ANOVA analyses for maximum force.

IRM	Machine	Mean (kN)	F	F_{crit}	P	Test result
LL-137	TO2	32.47	117.91	2.78	< 0.001	Machines are very significantly different
	TK	32.18				
	TO3	31.52				
	MPM	31.44				
HH-107	TO2	26.84	470.68	2.72	< 0.001	Machines are very significantly different
	TK	23.36				
	TO3	29.05				
	MPM	26.01				
SH-38	TO2	33.33	2449.44	2.73	< 0.001	Machines are very significantly different
	TK	30.06				
	TO3	33.98				
	MPM	31.58				

7.2 One-on-one machine comparisons: *t*-tests

Student's *t*-tests were run on every two machine combination for both absorbed energy (KV) and maximum force (F_m). The results are pictorially represented in Figure 18:

- "NO" on a green background means the two machines are not statistically different;
- "YES" on a pink background means they are different but not significantly (*i.e.*, t and t_{crit} belong to the same decade);
- "YES" on a red background means they are significantly different (*i.e.*, t and t_{crit} differ by only one decade);
- "**YES**" (in bold) on a red background means they are very significantly different (*i.e.*, t and t_{crit} differ by more than one decade).

7.2.1 Absorbed energy

The results concerning absorbed energy values are on the left side of Figure 18.

At the low-energy level, most comparisons yielded statistical differences except for:

- TO2 vs. TO3: not surprisingly, since the two machines are quite similar – the main difference being a higher capacity for TO3, due to a higher falling angle (Table 1);
- TO3 vs. MPM.

The TK machine was found very significantly different from all the other machines, as a confirmation of the bias already observed in the analysis of historical data presented in Section 4.

At the high-energy level, only the high-capacity MPM machine was found significantly different from all the other machines. It shows a tendency to yield higher absorbed energies (Table 26). The remaining four machines (and in particular, the three master machines) do not exhibit statistically different behaviors.

At the super-high-energy level, the overall situation is less clearly defined: however, the only statistically significant differences were found when the non-master machines (TO3 and MPM) were involved in the comparisons. It's interesting to note that, based on SH-38 results, TO2 and TO3 are significantly different.

Very significant statistical differences were only obtained at the low-energy level, as a consequence of the excellent homogeneity of LL-137. As previously mentioned, the better is the material homogeneity, the more likely it is to observe statistical differences between the machines with a given sample size.

7.2.2 Maximum force

The results concerning maximum forces are on the right side of Figure 18.

At the low-energy level, most comparisons yielded statistical differences, although most of them were not significant (pink cells in Figure 18). The only equivalences were found between TO2 and TK and between TO3 and MPM.

At the high-energy level, all the machines were found statistically different from each other. In most cases, the differences were very significant.

At the super-high-energy level, only TO2 and MPM were found to be statistically not different. Most of the other differences were found to be very significant.

It's interesting to remark that the TO2 and TO3 machines have been found to be statistically different at all energy levels, despite the very similar design and the fact that they share the same instrumented striker (TO-JS1). The reason of such difference has to be attributed to the calibration of the instrumented striker.

LL-137 - *KV*	SI3	TO2	TK	TO3	MPM
SI3		YES	**YES**	YES	YES
TO2	YES		**YES**	NO	YES
TK	**YES**	**YES**		**YES**	**YES**
TO3	YES	NO	**YES**		NO
MPM	YES	YES	**YES**	NO	

LL-137 - F_m	SI3	TO2	TK	TO3	MPM
SI3					
TO2			NO	YES	YES
TK		NO		YES	YES
TO3		YES	YES		NO
MPM		YES	YES	NO	

HH-107 - *KV*	SI3	TO2	TK	TO3	MPM
SI3		NO	NO	NO	YES
TO2	NO		NO	NO	YES
TK	NO	NO		NO	YES
TO3	NO	NO	NO		YES
MPM	YES	YES	YES	YES	

HH-107 - F_m	SI3	TO2	TK	TO3	MPM
SI3					
TO2			**YES**	**YES**	YES
TK		**YES**		**YES**	YES
TO3		**YES**	**YES**		**YES**
MPM		YES	YES	**YES**	

SH-38 - *KV*	SI3	TO2	TK	TO3	MPM
SI3		NO	NO	NO	YES
TO2	NO		NO	YES	NO
TK	NO	NO		YES	NO
TO3	NO	YES	YES		YES
MPM	YES	NO	NO	YES	

SH-38 - F_m	SI3	TO2	TK	TO3	MPM
SI3					
TO2			**YES**	YES	NO
TK		**YES**		**YES**	**YES**
TO3		YES	**YES**		**YES**
MPM		NO	**YES**	**YES**	

Figure 18 – Results of the machine-to-machine comparisons by means of t-tests.

8. Discussion: calibration of an instrumented striker

A thoroughly reliable force calibration of instrumented Charpy strikers is a longstanding and basically unresolved problem [13-15].

The main issue is the different behavior of the striker when force is applied statically (during static calibration) and dynamically (during the impact test). The difference between the static and dynamic response of the striker generally depends on the design of the striker (type and location of the strain-gages, as well as their distance from the striking edge) [16].

