JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 87, No. 1, January-February 1982

Contents

Library of Congress Catalog Card Number: 62-37059

For sale by the Superintendent of Documents, U.S. Government Printing Office,
Washington, DC 20402.
Single copy price $4.25 Domestic; $5.35 Foreign.
Subscription price: $18.00 a year; $22.50 for foreign mailing.

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON: 1982

JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 87, No. 1, January–February 1982

Absolute Isotopic Abundance Ratios and Atomic Weight of a Reference Sample of Strontium

L. J. Moore,* T. J. Murphy,* I. L. Barnes,* and P. J. Paulsen*

National Bureau of Standards, Washington, DC 20234

July 15, 1981

Absolute values have been obtained for the isotopic abundance ratios of a reference sample of strontium using solid sample thermal ionization mass spectrometry. Samples of independently known isotopic composition prepared from chemically pure and nearly isotopically pure separated strontium isotopes were used to calibrate the mass spectrometry. The resulting absolute ^{88}Sr/^{86}Sr, ^{87}Sr/^{86}Sr, and ^{84}Sr/^{86}Sr ratios are 8.3786 ± 0.0033, 0.71034 ± 0.00026, and 0.05655 ± 0.00014 respectively which yields atom percents of: ^{88}Sr = 82.5845 ± 0.0066, ^{87}Sr = 7.0015 ± 0.0026, ^{86}Sr = 9.8566 ± 0.0034, and ^{84}Sr = 0.5574 ± 0.0015. The atomic weight calculated from these abundances is 87.61681 ± 0.00012. The indicated uncertainties are overall limits of error based on 95 percent confidence limits for the mean and allowances for the effects of possible systematic error.

Key Words: Absolute ratios; atomic weight; isotopic abundances; strontium.

1. Introduction

The analytical mass spectrometry group of the National Bureau of Standards is conducting a long term program of absolute abundance ratio and atomic weight determinations on polynuclidic elements using predominantly thermal ionization mass spectrometry though, on occasion, electron impact ionization has been used. Previous elements studied include silver [1],[1] chlorine [2], copper [3], bromine [4], chromium [5], magnesium [6], lead [7], boron [8], rubidium [9], rhenium [10], silicon [11], potassium [12], and thallium [13]. The present work extends the study to strontium.

Natural strontium consists of four isotopes, ^{88}Sr, ^{87}Sr, ^{86}Sr, and ^{84}Sr. One of these, ^{87}Sr, is constantly, though slowly, accumulating as the end product of the decay of ^{87}Rb; the other three are believed to be of stable abundance.

The present atomic weight of strontium, 87.62, is based on the relative isotopic measurements of A. O. Nier [14] on a piece of pure strontium metal. Since this early work there has been no serious effort focused toward the determination of the absolute isotopic abundances of strontium. In the interim, the value of 0.1194 determined by Nier for the ratio ^{86}Sr/^{88}Sr has been nearly universally adopted by the geological community as an interlaboratory standardization value in the mass spectrometric measurement of radiogenic ^{87}Sr.

The large limit of error (±2 percent) associated with Nier's measurement of ^{86}Sr/^{88}Sr left some uncertainty about a possible systematic error associated with the currently accepted procedure of correcting for instrumental fractionation of the isotopes by normalizing to Nier's ^{86}Sr/^{88}Sr ratio (=0.1194). Thus, one primary purpose for the present work was to provide absolute values for the strontium isotopic ratios which might be used as a reference for the hundreds of papers published annually dealing with strontium isotope geology and geochronology.

In the present study the mass spectrometers were calibrated for bias by the use of samples of independently known ^{88}Sr/^{86}Sr ratios, prepared from chemically pure and nearly isotopically pure ^{88}Sr and ^{86}Sr solutions. These measured biases were then used to correct the raw data obtained on a reference sample of strontium thus yielding absolute values for this sample. The reference sample selected for this work was SRM 987, Strontium Carbonate, which is a highly purified material previously certified for strontium content.

2. Experimental Procedure

2.1. Mass Spectrometry

Isotopic measurements were performed using a triple filament rhenium ion source. The ionizing filaments used in the measurements were rigorously degassed at 5 A (>2000 °C) for 2 h to reduce any background ion contributions to the ^{86,87,88}Sr mass positions to <0.01 percent of the ion beam intensities used in the measurements of the "natural" strontium, SRM 987. The viability of this procedure was supported by scanning these mass positions while heating a "blank" ionizing filament at temperatures in excess of the usual operating temperature (1600 °C). Any possible contribution to these masses under actual analysis conditions was evaluated

---

* Center for Analytical Chemistry, National Measurement Laboratory.

[1] Figures in brackets indicate literature references at the end of this paper.

by determining the absence of minor isotope perturbations in the analyses of SRM 988, ^{84}Sr spike, which has ^{86}Sr, ^{87}Sr, and ^{88}Sr atom fractions of 0.00059 ± 0.00001, 0.00010 ± 0.00001, and 0.00039 ± 0.00001, respectively [15]. Using this technique, any contribution to the ^{85}Rb mass position was observed to be $<10^{-15}$ A, and a factor of 2.6 less for the ^{87}Rb based on a natural ^{85}Rb/^{87}Rb ratio of 2.59 [9].

Details of the strontium isotopic analysis technique have been published [16]. Generally, the procedure consists of adding a drop of solution [500 μg Sr/mL in 2% (V/V)HNO_3] to each of two degassed 0.001 inch × 0.030 inch regular rhenium filaments, drying under an infrared lamp and heating the resulting deposit with sequentially higher alternating currents to a final dull red heat. In the latter step the deposit "collapses" on the filament and undergoes an apparent reaction with the rhenium filament material to form a compound probably analogous to the $Ca_3Re_2O_9$ and $CaReO_4$ detected by electron diffraction earlier during Ca analyses [17].

This technique has been used extensively in this laboratory for hundreds of strontium analyses, and has been adopted by several geological laboratories around the world. Experience has shown that inter-analysis variations in the unnormalized ^{86}Sr/^{88}Sr ratio can be held to <0.05 percent, relative, and usually substantially better using rigidly controlled analysis parameters.

The two spectrometers used in these measurements were nominally identical 90° sector, 30 cm radius of curvature instruments equipped with thin lens Z-focusing ion sources and a conventional NBS-designed collector appropriately biased and baffled to minimize problems due to secondary particles [18]. Both measuring systems used Cary 401 MR[2] vibrating reed electrometers whose attenuator resistors and amplifier linearities were calibrated relative to each other within ± 0.01 percent by placing precisely known relative voltages into the feedback loop of the electrometer. The 1 V analog output of the electrometer was digitized using two different systems: the instrument of operator I used a Hewlett-Packard 2212B voltage-to-frequency convertor (full scale = 100 kHz) coupled to a high speed ATEC scaler; operator II's instrument used a Teledyne (4501A) voltage-to-frequency convertor unit in tandem with a high speed scaler of NBS design. Data acquisition and instrument control for both instruments were achieved with Hewlett-Packard 9830A(B) programmable calculators, using software developed in this laboratory. In the absence of a priority interrupt capability for the calculator, the software was arranged such that the data averaging and printing segment at the end of each peak top measurement was completed in time for the calculator to be ready for the next data transmission sequence from the scaler-VFC unit.

In the process of thermally vaporizing and ionizing samples, a mass dependent isotopic discrimination occurs that produces an observed isotope ratio that is not characteristic of the sample. This discrimination occurs among sample loadings and during a sample analysis. To minimize any discrimination-induced skewing of the relative isotope ratios with respect to a common time base during an analysis, the isotope ratios were measured symmetrically with respect to mass. Typically, each peak was monitored for 10 precisely timed 1 s integrations, and 8–10 s (30–60 s for the separated isotope mixtures) were allowed between signal integration periods to allow for the RC decay of the electrometer as well as for a settling time for the magnetic field. Normally 10 to 20 ratios were measured for each isotope per analysis.