Instrumented Charpy strikers are typically calibrated statically (*i.e.*, applying a static force and recording the output of the strain-gages). Although both ASTM E2298 and ISO 14556 recommend an *in-situ* calibration (with the striker attached to the pendulum assembly, Figure 19), it is much more common for the calibration to be performed *ex-situ* (with the striker removed from the pendulum and installed on a universal testing machine, Figure 20).

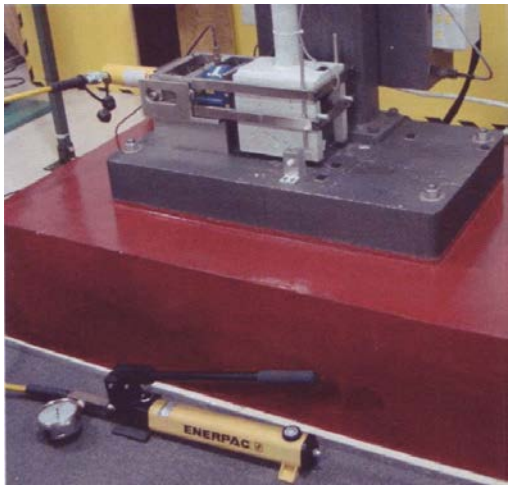

Figure 19 - In-situ calibration of an instrumented Charpy striker.

Figure 20 - Ex-situ calibration of an instrumented Charpy striker.

The comparison between the absorbed energy measured by the machine encoder (*KV*) and calculated by integrating the area under the instrumented force/displacement test record (W_t) is considered to be a good indicator of the reliability of the striker calibration [15].

The two measures of absorbed energy should be similar, but not identical, for a Charpy test. Specifically, some energy components that affect the energy measured by the encoder are unrelated to specimen fracture and should therefore not be included in W_t [17]:

(a) Residual vibrational energy of the pendulum hammer after specimen fracture.

(b) Energy associated with post-fracture impacts.

Manahan *et al.* [8] quantified these energy losses through numerical simulations and impact experiments conducted at NIST on verification specimens of low, high, and super-high energy. For a U-shaped hammer with 400 J capacity (the TO3 machine), the vibrational energy losses were estimated to be on the order of 1 % – 2 % of *KV*. Post-fracture interactions between low-energy specimen halves exiting the front of the machine and the swinging hammer were shown to add an average of 3 J to 4.5 J to the encoder energy. Unfortunately, such post-fracture interactions are of a random nature, and their occurrence is not the same for all tests.

Additional effects which were identified in the study were: force distribution in the striker, machine and striker inertia, and windage/friction correction of the encoder energy.

Vibrational phenomena, as well as some of the other factors, are generally sensitive to the design and the stiffness of the impact machine and the striker.

From the discussion reported above, it results that theoretically the instrumented energy (W_t) should be lower than the encoder energy, and the magnitude of the difference would depend on the machine design and the energy level of the material being tested.

The two major instrumented impact test standards address the difference between *KV* and W_t:

- ASTM E2298, Section 7.2.6, *Requirements on Absorbed Energy*: if the difference between *KV* and W_t is outside the larger of ±15 % of *KV* or ±1 J but less than the larger of ±25 % of *KV* or ±2 J, force values must be adjusted until $KV = W_t$. If the difference exceeds the latter limits, the calibration of the instrumented striker must be questioned.
- ISO 14556, Section 6.1, *Testing machine*: if *KV* and W_t differ by more than ±5 J, the user should investigate the friction of the machine, the calibration of the measuring system, and the software used.

The procedure adopted by ASTM E2298 is sometimes called “Dynamic Force Adjustment”, and was shown to be beneficial in reducing the scatter of instrumented Charpy results [15].

An investigation performed at NIST in 2009 [14] showed that the response of an instrumented striker depends on the loading rate, and that the conversion factor (kN/V) of a striker tends to decrease as the loading rate increases. However, the sensitivity to loading rate can be very different depending on the striker design: the decrease of the conversion factor was observed to be more significant for a striker where the strain-gages were located laterally (“left-right” design) in comparison to a striker where the strain-gages were located above and below the striking edge (“top-bottom” design).

An additional option that was investigated in a study conducted at NIST in 2015 [9] is the obtainment of the calibration factor for an instrumented impact striker through the measurement of the compliance of the Charpy machine.

The machine compliance (C_M) is one of the two components that add up to yield the elastic slope of an instrumented Charpy force/displacement curve. The other component is the specimen compliance (C_S), which can be calculated analytically from the material's Young's modulus E, the distance between machine anvils S (span), and the specimen dimensions (thickness B, width W, notch depth a). For a steel specimen with nominal dimensions ($B = W = 10$ mm, $a = 1$ mm) tested at room temperature ($E = 207{,}000$ MPa) on a machine with $S = 40$ mm, the compliance of the specimen is $C_S = 0.0114$ mm/kN.

The machine compliance can be measured according to Ireland [18] through different approaches, all based on performing a low-blow impact test on an unnotched specimen. A low-blow test is an impact test wherein the pendulum is dropped from a small angle, so that no plastic deformation is produced on the specimen. The analysis of the instrumented force/time curve of the low-blow test (Figure 21) can be performed in three different ways, all leading to an estimate of C_S. The average of these three determinations is then taken as the compliance of the Charpy machine. The compliance values obtained in [9] for the NIST Charpy machines are summarized in Table 28. Note that TK and MPM were found to be significantly stiffer than TO2 and TO3.

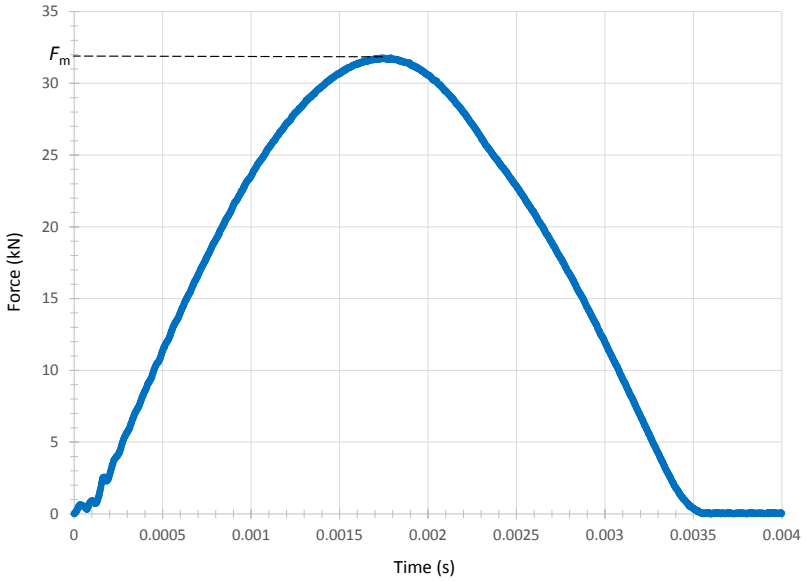

Figure 21 - Instrumented force/time record for a low-blow test on an unnotched specimen.

Table 28 - Compliance values measured in 2015 for the NIST Charpy machines.

Impact machine	Instrumented striker	C_M (mm/kN)
TO2	TO-JS1	0.01252
TO3	TO-JS1	0.01224
MPM	MPM-JS2	0.00980
T-K	S-79-R	0.00927

One of the methods used for determining machine compliance [19] is based on the assumption that the interaction between hammer and specimen can be represented by a vibrating mass on a spring, and therefore the force/time record of the low-blow test becomes a half-oscillation of the system. The method (which Ireland called "dynamic calibration of an instrumented tup" in [18]) consists in matching a calculated peak impulse force, from an elastic low-blow test, with that obtained from the instrumented signal.

If the impact is entirely elastic, the maximum force F_m is calculated from the following relationship, which is valid for elastic absorption:

$$F_m = \sqrt{\frac{2W_0}{C_M + C_S}} \quad , \tag{9}$$

where W_0 is the initial potential energy, corrected for air resistance and friction losses. W_0 is obviously lower than the machine capacity because the hammer is dropped from a lower height. By substituting the corresponding values of W_0, C_M, and C_S, the dynamic conversion factors shown in Table 29 were obtained.

Table 29 – Dynamic conversion factors determined in 2015 for the NIST Charpy machines.

Impact machine	Instrumented striker	Dyn. conv. factor (kN/V)	Static conv. factor (kN/V)	Ratio dyn/stat
TO2	TO-JS1	9.9336	10.188	0.975
TO3	TO-JS1	9.7021	10.188	0.952
MPM	MPM-JS2	8.1622	8.2166	0.993
T-K	S-79	10.8708	11.4110	0.953

The calculated dynamic conversion factors are between 1 % and 5 % lower than the corresponding static conversion factors obtained from the static calibrations. This confirms the outcome of the 2008 investigation [14], which showed that conversion factors tend to decrease as the loading rate increases. Note also that the loading rate in a low-blow test (impact velocity $v \approx$ 0.5-1 m/s) is lower than in a conventional Charpy test ($v \approx$ 5-5.5 m/s), so a further decrease can be anticipated if the loading rate corresponds to that of an actual Charpy test.

The conversion factors in Table 29 were obtained with a specific configuration of the acquisition system (amplifier, signal conditioner, A/D converter). If the configuration changes, the effective amplification factor for the striker signal (gain) will be different, and a new conversion factor has to be obtained. The instrumented tests documented in this report were performed with a different acquisition system than in [9], and therefore the specific calibration factors (both static and dynamic) are different. Nevertheless, it is reasonable to assume that the ratio between dynamic and static (shown in the last column of Table 29) would remain the same if a different acquisition system was used.

In the following sections, we re-analyzed the instrumented tests performed on the IRMs on the different Charpy machines and compared the values of maximum force after:

(a) applying the "dynamic force adjustment" (*i.e.*, correcting the forces until $W_t = KV$ or $W_t = KV_{mod}$, where KV_{mod} is KV minus vibrational and post-fracture components), and

(b) correcting the forces for the effect of increased loading rate by means of the ratio in the last column of Table 29.

8.1 Comparison between *KV* and W_t

As explained above, the comparison between the two measures of absorbed energy (*KV* and W_t) provides a qualitative assessment of the instrumented striker calibration. *KV* should be larger than W_t, since it includes vibrational components, post-fracture interactions, and possibly other extraneous components. The difference should not exceed a few percent of *KV*.

Table 30 provides an overview of the comparison between *KV* and W_t for all the instrumented tests performed on the NIST IRMs. For every combination machine/striker/material, the following information is given: average values and coefficients of variation for *KV* and W_t; average values of the difference between *KV* and W_t; average values and coefficients of variation for the ratio KV/W_t.