2.2. Purification of the Separated Isotopes

To prepare accurate isotopic standards for strontium, it is necessary to know, as accurately as possible, the concentration of strontium in solutions of separated isotopes of strontium. To accomplish this objective a purification method and an assay procedure were developed that enabled us to determine the concentration of strontium in a solution with an accuracy of ±0.01 percent.

Electromagnetically separated ^{86}Sr and ^{88}Sr isotopes in the form of strontium nitrate were obtained from the Nuclear Division of the Oak Ridge National Laboratory. The $^{86}Sr(NO_3)_2$ was designated series 136801 and the $^{88}Sr(NO_3)_2$ was designated series 137001. The certificates of analysis which accompanied each sample included a semi-quantitative spectrographic analysis which showed that several impurity elements could be present at the 0.02 to 0.05 percent level.

To reduce those impurities to a level low enough so that they could not cause a significant error in the assay of strontium, the separated isotope samples were further purified. The purification procedure used was based on the relative insolubility of strontium nitrate in concentrated nitric acid. Only two elements, barium and lead, are known to co-precipitate under this condition [19]. If a dilute nitric acid solution is evaporated, the nitric acid concentration increases until the azeotropic solution containing 68 percent HNO_3 is reached [20]. In 68 percent HNO_3 the solubility of strontium nitrate is almost zero [21] while the solubility of most cations except barium and lead is sufficient to keep them in solution [19]. Lead can be efficiently removed by anodic electrodeposition but barium is not separated and remains as an impurity in the strontium nitrate.

The effectiveness of the purification procedure described below was first tested by the purification of natural strontium nitrate which had been doped with 1000 ppm each of 30 common impurity elements.

The results of the analysis of this material by spark source and thermal isotope dilution mass spectrometry are shown in

---

[2] Certain commercial equipment, instruments, or materials are identified in this paper to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the National Bureau of Standards nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

table 1. The methods of analysis are also described below. Only sodium was detected at a concentration greater than 1 µg/g while all of the other elements were at the sub-ppm level. Barium was not determined since the purification method would not remove it and a correction for its concentration in each separated isotope is necessary.

Each strontium separated isotope was purified as follows: The strontium nitrate, $Sr(NO_3)_2$, (about 5.0 g) was transferred to a 100 mL Teflon-FEP beaker and dissolved in 50 mL of water. One mL of 1 mol/L hydrochloric acid was added and the solution was heated at about 90 °C on a hot plate for two hours. The solution was filtered through a close-textured filter paper into a Teflon-FEP beaker. Five mL of high-purity nitric acid was added to the filtrate and the solution was heated to about 90 °C on a hot plate. Strontium nitrate was crystallized by allowing the solution to evaporate until only about 3 mL of solution remained (68% HNO_3). The beaker and contents were allowed to cool to room temperature. The solution was withdrawn from the crystals of strontium nitrate by using a plastic hypodermic syringe with a platinum needle. The crystals were washed twice with 1 mL portions of nitric acid which were withdrawn in the same manner. The excess nitric acid was then removed by heating the beaker until the crystals were dry.

TABLE 1. *Analysis of Purified Strontium*

Element	Spike Isotope	Natural $Sr(NO_3)_2$ µg/g	^{86}Sr $Sr(NO_3)_2$ µg/g	^{88}Sr $Sr(NO_3)_2$ µg/g
Ag	^{109}Ag	0.4[a]	0.03	0.1
Al	^{26}	0.4[a]	0.1	0.5
As	^{86}Se	0.03[a]	—	—
Ba[b]	^{135}Ba	—	12.5	0.73
Bi	^{203}Tl	0.04[a]	—	—
Ca[b]	^{42}Ca	0.10[a]	0.78	0.30
Cd	^{111}Cd	0.01[a]	0.01	0.01
Ce	^{142}Ce	0.004[a]	0.02	0.2
Co	^{62}Ni	0.01[a]	—	—
Cr	^{53}Cr	0.05[a]	0.07	0.06
Cu	^{65}Cu	0.04[a]	0.30	0.2
Fe	^{54}Fe	0.3[a]	0.54	1.2
Ga	^{71}Ga	0.02	0.01	0.01
In	^{113}In	0.01	0.01	0.01
K	^{41}K	0.05[a]	0.2	0.4
La	^{142}Ce	0.02[a]	—	—
Mg	^{26}Mg	0.2[a]	0.06	0.1
Mn	^{54}Fe	0.04[a]	—	—
Mo	^{97}Mo	0.12[a]	0.03	0.02
Na	^{41}K	2[a]	0.7	0.7
Nd	^{145}Nd	0.1	0.01	0.03
Ni	^{62}Ni	0.01[a]	0.03	0.06
Pb[b]	^{206}Pb	0.04[a]	0.13	0.06
Sb	^{117}Sn	0.1[a]		
Se	^{82}Se	0.01[a]	0.01	0.01
Sn	^{117}Sn	0.3[a]	0.1	0.1
Te	^{125}Te	0.1[a]	0.05	0.08
Tl	^{203}Tl	0.05[a]	0.07	0.1
Ti	^{47}Ti	0.05[a]	0.02	0.06
V	^{53}Cr	0.01[a]	0.1	8.8
Zn	^{67}Zn	0.03[a]		
Zr	^{91}Zr	0.0021	0.01	0.01

[a] Doped with 1000 µg/g of each element before purification.

[b] Analyzed by thermal ionization mass spectrometry.

Lead was removed from the crystallized strontium nitrate by dissolving it in about 50 mL of water and electrodepositing overnight onto platinum wire electrodes at an applied voltage of 2.0 V. Under those conditions, lead will deposit as PbO_2 onto the anode.

Only a slight deposit (<5 µg Pb) was noted on the anode from the "Sr 86" solution but there was a heavy deposit on the anode from the "Sr 88" solution with considerable PbO_2 on the bottom of the beaker. The "Sr 88" solution was filtered thru acid washed close-textured filter paper. After the addition of 5 mL of nitric acid to each solution, strontium nitrate was recrystallized as described above and lead was removed by electrodeposition in the same manner. Only a faint deposit (about 1 µg) was noted on the anode from the "Sr 86" solution but once again the deposit on the anode from the "Sr 88" was heavy and particles of PbO_2 were on the bottom of the beaker, necessitating filtration of the solution.

The cycle of crystallization and electrodeposition was repeated four times for the "Sr 88" solution until the anode showed no evidence of PbO_2 after electrodeposition. (The total amount of lead from the "Sr 88" solution amounted to 0.1% of the starting weight of $^{88}Sr(NO_3)_2$.)

The final solutions from the lead separations were evaporated to dryness and the salts were transferred to two platinum crucibles, dried, and weighed. Calculations based on the starting weights of $^{86}Sr(NO_3)_2$ and $^{88}Sr(NO_3)_2$ and the weights of the corresponding purified salt showed that about 99 percent of the strontium was recovered in each case.

The acids and water used in those purifications were produced at NBS by sub-boiling distillation [22] and have been shown to be extremely low in trace cation contamination. Apparatus such as beakers and filters were cleaned with high-purity acid before use.

2.3. Analysis of Purified Strontium Separated Isotopes

About 0.5g samples of each purified strontium isotope as $Sr(NO_3)_2$ were dissolved in 10 mL of water in a Teflon beaker and spiked with 10^{-7}g of ^{109}Ag, ^{111}Cd, ^{142}Ce, ^{53}Cr, ^{65}Cu, ^{54}Fe, ^{71}Ga, ^{113}In, ^{26}Mg, ^{97}Mo, ^{145}Nd, ^{62}Ni, ^{82}Se, ^{117}Sn, ^{125}Te, ^{47}Ti, ^{203}Tl, ^{67}Zn, and ^{91}Zr. Each solution was treated as follows: 1 mL of nitric acid was added and the solution was evaporated until most of the strontium crystallized and about 0.5 mL of nitric acid remained. The nitric acid was removed from the crystallized strontium nitrate with a polypropylene hypodermic syringe equipped with a platinum needle and transferred to a Teflon beaker. The crystals of $Sr(NO_3)_2$ were washed twice with 0.5 mL portions of nitric acid which were removed in the same manner and combined with the acid from the crystallization. This nitric acid was concentrated to a small drop, evaporated to dryness on high purity gold wires, and analyzed by spark-source mass spectrometry. The results of this analysis are also shown in table 1.