Table 30 - Comparison between encoder and instrumented energies for the tests on IRMs.

Charpy machine	Instr. striker	IRM	KV_{mean} (J)	CV_{KV}	$W_{t,mean}$ (J)	CV_{Wt}	$(KV\text{-}W_t)_{mean}$ (J)	$(KV/W_t)_{mean}$	$CV_{KV/Wt}$
TO2	TO-JS1	LL-137	18.45	2.0%	19.33	2.0%	-0.80	0.959	0.9%
		HH-107	109.12	5.0%	110.19	4.2%	-1.07	0.990	2.9%
		SH-38	175.64	6.0%	176.61	5.0%	-1.52	0.991	1.4%
TO3		LL-137	18.34	1.9%	17.59	3.4%	0.79	1.046	3.4%
		HH-107	109.14	4.4%	105.41	3.7%	3.37	1.032	2.2%
		SH-38	166.99	3.8%	161.54	3.3%	5.29	1.033	1.7%
TK	S-79-R	LL-137	16.53	1.4%	12.09	1.9%	4.44	1.368	1.3%
		HH-107	111.04	4.4%	102.58	4.4%	8.46	1.083	1.4%
		SH-38	175.52	4.1%	165.03	3.8%	10.49	1.063	0.8%
MPM	MPM-JS2	LL-137	18.19	1.9%	16.35	2.0%	1.84	1.113	1.1%
		HH-107	119.16	5.3%	121.64	4.6%	-2.00	0.983	1.0%
		SH-38	178.78	4.7%	182.87	4.6%	-4.09	0.978	0.7%

If the information presented in Table 30 is compared with the requirements in 7.2.6 of ASTM E2298, all the combinations fulfill the requirements and would therefore need no adjustment of the measured forces, except for the low-energy tests performed on the TK machine. In this case, the mean difference between *KV* and W_t amounts to 37 % or 4.44 J, and would cause these results to be deemed unacceptable and the striker calibration to be repeated. Even at the high- and super-high-energy levels, the differences between the two measures of absorbed energy are the highest of all the machine/striker combinations. Therefore, it is reasonable to maintain that the static calibration is not an adequate option for the instrumented striker of the TK machine[9]. We should mention here that the difference in absorbed energies for the TK machine at the low-energy level (Section 7) was found to be between 7 % and 11 %, and therefore it's not sufficient to justify such a large difference between this machine and the others.

The examination of the information presented in Table 30 yields the following additional observations.

1. In percent, the difference between *KV* and W_t tends to decrease with increasing absorbed energy. Conversely, the difference increases in absolute terms.

[9] Another plausible hypothesis is that the striker design is not adequate.

2. For five out of twelve machine/striker/material combinations in Table 30, W_t is smaller than *KV*, contrary to the physical evidence which identified additional energy components in *KV* which should not affect W_t [8,17]. Under such circumstances, the adequacy of the static calibration of the instrumented strikers can be questioned.
3. In most cases, the scatter of W_t values was found to be slightly lower than the scatter of *KV* values. This is likely related to the random nature of the additional energy components that affect *KV* but not W_t.

8.2 Effect of dynamic force adjustment

We applied the so-called "dynamic force adjustment" [15] to the instrumented test results obtained in this study, in order to assess whether the agreement between machine/striker combinations would improve, as the recommendation in 7.2.6 of ASTM E2298 implies.

The adjustment consisted in iteratively correcting instrumented forces, until equivalence between W_t and *KV* is achieved within ±0.1 %. This procedure is based on the assumption that the two measures of absorbed energy should be identical, which, as noted above, is not entirely accurate. However, in case of large discrepancy between *KV* and W_t and therefore significant doubts on the reliability of the striker calibration, it is considered an acceptable/reasonable approach from an engineering perspective.

The effect of the dynamic force adjustment (DFA) was assessed based on the maximum force values measured for every machine/striker/material combination. Average values and coefficients of variation for the original ($F_{m,or}$) and adjusted ($F_{m,DFA}$) maximum forces are shown in Table 31. A general increase of the scatter is noticeable, as a result of the additional variability introduced by the adjustment factors that are determined individually for each impact test.

Table 31 - Effect of dynamic force adjustment (DFA) on maximum forces.

Charpy machine	Instr. striker	IRM	$F_{m,or}$ (kN)	$CV_{Fm,or}$	$F_{m,DFA}$ (kN)	$CV_{Fm,DFA}$
TO2	TO-JS1	LL-137	32.47	1.5%	31.12	1.8%
		HH-107	26.84	1.2%	26.58	2.9%
		SH-38	33.33	1.0%	33.03	1.7%
TO3		LL-137	31.52	3.6%	33.01	3.4%
		HH-107	29.06	2.5%	29.98	2.7%
		SH-38	33.98	0.4%	30.87	0.8%
TK	S-79-R	LL-137	25.87	1.2%	35.39	1.6%
		HH-107	23.36	0.7%	25.29	1.5%
		SH-38	28.51	0.4%	30.32	0.9%
MPM	MPM-JS2	LL-137	31.44	0.9%	34.98	1.5%
		HH-107	26.01	1.6%	25.58	2.2%
		SH-38	31.58	0.4%	30.87	0.8%

In order to assess whether the dynamic force adjustment is beneficial or detrimental to the agreement between the NIST machines at the different energy levels, we compared the statistical *F* factors obtained from the ANOVA analyses performed on the $F_{m,or}$ (Table 27) and $F_{m,DFA}$ test results (Table 32). An increase in *F* factor after adjustment would indicate a detrimental effect of the adjustment procedure.

At all energy levels, application of DFA improves the agreement between machine/striker combinations, although they remain statistically different.

Table 32 - Effect of dynamic force adjustment on the agreement among NIST machines.

IRM	Machine	$F_{m,or}$ (kN)	F	$F_{m,DFA}$ (kN)	F	F_{crit}	Effect of dynamic force adjustment
LL-137	TO2	32.47	117.90	31.12	116.03	2.78	Negligible
	TO3	31.52		33.01			
	TK	26.21		35.39			
	MPM	31.44		34.98			
HH-107	TO2	26.84	470.68	26.58	170.31	2.72	Beneficial
	TO3	29.06		29.98			
	TK	23.36		25.29			
	MPM	26.01		25.58			
SH-38	TO2	33.33	2449.44	33.03	414.62	2.73	Very beneficial
	TO3	33.98		35.09			
	TK	28.51		30.32			
	MPM	31.58		30.87			

The analysis was repeated after excluding the TK machine and striker, for which the static calibration factor had been found completely inadequate.

Table 33 shows that, after the "outlier" machine/striker combination is excluded, the effect of applying DFA is less pronounced, and specifically at the low-energy level (where the TK striker had exhibited a particularly poor performance) the agreement is significantly worse.

Table 33 - Effect of dynamic force adjustment on the agreement among NIST machines, after excluding TK.

IRM	Machine	$F_{m,or}$ (kN)	F	$F_{m,DFA}$ (kN)	F	F_{crit}	Effect of dynamic force adjustment
LL-137	TO2	32.47	9.05	31.12	93.46	3.26	Detrimental
	TO3	31.52		33.01			
	MPM	31.44		34.98			
HH-107	TO2	26.84	193.40	26.58	185.25	3.15	Negligible
	TO3	29.06		29.98			
	MPM	26.01		25.58			
SH-38	TO2	33.33	370.25	33.03	265.54	3.18	Beneficial
	TO3	33.98		35.09			
	MPM	31.58		30.87			

Based on these results, we see significant benefit from the application of DFA only when the reliability of the calibration of an instrumented striker is clearly questionable (as in the case of TK, see Table 30).

8.3 Modified dynamic force adjustment

The already mentioned investigation conducted by Manahan et al. [8] on the nature of the differences between absorbed energies measured by the encoder (*KV*) and calculated from the area under the instrumented force-displacement record (W_t) identified two significant energy losses which affect *KV* but do not contribute to the fracture process:

(a) Vibrational energy of the pendulum hammer, which was quantified by means of finite element calculations as 1 % of *KV* (slightly higher for U-type hammers, such as TO2 and TO3).

(b) Only for low-energy specimens exiting the front of the machine, energy associated with post-fracture impacts between specimen halves and hammer/striker assembly, which amounts to 3-4.5 J (once again, the higher end of the range refers to U-type hammers).

We decided to apply once more the dynamic force adjustment procedure to our test results, this time imposing equivalence between W_t and a modified encoder energy KV_{mod}, where:

$$KV_{mod} = 0.99\,KV - 3.75\,\text{J} \tag{10}$$

for tests on LL-137 (3.75 J is the average between 3 J and 4.5 J, and all specimens exit the front of the machine), and

$$KV_{mod} = 0.99\,KV \tag{11}$$

for tests on HH-107 and SH-38. The results are presented in Table 34 (including the TK machine) and Table 35 (excluding the TK machine).

Table 34 - Effect of modified dynamic force adjustment (DFAmod, based on the equivalence between W_t and KV_{mod}) on the agreement among NIST machines.

IRM	Machine	$F_{m,or}$ (kN)	F	$F_{m,DFAmod}$ (kN)	F	F_{crit}	Effect of dynamic force adjustment
LL-137	TO2	32.47	117.90	24.51	68.98	2.78	Beneficial
	TO3	31.52		25.93			
	TK	26.21		27.01			
	MPM	31.44		27.42			
HH-107	TO2	26.84	470.68	26.31	170.31	2.72	Beneficial
	TO3	29.06		29.68			
	TK	23.36		25.04			
	MPM	26.01		25.32			
SH-38	TO2	33.33	2449.44	32.70	394.80	2.73	Very beneficial
	TO3	33.98		34.74			
	TK	28.51		30.02			
	MPM	31.58		30.56			

Table 35 - Effect of modified dynamic force adjustment (DFAmod, based on the equivalence between W_t and KV_{mod}) on the agreement among NIST machines (excluding TK).

IRM	Machine	$F_{m,or}$ (kN)	F	$F_{m,DFAmod}$ (kN)	F	F_{crit}	Effect of dynamic force adjustment
LL-137	TO2	32.47	9.05	24.51	69.52	3.26	Detrimental
	TO3	31.52		25.93			
	MPM	31.44		27.42			
HH-107	TO2	26.84	185.25	26.31	185.25	3.15	Negligible
	TO3	29.06		29.68			
	MPM	26.01		25.32			
SH-38	TO2	33.33	265.54	32.70	251.45	3.18	Slightly beneficial
	TO3	33.98		34.74			
	MPM	31.58		30.56			

When compared with Table 32, the information in Table 34 shows an improvement in the within-machine agreement (decrease of F statistic) when an "effective" encoder energy is used, which filters out the fracture-unrelated components in KV. This improvement, however, is substantial only for the LL-137 tests. At the high-energy level, the calculated values of F in Tables 32 and 34 are identical.