In addition to the spiked elements, a number of other elements were determined by comparison to spiked nuclides and assuming that the sensitivities of the natural element and spike nuclide are the same.

Three elements, Ba, Pb, and Ca were determined by thermal ionization mass spectrometry since they were not obtained from the spark source analysis. Barium and lead are co-crystallized with the $Sr(NO_3)_2$ and calcium could not be determined because of Sr interference ($M/Z = 44$ for $^{88}Sr^{++}$). About 0.2 g samples were spiked with ^{206}Pb, ^{42}Ca, and ^{135}Ba. Calcium was separated in the same manner as the separation for the spark source analysis. Lead was separately electrodeposited from a nitrate solution onto a platinum anode. Barium was separated by adding lead and co-precipitating it with lead chromate by the addition of ammonium chromate. The chromates were caught on a filter and dissolved with 1.5 mol/L HCl. Lead and chromate ions were removed by passing the solution through a strongly basic anion exchange column, and the barium was recovered by evaporation of the eluate. The concentration of each analyte was determined from isotope ratio measurements using thermal ionization mass spectrometry.

The results of these analyses are also shown in table 1. Barium, which was not removed by the purification, was determined to be 12.5 µg/g in the $^{86}Sr(NO_3)_2$ and 0.73 µg/g in the $^{88}Sr(NO_3)_2$. This would cause an error of about 0.002 percent in the ^{86}Sr assay and <0.001 percent for the ^{88}Sr assay. Since these errors are not significant, corrections were not applied. No other element was found at a level high enough to cause a possible significant error in the assay procedure.

2.4. Preparation and Assay of the Separated Isotopic Solutions

The 4.3 g of purified $^{86}Sr(NO_3)_2$ was dissolved in about 50 mL of 0.5 mol/L HNO_3, transferred to a specially constructed and tared 500 mL quartz flask and diluted to about 500 mL with 0.5 mol/L HNO_3. This flask was constructed from a 500 mL quartz flask by cutting the neck of the flask about 1 cm from the body and tooling the neck for a number zero polyethylene stopper and standard aluminum serum cap.

The 4.3 g of purified $^{88}Sr(NO_3)$ was also dissolved in about 50 mL of 0.5 mol/L HNO_3, transferred to a specially constructed 200 mL quartz flask and diluted to about 200 mL with 0.5 mol/L HNO_3. This flask was constructed in the same manner as described for the 500 mL flask. The bottles and contents were weighed to ±0.2 mg and the preliminary strontium concentration was calculated. The solution of $^{86}Sr(NO_3)_2$ was designated "Sr 86" and the solution of $^{88}Sr(NO_3)_2$ was designated "SR 88".

Four weighed portions each containing about 1.6 mmol of strontium (~16 g each for the ^{88}Sr solution and ~40 g each for the ^{86}Sr solution) were withdrawn from each separated isotope solution in the following manner: a 10-cm platinum needle was inserted through a No. 0 polyethylene stopper which was used to replace the cap in the bottle. A 10-mL polyethylene hypodermic syringe with the plunger covered with a thin sheet of Teflon was attached to the Kel-F hub of the needle and the desired amount of solution was withdrawn. The syringe was then disconnected from the hub and the tip was capped with a Kel-F cap. Any static charge that might be present on the plastic syringe was dissipated by wiping it with a damp lintless towel. The syringe and contents were weighed on a semimicrobalance to ±0.02 mg. The solution was then delivered from the syringe into a 100 mL Teflon-FEP beaker and the syringe was again capped, wiped, and weighed. The weight of the sample was determined from the weight of the syringe before and after delivery of the sample. Two assay samples were withdrawn from each solution before and after withdrawing the calibration samples to ensure that no change in concentration had occurred during this time interval (about 3 h).

Each portion was then assayed as follows: 1 mL of perchloric acid was added and the solution was evaporated at low heat to fumes of $HClO_4$. The solution was then cooled, taken up in a few mL of water, and again evaporated to fumes of $HClO_4$. This procedure was repeated and the excess perchloric acid was removed by evaporation. (This procedure converts $Sr(NO_3)_2$ to $Sr(ClO_4)_2$ and is necessary because nitrate causes significant errors in the determination of strontium as sulfate.)

The residue was taken up in 20 mL of water and heated. Fifteen mL of warm (1 + 99) H_2SO_4 was added dropwise to the hot solution while swirling the beaker to precipitate $SrSO_4$. The sulfuric acid was added at a rate that required about 15–20 min to complete the addition. The solution was heated at 80–90 °C for 2 h and evaporated to about 15 mL of solution. Fifteen mL of 95 percent ethyl alcohol was added and mixed with the solution. After tightly covering the beaker with plastic film, it was allowed to stand overnight to complete the precipitation.

The solution was then filtered through a tared platinum Monroe crucible and the precipitated strontium sulfate was caught on the platinum mat. Any $SrSO_4$ adhering to the walls of the beaker was freed by scraping with a Teflon "policeman" and transferred to the crucible with a stream of 95 percent ethyl alcohol. The filtrate, washings, and Teflon "policeman" were transferred back to the original beaker and reserved for the determination of dissolved and untransferred strontium. The crucible and contents were dried for 2 h at 105 °C, ignited at 800 °C for 8 h, cooled in a desiccator, transferred to the case of a micro-balance and allowed to stand for at least 2 h. The crucible and contents were weighed on the micro-balance to ±0.002 mg. A buoyancy correction for platinum crucibles was made by averaging three empty tare crucibles. (The drying, cooling, and weighing were repeated to insure constant weight.) The air weight of the $SrSO_4$ was then determined and converted to vacuum weight using 3.92 for the density of $^{86}SrSO_4$ and 3.97 for the density of $^{88}SrSO_4$. These densities were calculated by assuming that they were proportional to the density of natural $SrSO_4$, 3.96 in the same relationship as their molecular weight. The vacuum weight of the strontium sulfate was converted to millimoles of strontium using a calculated atomic weight for strontium and the 1977 atomic weight values for sulfur and oxygen. The formula weights used were 182.0066 for the $^{86}SrSO_4$ and 183.9612 for the $^{88}SrSO_4$.

To determine the soluble and untransferred strontium, the filtrate and washings which had been reserved in the original beaker were spiked with about 1.5 mmol of ^{84}Sr. The ^{84}Sr was NBS SRM 988, Strontium 84 Spike Assay and Isotopic Solution Standard. The solution was heated and concentrated to about 10 mL to remove ethyl alcohol. Fifteen mL of HCl and 5 mL of (1 + 99) H_2SO_4 were added, and the solution was diluted to 100 mL with water. The solution was then heated for two hours to ensure equilibration of the natural and spike strontium. The solution was evaporated to fumes of perchloric acid and then the heat was increased to fumes of sulfuric acid. After cooling, the residue was taken up in 15 mL of water and passed through a cation exchange column, containing 5 mL of AG50Wx8, resin. The column was then washed with 50 mL of 0.5 mol/L HCl and the Sr was eluted with 25 mL of 6 mol/L HCl. This solution was evaporated to dryness. The residue was converted to the nitrate by the addition of a few drops of nitric acid followed by evaporation. The residue was taken up and the 84/86 or 84/88 ratio was determined by thermal ionization mass spectrometry. The Sr found as mmol Sr was added to the strontium from the gravimetric determination to yield the total strontium in the sample.

This method of determining the strontium was first tested on solutions containing known amounts of strontium. Ten solutions were prepared from high purity strontium carbonate. The material used was SRM 987, Strontium Carbonate, which has an assay value of 99.98 ± 0.02 percent. The solutions were prepared in the approximate concentration range of 0.10 to 0.12 mmol/g. Four samples containing from 1.50 to 1.79 mmol were withdrawn from each solution and the strontium was determined as described above. Comparison of the calculated to measured concentration detected a small positive bias of about 0.01 percent, but this would have a negligible effect on ratios.