After excluding the TK machine, the comparison between Table 33 and Table 35 reiterates the observation that DFA is beneficial only in the presence of a questionable striker calibration.

8.4 Force correction based on machine compliance

As detailed in section 8 (Figure 21 and Table 28), the conversion factor between applied force and striker output can be determined through the measurement of machine compliance. When measuring the compliance of NIST Charpy machines [9], a different acquisition system was used than that used for the current investigation.

The conversion factors calculated in [9] for low-blow impact tests (0.5-1 m/s) were systematically lower than those determined from static calibration. The decrease ranged from 1 % to 5 % approximately (Table 29).

The maximum forces obtained from the tests performed within this study were corrected by multiplying them by the corresponding ratios between dynamic and static conversion factors shown in the last column of Table 29, under the assumption that the effect of loading rate is independent from the specific acquisition system used (amplifier, signal conditioner, A/D converter) within the applicable range. The resulting maximum forces are labeled $F_{m,CBC}$ in Table 36. The TK machine and striker were excluded from this analyses, since the results in Table 29 had been obtained before having the striker re-gaged.

Once again, the effectiveness of this compliance-based force correction (CBC) was assessed in terms of variation of the F statistic yielded by the ANOVA single-factor analysis (F increases $\rightarrow$ effect is detrimental; F decreases $\rightarrow$ effect is beneficial).

Table 36 - Effect of compliance-based force correction on the agreement among NIST machines and instrumented strikers (excluding TK).

		Original data		Compliance correction			
IRM	**Machine**	$F_{m,or}$ **(kN)**	F	$F_{m,CBC}$ **(kN)**	F	F_{crit}	**Effect of dynamic force adjustment**
LL-137	TO2	32.47	9.05	31.65	20.49	3.25	Detrimental
	TO3	31.52		30.01			
	MPM	31.44		31.22			
HH-107	TO2	26.84	193.40	26.17	77.50	3.15	Beneficial
	TO3	29.06		27.66			
	MPM	26.01		25.83			
SH-38	TO2	33.33	370.25	32.50	106.66	3.18	Beneficial
	TO3	33.98		35.35			
	MPM	31.58		31.36			

Except for the low-energy level (where the effect is only moderately detrimental), it appears that this compliance-based correction procedure had a beneficial effect in improving the between-machine consistency.

8.5 Overview of force correction results

The values of the *F* statistic calculated by means of the ANOVA analyses for the uncorrected and corrected (DFA, DFA_{mod}, and CBC) maximum forces are illustrated in Figure 19 for the three energy levels (excluding the TK machine in all cases). The *F* statistic is inversely proportional to the degree of between-machine agreement.

The most homogeneous IRM (LL-137) shows the best agreement between machines, while the least homogeneous (SH-38) shows the worst.

Even though the calculated *F* values do not approach the critical values (3.1-3.3) that would indicate no statistical difference between the machines, the between-machine agreement improves at all energy levels when different correction/adjustment procedures are applied to the original maximum forces, which derive from the static calibration of the instrumented strikers.

The compliance-based correction appears to be the most effective, and the fact that it represents an attempt to take dynamic loading effects into account is particularly noteworthy. Indeed, from the experience gained by the authors in several decades of performing instrumented Charpy tests, everything seems to indicate that the most effective and reliable type of calibration for a Charpy instrumented striker should be a truly "dynamic" calibration, in which known force values would be applied to the striker at the same (or equivalent) loading rates experienced during Charpy tests (on the order of 10^5 kN/s).

We also observe that the super-high-energy specimens provide the highest *F* values (*i.e.*, the worst between-machine agreement) for both original and corrected forces, irrespective of the correction type used. This can be explained by the fact that these specimens tend to wrap around the striker before being ejected from the machine, and these contact conditions cannot be reproduced by any static calibration, whether corrected or not.

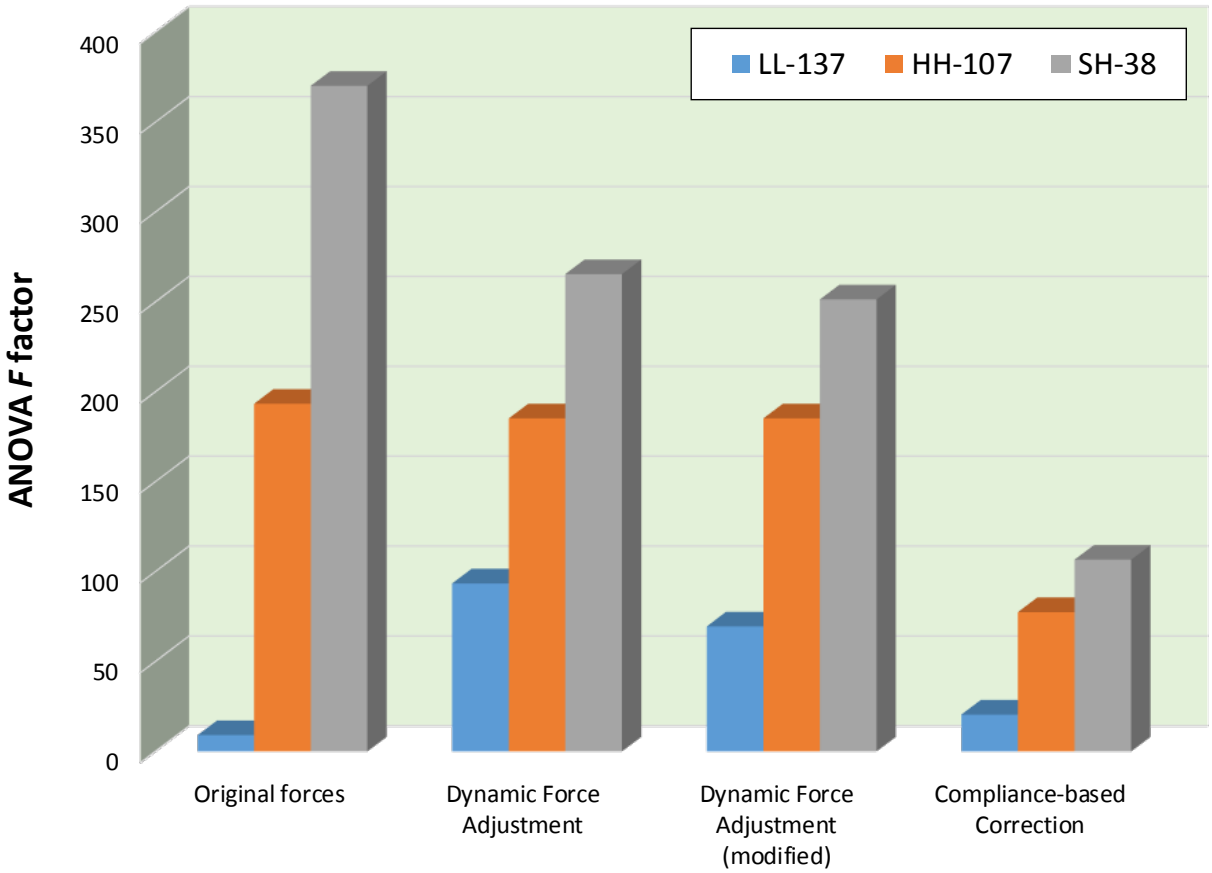

Figure 19 - Values of the F statistic calculated by means of the ANOVA analyses.

At the time this report is compiled, the Structural Materials Group at NIST Boulder is starting a project aimed at establishing a procedure for a true "dynamic calibration" of instrumented Charpy strikers, in cooperation with the Mass and Force Group at NIST Gaithersburg.

9. Conclusions

We have conducted and documented a comparative study of the Charpy machines at NIST Boulder (three master machines and two additional machines used for research purposes), considering both non-instrumented (absorbed energy) and instrumented (maximum forces) parameters.

Historical data collected between 1995 and 2016 on the three master machines, used to certify reference specimens for the indirect verification of Charpy machines all around the world, were statistically analyzed to detect possible trends or systematic differences in terms of absorbed energy and/or data scatter.

Statistically significant differences were not observed among the three master machines at any of the three energy levels (low, high, or super-high), with the sole exception of the TK machine, which consistently yields lower values at the low-energy level. This is suspected to be due to the higher stiffness of the TK machine, which coupled with its unique design (C-type hammer) promotes an increase of its vibrational frequencies and causes fracture forces to be reached at smaller displacements, *i.e.*, at lower absorbed energies. It is noted here that some of the trends observed in the historical grand averages of the results might also be due to machine repairs, such as the replacement of anvils and/or striker, which could affect the relationship between the master machines.

Three lots of verification specimens (low, high, and super-high) were selected and qualified as Internal Reference Materials, to be used in the future for the verification and qualification of NIST impact machines and instrumented strikers. A large number of room temperature impact tests (299, of which 214 were instrumented) were conducted on the five NIST machines, in order to establish reference values and expanded uncertainties for both absorbed energy (*KV*) and maximum force (F_m).

A statistical analysis of the *KV* and F_m test results, by means of Fisher's analysis of variance (ANOVA) and Student's *t*-test, showed significant statistical differences among the NIST machines for both parameters. With regard to the absorbed energy (*KV*) results from the master machines (SI3, TO2, TK), the difference between TK and the other master machines at the low-energy level that had been observed in the historical data was confirmed. At the high- and super-high-energy levels, *KV* data from the master machines were in statistical agreement, as in the case of the historical data.

This large database of instrumented Charpy tests was also used to investigate the relationship between two different measures of absorbed energy: the value returned by the machine encoder (*KV*) and the value obtained by integrating the area under the instrumented force/displacement record (W_t). The two measures should be in general agreement, although from a physical standpoint it is expected that $KV > W_t$ on account of energy components that are unrelated to specimen fracture (such as hammer vibrations and post-fracture secondary impacts).

A direct comparison between *KV* and W_t for the four instrumented machines (TK, TO2, TO3, MPM) showed acceptable agreement at the different energy levels, with the exception of TK (particularly at the low-energy level). For this machine, the conversion factor between force and striker output obtained from the static calibration proved to be inadequate.

We also investigated the effect of several force correction/adjustment procedures on the between-machine consistency, using the *F* statistic returned by the ANOVA analysis. The main conclusions are listed below.