The results of the assay of the separated isotope solutions are shown in table 2. Pooling the results of the separated isotope solutions and the ten sets described above, yields an uncertainty of 0.119×10^{-5} mmol/g for the concentration of the "Sr 86" solution and 0.297×10^{-5} mmol/g for the concentration of the "Sr 88" solution.

2.5. Preparation of Calibration Samples

Six calibration samples were prepared by mixing weighed portions of "Sr 86" and the "Sr 88" solutions to produce 88/86 ratios, ranging from 8 to 0.7. Four of the calibration samples were within a few percent of the natural 88/86 ratio of 8.38 and two bracketed the natural 87/86 ratio of 0.71. The portions were withdrawn from the flasks and weighed in the manner previously described for assay of the solution. To eliminate any possibility of change in the concentration of the isotope solution with time, the portions for the calibration samples were withdrawn from the flasks between the samples taken for assay over a period of about 6 h.

TABLE 2. *Assay of Strontium Separated Isotope Solutions Separate I*

Solution	Sample No.	Weight $SrSO_4$ (vac) g	Sr from $SrSO_4$ mmol	Sr from Filtrate mmol	Total Sr mmol	Weight Sample g	Conc. Sample mmol/g
^{86}Sr	1	0.293588	1.613062	0.002681	1.615743	40.89026	0.0395141
	2	.296739	1.630375	.001504	1.631878	41.30356	.0395094
	3	.293032	1.610006	.003275	1.613281	40.82803	.0395141
	4	.298813	1.609352	.002365	1.611717	40.78996	.0395126
	Average						.0395175
^{88}Sr	1	0.320318	1.741227	0.001667	1.742894	17.64999	0.098748
	2	.302946	1.646795	.001854	1.648649	16.69340	.098761
	3	.301413	1.638460	.001616	1.640076	16.60703	.098758
	4	.310012	1.685206	.002920	1.688126	17.09511	.098749
	Average						.098754

Table 3 shows the composition of these calibration samples. The isotope ratio of each calibration sample was calculated from the isotopic analysis of the separated isotopes and from the mmols of strontium from each separated isotope solution as determined from the assay and weight of solution taken.

Each calibration sample was thoroughly mixed and evaporated to dryness at low heat on a hot plate. The calibration samples were taken up in (1 + 49) nitric acid so that 1 mL of solution contained 500 µg of strontium.

3. Results and Discussion

Isotopic compositions of the purified ^{86}Sr and ^{88}Sr were each measured by Operator #I, using 7–10 sample analyses for each separated isotope. Before and after each set of separated isotope analyses, the ion source was removed from the instrument and meticulously cleaned. Prior to each set of separated isotope analyses, the source was carefully checked to insure the absence of any cross contamination. The results of the measurement of the separated isotope solutions are shown in table 4.

The six synthetic isotope mixes were selected randomly and measured two or three times by each operator for its ^{88}Sr/^{86}Sr ratio. SRM 987 was analyzed after every two mixes, and the average of these SRM analyses provided the data used in the final computation of the absolute abundance ratios of the reference sample.

The results of the measurements of the separated isotope mixtures as well as the correction factors for each operator are shown in table 5.

In Table 6 are shown ratios for the reference material (SRM 987–Strontium Carbonate) for each operator.

The calculations of the atomic weight of the reference sample of strontium are shown in table 7. The value calculated is only applicable to the reference material (SRM 987) since samples in nature with large variations in the ^{87}Sr/^{86}Sr ratios

TABLE 3. *Isotopic Composition of Calibration Samples*

Solution No.	Isotope Solution	Weight Solution g	Sr from Solution *mmol*	^{86}Sr from Solution *mmol*	^{88}Sr from Solution *mmol*	Total ^{86}Sr Solution mmol	Total ^{88}Sr Solution mmol	Ratio 88/86 Solution
1	"Sr 86"	5.05855	0.199876	0.195191	0.003349			
	"Sr 88"	16.28401	1.608108	.000741	1.605506	0.195932	1.608855	8.211280
2	"Sr 86"	4.94747	0.195487	.190905	0.003276			
	"Sr 88"	16.45020	1.594893	.000735	1.592313	.191640	1.595589	8.325967
3	"Sr 86"	4.98057	0.196795	.192182	0.003298			
	"Sr 88"	16.50848	1.630275	.000751	1.627637	.192934	1.630935	8.453349
4	"Sr 86"	5.20231	0.205557	.200738	0.003444			
	"Sr 88"	17.41636	1.719931	.000793	1.717149	.201531	1.720593	8.537609
5	"Sr 86"	17.91602	1.707908	.691315	0.011862			
	"Sr 88"	5.02571	0.496308	.000229	0.495505	.691543	0.507367	0.733673
6	"Sr 86"	18.24981	0.721097	.709195	0.012083			
	"Sr 88"	4.89190	0.483094	.000223	0.482312	.704417	.494395	.7018501

TABLE 4. *Isotopic Composition of Separated Strontium Isotopes Used in Calibration Samples*

Separated Isotope	Isotopic Composition Atom Percent	
"Sr-86"	88	1.673483 (365)[a]
	87	0.655174 (384)
	86	97.658689 (721)
	84	0.012651 (507)
"Sr-88"	88	99.838084 (260)
	87	0.115774 (234)
	86	0.046142 (110)
	84	0 (<4 ppm)

[a] Numbers in brackets are the errors (95% confidence limits) on the last digits of the numbers given.

are known. Subsequent to the Faraday cage measurements described here, high sensitivity pulse counting measurements in this laboratory have indicated the possibility of sporadic backgrounds at the $M/Z = 86$ position of $\sim 10^{-16}$ A. Although the background has tentatively been identified as a hydrocarbon residue resulting from previous exposure to resin beads in the specific pulse counting system (not either instrument used in the Sr measurements), an extra allowance of 0.01 percent has been added to the systematic error of the ^{88}Sr/^{86}Sr and ^{87}Sr/^{86}Sr ratios as a precaution.[3]

It should be noted that the value of the ^{86}Sr/^{88}Sr ratio of 0.119351 is very close to the less precise value of 0.1194 universally used. The continued use of this historical value seems justified except where the very highest accuracy is required.

---

[3] Recently published research [23] has demonstrated the formation of molecular parent and decomposition ions from quaternary ammonium salts (e.g., $M/Z = 86 = C_5H_{12}N^+$) in a conventional thermal ionization source. These data support the observation that exposure of the pulse counting mass spectrometer to anion exchange resin beads (quaternary ammonium salts) was the probable cause for background peaks in the strontium mass region for that specific mass spectrometer. Since: (a) the mass spectrometers used for strontium were never exposed to these resin beads; (b) all other measurement evidence in the strontium data do not reflect a statistically significant background; (c) no higher intensity molecular ion peaks were observed that one would expect to accompany those of the strontium mass region, we believe the isotope ratio data for SRM 987 to be free of systematic bias well within the stated error limits.

TABLE 5. *Determination of Mass Spectrometer Bias*

Calibration Sample No.	Isotopic Ratio ^{88}Sr/^{86}Sr Calculated	Isotopic Ratio ^{88}Sr/^{86}Sr Operator I	Isotopic Ratio ^{88}Sr/^{86}Sr Operator II	Correction Factor Operator I	Correction Factor Operator II
1	8.211280	8.200908	8.197613	1.001265	1.001667
2	8.325967	8.315727	8.315177	1.001231	1.001298
3	8.453349	8.442254	8.443429	1.001314	1.001175
4	8.537609	8.527686	8.527326	1.001164	1.001201
5	0.733673	0.732628	0.732525	1.001426	1.001567
6	.701850	.700835	.700891	1.001448	1.001368
		Mean Values of Correction Factors		1.001308	1.001379
	Average			1.001344	

TABLE 6. *Isotopic Ratios of SRM 987, SrCO$_3$*

Ratio	Operator 1	Operator 2	Average Ratio[a]	Correction Factor	Corrected Ratio
^{88}Sr/^{86}Sr	8.3678911	8.36689550	8.3673671	1.0013439	8.378612
^{87}Sr/^{86}Sr	0.7099313	0.7098806	0.7098619	1.0006719	0.7103389
^{84}Sr/^{86}Sr	0.0566361	0.0566166	0.0566253	0.9986579	0.05654927

[a] Because no statistically significant differences were found between the data of operators 1 and 2 they were combined into a single set for the calculation of average ratio, correction factor, and corrected ratio. The appropriate error limits are given in Table 7.