(a) The dynamic force adjustment, based on the equivalence $W_t = KV$, was beneficial in improving the agreement between the machines. However, the effect was significant only when the "outlier" machine (TK) was included in the comparison.
(b) A modified dynamic force adjustment, in which forces were corrected until W_t corresponded to a modified energy *KV* (obtained by removing from *KV* the contributions given by hammer vibrations and secondary impacts), resulted in a more significant improvement of between-machine consistency.
(c) The use of a conversion factor based on machine compliance, which reflects the response of the instrumented striker to forces applied at a higher loading rate than during static calibration, was found to be the most effective approach used to improve the between-machine consistency.

The outcome of this investigation supports the need for a truly "dynamic" calibration, in which force would be applied at loading rates which are comparable to those experienced by the striker during a Charpy test (in the order of 10^5 kN/s). We are presently initiating a collaborative effort to develop a procedure for dynamically calibrating instrumented Charpy strikers.

Acknowledgements

The authors would like to express their gratitude to Ako Chijioke, whose excellent review of the first version of the manuscript improved significantly the quality of this report.

References

[1] S. B. Russell, "Experiments with a New Machine for Testing Materials by Impact," *Transactions*, American Society of Civil Engineers, Vol. 39, No. 826, 1898, pp. 237-250.
[2] G. Charpy, "Note sur l'essai des métaux à la flexion par choc de barreaux entaillés," *Mémoire et compte-rendus de la Société des ingénieurs civils de France*, Paris, 1901.
[3] M. L. Williams and G. A. Ellinger, "Investigation of Fractured Steel Plates Removed from Welded Ships," National Bureau of Standards Report, December 9, 1948.
[4] N. H. Fahey, "The Charpy Impact Test – Its Accuracy and Factors Affecting Test Results," *Materials Research Standards*, Vol. 1, No. 11, Nov 1961, pp. 872-876.
[5] D. E. Driscoll, "Reproducibility of Charpy Impact Test," in *Symposium on Impact Testing*, ASTM STP 176, American Society for Testing and Materials, Philadelphia, PA, 1955, pp. 70-75.
[6] E. Lucon, "Miniaturized Charpy Specimens for the Indirect Verification of Small-Scale Charpy Machines: Initial Qualification Phase," NIST Technical Note 1562-1, July 2012.
[7] R. A. Fisher, "On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample," *Metron* 1, 1921, pp. 3-32.
[8] M. P. Manahan, Sr., R. B. Stonesifer, T. A. Siewert, C. N. McCowan, and D. P. Vigliotti, "Observations on Differences between the Energy Determined Using an Instrumented Striker and Dial/Encoder Energy," in *From Charpy to Present Impact Testing*, ESIS Publication 30, Elsevier, Oxford, UK, 2002, pp. 229-236.
[9] E. Lucon, "Determination of the Compliance of NIST Charpy Impact Machines," NIST Internal Report NISTIR 8043, Feb 2015.

[10] C. N. McCowan, T. A. Siewert, and D. P. Vigliotti, “The NIST Charpy V-Notch Verification Program: Overview and Operating Procedures,” in *Charpy Verification Program: Reports Covering 1989-2002*, NIST Technical Note 1500-9, Sep 2003, pp. 3-42.
[11] J. D. Splett and C. M. Wang, “Uncertainty in Reference Values for the Charpy V-notch Verification Program,” *Journal of Testing and Evaluation*, Vol. 34, No. 3, 2006, pp. 1-4.
[12] Student (W. S. Gossett), “The probable error of a mean,” *Biometrika* 6(1), Mar 1908, pp. 1-25.
[13] E. Lucon, R. Chaouadi, and E. van Walle, “Different Approaches for the Verification of Force Values Measured with Instrumented Charpy Strikers,” *Journal of ASTM International*, Vol. 3, No. 3, 2006, pp. 1-7.
[14] E. Lucon, M. Scibetta, J. D. McColskey, and C. N. McCowan, “Influence of Loading Rate on the Calibration of Instrumented Charpy Strikers,” *Journal of Testing and Evaluation*, Vol. 37, No. 6, 2009, pp. 1-11.
[15] E. Lucon, “On the Effectiveness of the Dynamic Force Adjustment for Reducing the Scatter of Instrumented Charpy Results,” *Journal of ASTM International*, Vol. 6, No. 1, 2009, pp. 1-9.
[16] J. Schuurmans, M. Scibetta, E. Lucon, and J.-L. Puzzolante, “Influence of Strain Gage Position on the Static and Dynamic Performance of Instrumented Impact Strikers,” *Journal of Testing and Evaluation*, Vol. 37, No. 2, 2008, pp. 1-7.
[17] M. P. Manahan, Sr. and R. B. Stonesifer, “The Difference Between Total Absorbed Energy Measured Using an Instrumented Striker and that Obtained Using an Optical Encoder,” in *Pendulum Impact Testing: A Century of Progress*, ASTM STP 1380, West Conshohocken, PA, 2000, pp. 181-197.
[18] D. R. Ireland, “Procedures and Problems Associated with Reliable Control of the Instrumented Impact Test,” in *Instrumented Impact Testing*, ASTM STP 563, ASTM, Philadelphia PA, 1974, pp. 3-29.
[19] A. H. Priest and M. J. May, “Fracture Toughness Testing in Impact,” NDACSS A86, British Iron and Steel Research Association, London, UK, 1969.