TABLE 7. *Atomic Weight, Atom Percent, and Isotopic Ratios of Strontium (corrected for systematic errors)*

	Value	Total Uncertainty	Due to Error in Calibration Factor	Due to Error in Ratio Determination	Due to Systematic Error in Determination	Due to Error in Nuclidic Masses of Isotopes
Atomic Weight	87.616814	0.000117	0.0000477	0.0000266	0.0000397	0.0000029
Atom Percent						
Sr-84	0.55738	0.00155	0.000211	0.000412	0.00093	
Sr-86	9.85659	.00337	.001718	.000724	.00093	
Sr-87	7.00152	.00263	.000506	.000862	.00126	
Sr-88	82.58451	.00657	.002435	.001458	.00268	
Isotopic Ratios						
84/86	0.056549	0.000143	0.0000115	0.0000420	0.00009	
87/86	.710339	.000261	.0000725	.0000760	.00007	
88/86	8.37861	.003248	.0017070	.0007409	.00080	

The authors are indebted to Drs. John Orban and Keith Eberhardt for the statistical analysis of the experimental data and to Mr. W. A. Bowman for assistance with instrument maintenance. We also thank Mrs. Joy Shoemaker and Teresa Sperow for their painstaking typing of the manuscript.

4. References

[1] Shields, W. R., Garner, E. L., and Dibeler, V. H., Absolute Isotopic Abundance of Terrestrial Silver, J. Res. Nat. Bur. Stand. (U.S.) **66A** (Phys. and Chem.) 1 (1962).

[2] Shields, W. R., Murphy, T. J., Garner, E. L., and Dibeler, V. H., Absolute Isotopic Abundance Ratio and the Atomic Weight of Chlorine, J. Am. Chem. Soc., **84**, 1519 (1962).

[3] Shields, W. R., Murphy, T. J., and Garner, E. L., Absolute Isotopic Abundance Ratio and the Atomic Weight of a Reference Sample of Copper, J. Res. Nat. Bur. Stand. (U.S.) **68A** (Phys. and Chem.) 589 (1964).

[4] Catanzaro, E. J., Murphy, T. J., Garner, E. L., and Shields, W. R., Absolute Isotopic Abundance Ratio and the Atomic Weight of Bromine, J. Res. Nat. Bur. Stand. (U.S.) **68A** (Phys. and Chem.) 593 (1964).

[5] Shields, W. R., Murphy, T. J., Catanzaro, E. J., and Garner, E. L., Absolute Isotopic Abundance Ratio and the Atomic Weight of a Reference Sample of Chromium, J. Res. Nat. Bur. Stand. (U.S.) **70A** (Phys. and Chem.) 193 (1966).

[6] Catanzaro, E. J., Murphy, T. J., Garner, E. L., and Shields, W. R., Absolute Isotopic Abundance Ratios and Atomic Weight of Magnesium, J. Res. Nat. Bur. Stand. (U.S.) **70A** (Phys. and Chem.) 453 (1966).

[7] Catanzaro, E. J., Murphy, T. J., Shields, W. R., and Garner, E. L., Absolute Isotopic Abundance Ratios of Common, Equal-Atom, and Radiogenic Lead Isotopic Standards, J. Res. Nat. Bur. Stand. (U.S.) **72A** (Phys. and Chem.) 261 (1968).

[8] Catanzaro, E. J., Champion, C. E., Garner, E. L., Marinenko, G., Sappenfield, K. M., and Shields, W. R., Nat. Bur. Stand. (U.S.) Spec. Publ. 260-17, 70 pages (Feb. 1969).

[9] Catanzaro, E. J., Murphy, T. J., Garner, E. L., and Shields, W. R., Absolute Isotopic Abundance Ratio and Atomic Weight of Terrestrial Rubidium, J. Res. Nat. Bur. Stand. (U.S.) **73A** (Phys. and Chem.) 511 (1969).

[10] Gramlich, J. W., Murphy, T. J., Garner, E. L., and Shields, W. R., Absolute Isotopic Abundance Ratio and Atomic Weight of a Reference Sample of Rhenium, J. Res. Nat. Bur. Stand. (U.S.) **77A** (Phys. and Chem.) 691 (1973).

[11] Barnes, I. L., Moore, L. J., Machlan, L. A., Murphy, T. J., and Shields, W. R., Absolute Isotopic Abundance Ratios and the Atomic Weight of a Reference Sample of Silicon, J. Res. NBS (U.S.), **79A** (Phys. and Chem.), 727–740 (1975).

[12] Garner, E. L., Murphy, T. J., Gramlich, J. W., Paulsen, P. J., and Barnes, I. L., Absolute Isotopic Abundance Ratios and the Atomic Weight of a Reference Sample of Potassium, J. Res. Nat. Bur. Stand. (U.S.) **79A** (Phys. and Chem.) 713 (1975).

[13] Dunstan, L. P., Gramlich, J. W., Barnes, I. L., and Purdy, W. C., Absolute Isotopic Abundance and the Atomic Weight of a Reference Sample of Thallium, J. Res. Nat. Bur. Stand. (U.S.) **85** (Phys. and Chem.) 1–10 (1980).

[14] Nier, A. O., The Isotopic Constitution of Strontium, Barium, Bismuth, Thallium, and Mercury, Phys. Rev. **5**, 275–278 (1938).

[15] Certificate of Analysis, SRM 988, Office of Standard Reference Materials, NBS, Washington, D.C. 20234.

[16] Moore, L. J., Moody, J. R., Barnes, I. L., Gramlich, J. W., Murphy, T. J., Paulsen, P. J., and Shields, W. R., Trace Determination of Rubidium and Strontium in Silicate Glass Standard Reference Materials, Anal. Chem. **45**, 2384–2387 (1973).

[17] Moore, L. J., Machlan, L. A., High Accuracy Determination of Calcium in Blood Serum by Isotope Dilution Mass Spectrometry, Anal. Chem. **44**, 2291–2296 (1972).

[18] Shields, W. R. (ed.), Nat. Bur. Stand. (U.S.), Tech. Note. 277, 99 (July 1966).

[19] Willard, H. W., and Goodspeed, E. W., Separations of Strontium, Barium, and Lead from Calcium and Other Metals, Ind. and Eng. Chem., Anal. Ed., **8**, 414–418 (1936).

[20] Roscoe, H. E., On the Composition of the Aqueous Acids of Constant Boiling Point, J. Chem. Soc., **13**, 146–154 (1860).

[21] Sieverts, A., and Petzold, W. Z., Das System Strontiumnitrat-Salpetersäure-Wasser, Anorg. Chem., **214**, 27–32 (1933).

[22] Kuehner, E. C., Alvarez, R., Paulsen, P. J., Murphy, T. J., Production and Analysis of Special High Purity Acids Purified by Sub-Boiling Distillation, Anal. Chem., **44**, 2050–2056 (1972).

[23] Cotter, R. J., Yergey, A. L., Thermal Desorption of Quaternary Ammonium Cations, J. Am. Chem. Soc. **103**, 1596–1598 (1981).

JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 87, No. 1, January–February 1982

The Absolute Isotopic Abundance and Atomic Weight of a Reference Sample of Silver

L. J. Powell,*,† T. J. Murphy,* and J. W. Gramlich*

National Bureau of Standards, Washington, DC 20234

August 12, 1981

The atomic weight of a reference sample of silver has been determined by mass spectrometry, with an uncertainty of one part in 10^6, using a single filament silica gel procedure. Accurately known quantities of chemically pure ^{107}Ag and ^{109}Ag were mixed to produce standards of known isotopic composition for calibration of the mass spectrometer. The absolute isotopic ratio of the reference sample of silver is $^{107}\text{Ag}/^{109}\text{Ag} = 1.07638 \pm 0.00022$ yielding an atomic weight of 107.86815 ± 0.00011. The indicated uncertainties represent an overall limit of error at the 95 percent confidence level which is the sum of the uncertainty components for the ratio determined and the components covering effects of known sources of possible systematic error.

Key words: Absolute ratios; atomic weight; Faraday Constant; isotopic abundance; mass spectrometry; silica gel; silver; silver iodide.

1. Introduction

The inorganic mass spectrometry group of the National Bureau of Standards has been conducting a long term program of absolute isotopic abundance ratio and atomic weight determinations using high precision isotope ratio mass spectrometry. Previous atomic weight determinations include silver [1],[1] chlorine [2], copper [3], bromine [4], chromium [5], magnesium [6], lead [7], boron [8], rubidium [9], rhenium [10], silicon [11], potassium [12], thallium [13], and strontium [14].

The present work, a redetermination of the atomic weight of silver, was undertaken in conjunction with the calculation of a more accurate Faraday constant. The Faraday is directly related to other physical constants including the Avogadro constant, the proton gyromagnetic ratio, the magnetic moment of the proton in nuclear magnetons, and the ratio of the NBS as-maintained ampere to the absolute or SI ampere. Over the years a problem arose in assigning a recommended value for the Faraday due to apparent discrepancies between the Faraday determined from electrochemical experiments and the Faraday calculated from other fundamental constants. Unfortunately, earlier electrochemical experiments were not sufficiently precise (6.8 ppm, one standard deviation, for the best silver determination) to either prove or disprove the existence of this discrepancy. As a result, the Faraday was excluded from the most recent (1973) least squares adjustment of the fundamental constants [28].

In 1980, Bower and Davis [15], using the same source of silver as was analyzed in this work, NBS Standard Reference Material (SRM) 748, published a new value for the electrochemical equivalent of silver with an accuracy of 1.28 ppm (one standard deviation). The calculation of the Faraday using this value and the 1962 atomic weight of silver [1], indicated that significant differences between the various Faraday calculations might indeed exist. In order to assess the significance of these differences, however, it became necessary to significantly reduce the uncertainty in the atomic weight of silver.

The method used for the determination of atomic weights at NBS, which has been briefly described in previous publications [1–14], and may be described as calibrated mass spectrometry, combines the techniques of high precision chemical assay with high precision mass spectrometry. The mass spectrometers to be used for the isotopic abundance measurements are calibrated for bias using synthetic mixes of known isotopic composition, prepared from nearly pure separated isotopes. Extensive research has demonstrated that this bias is due to non-linearities in the measurement circuit and mass-dependent isotopic fractionation [16]. In the case of ratios near one, this bias reduces to isotopic fractionation. These measured biases are then used to calculate the absolute isotopic abundance ratio and, ultimately, the atomic weight of the reference sample. A block diagram characterizing the atomic weight method is shown in figure 1.

The chemical research requires the development of an assay procedure for the pure element or reference sample precise to at least one part in 10^4, which will ultimately be used to assay the separated isotope solutions. In addition, a

---

* Center for Analytical Chemistry, National Measurement Laboratory.

† Previous publications of this author have been under the name of L. P. Dunstan.

[1] Figures in brackets indicate literature references at the end of this paper.

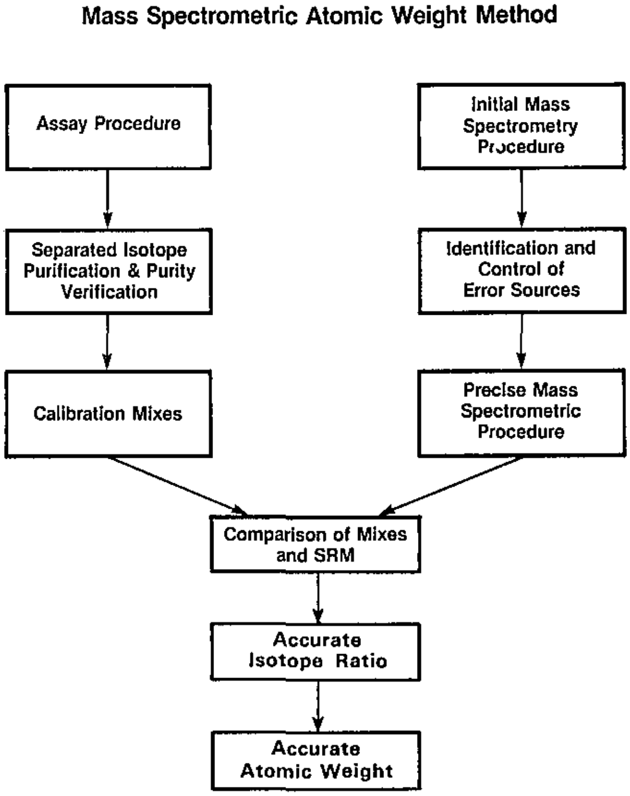

FIGURE 1.

procedure must be developed to purify the separated isotopes to eliminate impurities which could interfere with either the chemical or mass spectrometric procedures. Once these methods have been thoroughly tested using the reference sample, the separated isotopes are purified and dissolved. Aliquots of each isotope are then taken to produce calibration mixes which are blended to bracket the natural isotopic abundance ratio of the element. In addition, aliquots of each isotope are taken for assay so that accurate solution concentrations can be ascertained. A knowledge of the solution concentrations combined with aliquoting data allows the actual ratios of the calibration mixes to be calculated.

The mass spectrometric research involves the development of a high precision method for the analysis of the relative isotopic abundance ratios of the reference sample, followed by a comprehensive study to identify and control any sources of bias which could affect the final measurements.

After the above tasks have been accomplished, two complete sets of analyses of the calibration mixes and the SRM are made by two different operators using different mass spectrometers. The calibration mixes are randomly selected and are analyzed in an alternating pattern with the reference sample. A comparison of the relative isotopic abundance ratios obtained for the calibration mixes with the calculated or true value produces a calibration factor. This factor is applied to the relative isotopic ratio obtained for the SRM to produce the absolute isotopic abundance ratio. The atomic weight is then obtained by multiplying the fractional abundance of each isotope by its nuclidic mass[2] and summing the resultant products.

2. Experimental Procedure[3]

2.1 Mass Spectrometry

The isotopic ratio measurements were made using two nearly identical 90°, 30 cm radius of curvature solid sample mass spectrometers equipped with a "Z" lens focusing source [18]. The collector was a deep bucket Faraday cup equipped with a 50 percent transmission grid shadowing a series of suppression grids [18–20]. The measuring circuit consisted of two vibrating reed electrometers (VRE) and a voltage-to-frequency conversion system which transmits data to a computer for data acquisition. Prior to initiating the atomic weight ratio determinations, the measuring circuits of the mass spectrometers were calibrated and found to be linear to within one part in 10^4 over a range of 30–80 percent of full scale for each VRE scale. However, the linearity of the measurement electronics becomes insignificant because of the equal atom composition of the standard and most of the mixes. Under these conditions any non-linearities in the ratio measurement will cancel, and any possible effects would be indicated by the endpoint calibration mixes.

The relative isotopic composition of the calibration mixes and the reference sample were determined by surface ionization mass spectrometry using a platinum single filament silica gel technique.

The filaments used in this work were fabricated from 99.9 percent pure platinum ribbon (0.025 × 0.76 mm) and were cleaned by heating at 2.5 A for 1 h(r) in a vacuum and under a potential field.

The sample mounting was carried out in two stages, hereafter referred to as the low temperature drying stage and the high temperature drying stage. The low temperature drying stage was similar to the lead-silica gel method described by Barnes *et al.*, [21]. A 5 µL drop of silica gel suspension was placed on the filament surface and dried at 1.0 A for 5 min. A second drop of silica gel was added and the drying repeated. A 5 µL drop of silver as $AgNO_3$ in (1 + 9)[4] HNO_3 was placed

---

[2] Nuclidic masses, as published in Wapstra and Bos [17] are known to parts per billion accuracies.

[3] In order to describe adequately materials, instruments, equipment, and procedures, it was occasionally necessary to identify commercial products by manufacturer's name or label. In no instance does such identification imply endorsement by the National Bureau of Standards nor does it imply that the particular product or equipment is necessarily the best available for that purpose.

[4] A reagent dilution of (1 + 9) indicates 1 volume of concentrated reagent diluted with 9 volumes of pure water. If no dilution is specified, use of the concentrated reagent is implied. The acids and water used for these dilutions were produced at NBS by sub-boiling distillation [23].

on top of the silica gel and dried at 1.0 A for 5 min. A drop of 0.75 N H_3PO_4 was then added and dried at 1.5 A for 5 min then at 2.0 A for 5 min. This drying procedure was carried out in a class 100 clean air hood with an airflow of 100 linear ft/s per s using a programmable sample drier designed by Gramlich and Shideler [22]. Throughout the procedure a heat lamp, whose intensity was adjusted to yield a temperature of 70 °C at the filament surface, was used to aid the drying process.

The high temperature drying stage was accomplished using pyrometer adjustment of the filament temperature. This method was first used for the determination of the atomic weight of thallium [13] and has since proved to be an important key to controlling the reproducibility of isotopic ratios of a number of elements. The filament was covered with a bell jar and purged with nitrogen gas for 5 min. The nitrogen flow was stopped and the filament was adjusted to a temperature of 1040 °C for 60 s.

After loading the sample into the source of the mass spectrometer, the system was allowed to pump down to a pressure of at least 2×10^{-6} torr (300 μPa) before starting the analysis. Liquid nitrogen was then added to the source cold finger to reduce the pressure to less than 1×10^{-7} torr (10 μPa). At $t = 0$ min, the temperature of the filament was increased to 760 °C. This yielded a ^{107}Ag ion beam intensity $1 - 2 \times 10^{-12}$ A. At $t = 25$ min, the temperature of the filament was increased to 790 °C. At this temperature the intensity of the ^{107}Ag ion beam was between 5×10^{-12} A and 1×10^{-11} A. If the intensity of the ^{107}Ag ion beam was outside of these limits, the analysis was rejected. Baseline data were taken and ratio data were collected between $t = 30$ and 50 min.

2.2 Purification of Separated Silver Samples

Electromagnetically separated ^{107}Ag and ^{109}Ag isotopes in the form of silver metal powder were obtained from the Nuclear Division, Oak Ridge National Laboratory of the Union Carbide Nuclear Company. The ^{107}Ag was designated series R&D, sample 000101 and the ^{109}Ag was designated series R&D, sample 000201. The certificate which accompanied each sample showed enrichment to better than 99.9 percent for the major isotope. No information on chemical purity was given.

Since the method developed for the assay of silver was based on the precipitation of AgI from ammonical solution, it was necessary to develop a purification procedure that would reduce possible impurities to a level low enough so they would not cause a significant error.

Elements that form relatively insoluble iodides or hydroxides could possibly interfere. Included are Pb, and Tl which form insoluble iodides, and Al, Cr, Fe, Ga, In, Ti, Zr, and the rare earths which form insoluble hydroxides.

Each separated silver isotope was purified as follows: The silver isotope, as powdered silver metal (about 1g) was transferred to a 150 mL Teflon-FEP beaker, dissolved in 20 mL of (1 + 4) HNO_3, and the resulting solution evaporated to dryness. The residue was taken up in 50 mL of water and enough concentrated NH_4OH solution to dissolve the precipitated AgOH. The solution was digested on a warm hot plate for about one hour, allowed to cool, and filtered through a close textured paper. After washing the filter paper with dilute (1 + 49) NH_4OH the filtrate was heated to drive most of the NH_3 from the solution. The solution was made acidic with dilute (1 + 9) HNO_3 and 25 mL of (1 + 9) HCl was added to precipitate silver chloride. The solution and precipitate were digested on a hot plate for three hours in darkness and allowed to stand overnight at room temperature, also in darkness. Most of the supernatant slution was removed from the precipitated silver chloride by decantation and the last 10–15 mL was withdrawn by means of a plastic syringe equipped with a platinum needle. The precipitated AgCl was washed three times with a few mL of H_20 and the washings were removed with the syringe and needle.

The precipitated AgCl was then dissolved in 15 mL of NH_4OH and diluted with about 50 mL of water. The solution was heated to drive off excess NH_3 and made acid with (1 + 9) HNO_3 to reprecipitate AgCl. A few drops of (1 + 9) HCl were added and the solution was allowed to stand overnight in darkness. The supernatant liquid was then removed in the manner previously described. The precipitated AgCl was then dissolved in 20 mL of NH_4OH and transferred to a 350 mL platinum dish. A solution prepared by reacting 25 mL of (1 + 4) HNO_3 with 50 mL of (1 + 4) NH_4OH yielding NH_4NO_3, was added as an electrolyte and the resulting solution was diluted to approximately 250 mL. The platinum dish was placed on aluminum foil which was connected to the cathode of a dc-power source. The dish was covered with a Teflon-TFE cover fitted with a platinum rod through the middle which reached to within 1 cm of the dish bottom. This rod was connected to the anode of the dc-power source. Silver was then electrodeposited onto the dish by plating with a potential of 1.0 V for 24 h.

The electrolyte was then poured from the dish into a Teflon beaker and the electrodeposited silver in the dish was washed with dilute (1 + 9) NH_4OH.

The electrodeposited silver was dissolved with (1 + 4) HNO_3 diluted to about 150 mL and the solution was made ammonical with concentrated NH_4OH. Additional NH_4NO_3 electrolyte was added and the silver was again electrodeposited by plating at 1.0 V for 24 h.

The electrolyte was then poured into the Teflon beaker, and the silver was washed with water. The dish and contents were then dried and weighed. The electrodeposited silver was then dissolved with (1 + 4) HNO_3 and transferred to a Teflon beaker. The dish was washed with water, dried, and weighed to obtain the approximate weight of purified silver.

Calculations based on the starting weight of silver, and the weight of purified silver showed that more than 99 percent of the silver was recovered in each case.

The effectiveness of this purification procedure was first tested by purifying a solution of natural silver which had been doped at the 1000 μg/g level with each of thirty common impurity elements. The results of the analysis of this purified silver are shown in table 1. All of the impurity elements were reduced to 10 μg/g or less (the analytical method will also be described later).

Since the method was effective in purifying grossly contaminated silver, it was applied to each of the silver separated isotopes, which were of much higher initial purity.

The acids and water used in these purifications were produced at NBS by sub-boiling distillation [23] and have been shown to be extremely low in trace metal contamination. The NH_4OH was produced by saturating high purity cold water with high-purity NH_3 gas and allowing the resulting solution to warm to room temperature. Apparatus such as beakers and filters were cleaned with ACS Reagent Grade acids and rinsed with high purity water before use.

TABLE 1. *Analysis of impurities in silver separated isotopes.*

Element	Spike Isotope	Ag Doped[a] and Purified (μg/g)	^{107}Ag (μg/g)	^{109}Ag (μg/g)
Al	^{26}Mg	4	<14	<20
As	^{82}Se	0.3	—[b]	—
Ba	^{137}Ba	0.1	3.1	0.9
Bi	^{203}Tl	0.05	—	—
Ca	^{44}Ca	2.6	1.5	1
Cd	^{111}Cd	0.4	0.3	1.2
Co	^{62}Ni	0.5	—	—
Cr	^{53}Cr	0.2	0.2	0.1
Cu	^{65}Cu	1.2	2.2	0.7
Fe	^{54}Fe	5	6	6
Ga	^{71}Ga	0.2	<1.1	<1.6
K	^{41}K	2.7	5	1.2
Mg	^{26}Mg	1.8	0.8	1.2
Mn	^{54}Fe	0.03	—	—
Mo	^{97}Mo	0.01	0.06	0.06
Na	^{41}K	9	<50	<60
Nd	^{145}Nd	0.2	0.6	0.4
Ni	^{62}Ni	0.4	2	0.4
Pb	^{206}Pb	1.6	1.5	1.6
Pd	^{110}Pd	0.2	0.8	0.2
Sb	^{117}Sn	0.6	—	—
Se	^{82}Se	0.2	0.3	0.4
Sn	^{117}Sn	0.07	0.3	0.4
Sr	^{86}Sr	0.03	0.1	0.1
Te	^{125}Te	0.002	0.1	0.07
Tl	^{203}Tl	0.1	0.5	0.5
V	^{53}Cr	0.03	—	—
Zn	^{67}Zn	1	0.3	0.6
Zr	^{91}Zr	0.06	0.5	0.6

[a] Original concentration—1000 μg/g each.

[b] — denotes data not reported.

2.3 Preparation and Analysis of Separated Isotope Solutions

Each isotope (^{107}Ag and ^{109}Ag) was transferred to a specially constructed and tared 500 mL quartz flask. These flasks were constructed from a standard taper joint 500 mL quartz flask by cutting the neck of the flask about 1 cm from the body and tooling the neck to fit a number zero polyethylene stopper and a standard aluminum serum cap. Final solutions were diluted to about 350 mL and final acidity was about 0.5 N HNO_3. The solution of ^{107}Ag was labeled "Ag-107" and the solution of ^{109}Ag was labeled "Ag-109".

Portions of each separated isotope solution, equivalent to about 40 mg of silver, were taken for determination of impurities. Each sample was spiked with 10^{-7} g of ^{44}Ca, ^{111}Cd, ^{53}Cr, ^{65}Cu, ^{54}Fe, ^{71}Ga, ^{26}Mg, ^{97}Mo, ^{145}Nd, ^{62}Ni, ^{206}Pb, ^{110}Pd, ^{82}Se, ^{117}Sn, ^{86}Sr, ^{125}Te, ^{203}Tl, ^{67}Zn, and ^{91}Zr. Each spiked solution was then treated as follows: the solution was diluted to about 40 mL with H_2O and a slight excess of HCl was added to precipitate AgCl. The beaker and contents were allowed to stand in the dark overnight. Most of the supernatant liquid was removed from the precipitated AgCl by decantation into a second Teflon beaker. The last approximately 10 mL was removed by means of a plastic syringe and platinum needle and added to the decanted solution. The precipitate was washed with a few mL of (1 + 9) HNO_3 which was also combined with the decanted solution by use of the syringe. This solution was then evaporated to a few mL, removed from any remaining AgCl with the syringe and transferred to another beaker. This solution was evaporated to a large drop and transferred to specially prepared gold wire electrodes for analysis by spark source mass spectrometry using a peak switching electronic detection system.

The results of the anslysis of the purified separated silver isotopes are shown in table 1. In addition to the spiked elements, the concentrations of other elements were estimated by comparison to spiked nuclides using the assumption that the sensitivity of the natural element and spike nuclides are the same. The results show that most elements are at the low ppm level or less. The only elements found at concentrations of greater than 10 ppm were Al and Na. These contaminants may have resulted from the Pyrex glass flask used to store the spike isotopes. Even if they were present in the amounts indicated, the effect would be negligible since sodium would not interfere with the analysis at this level, and aluminum, as $Al(OH)_3$, would cause an error of only 0.001 percent in the assay of each isotope.

2.4 Assay of the Separated Isotope Solutions

The method developed for the high-precision assay of the silver solutions was based on a combination of gravimetry for the determination of most of the silver (greater than 99.5

percent) as AgI and isotope dilution mass spectrometry (IDMS) for the remainder. Silver iodide was selected as the gravimetric form since, of all the silver halides it is the least light sensitive, least soluble, and has the highest gravimetric factor. In fact, it was found that dry stoichiometric silver iodide was not detectably sensitive to fluorescent lighting since the material remained lemon yellow and no change in weight could be detected on prolonged exposure. The material was also found to be non-hygroscopic up to relative humidities of 90 percent.

The "Ag-107" solution and the "Ag-109" solution were each sampled for assay of silver in the following manner: four weighed portions of about 35 g, containing about 0.25 mmol Ag were withdrawn from each flask. A 10 cm platinum needle which had been inserted through a No. 0 polyethylene stopper was used to replace the stopper on the quartz flask. A 20 mL polyethylene hypodermic syringe, the plunger of which had been covered with a sheet of Teflon-FEP, was attached to the Kel-F hub of the needle and the desired amount of solution was withdrawn. The syringe was then disconnected from the hub and the tip was capped with a Kel-F cap. The syringe and contents were then weighed on a semi-micro balance to ±0.02 mg. (Any static charge on the syringe was dissipated by placing the syringe on the balance pan which was surrounded by several polonium anti-static sources and waiting 2 min before taking a weight.) The solution was then delivered to a 100 mL Pyrex beaker with a Teflon coated rim and the syringe was again capped and weighed. The weight of the sample taken was determined from the weight of the syringe before and after delivery of the sample. A second weighed portion was taken in the same manner and combined with the first portion for each assay sample. Two assay samples were withdrawn from each solution before and after withdrawing the calibration samples to ensure that no change in concentration had occurred during the time interval of withdrawing samples (about 6h).

Each sample was then assayed as follows: 20 mL of concentrated high-purity NH_4OH was added and approximately a 1 percent excess of 0.5 N NH_4I was added slowly from a small wash bottle. The solution was heated covered on a hot plate.

Precipitation from ammonical solution resulted in the formation of a white precipitate, $AgI \cdot 1/2\ NH_3$ [24]. Heating the solution drives off ammonia and results in the conversion of this white compound into yellow AgI. The AgI formed in this manner is composed of relatively large, easily filterable crystals that do not pass into the colloidal state when washed with water. The fact that AgI is slightly soluble in dilute NH_4OH also contributes to the formation of larger crystals.

After digesting overnight, the cover was removed and the solution was evaporated to a volume of 20 mL. Removal of NH_3 was tested by holding moist p-Hydrion paper over the hot solutions. If the paper indicated NH_3 in the vapor the solution was heated until NH_3 was no longer detected. The beaker was then removed from the hot plate, covered, and allowed to stand overnight.

The solution was then filtered through a tared platinum-Munroe crucible and the precipitated AgI was caught on the platinum mat. The beaker and precipitate were washed with several small increments of water. The filtrate and washing were transferred back to the original beaker and reserved for the determination of dissolved and untransferred silver. The crucible and contents were dried for 3 h at 200 °C. Pure AgI is only very slightly reduced by light when wet. To guard against any possible decomposition, the precipitation, filtration, and drying were carried out in orange light or darkness. (The pure compound is lemon yellow and even slight decomposition causes the AgI to darken.)

After heating, the crucible was cooled in a desiccator and transferred to the case of a micro-balance and allowed to stand for at least 2 h. The crucible and contents were then weighed on a micro-balance to ±0.002 mg. An effective buoyancy and adsorption correction for the platinum crucible was made by averaging three empty tare crucibles. (The drying, cooling, and weighing were repeated to ensure constant weight.) The apparent mass of the AgI was then determined and converted to true mass using a $5.66\ g \cdot cm^{-3}$ for the density of ^{107}AgI and $5.71\ g \cdot cm^{-3}$ for the density of ^{109}AgI. These densities were calculated by assuming that they are proportional to the density of natural αAgI, $D^{30} = 5.68\ g \cdot cm^{-3}$, in the same ratio as their molecular weights. The vacuum weight of the silver iodide was converted to millimoles of silver using the calculated atomic weight for silver and 120.9045 for iodine. The formula weights used were 233.8909 for ^{107}AgI and 235.8087 for ^{109}AgI.

To determine the soluble and untransferred silver, the filtrate and washings, which had been reserved in the original beaker, were spiked with 1 to 2 µmol of either ^{107}Ag or ^{109}Ag. Five mL of ammonium hydroxide solution and 0.25 g of potassium cyanide (ACS Reagent Grade) were added to dissolve silver iodide and the solution was well mixed by stirring with a magnetic stirring bar for a few hours. Silver was then separated electrolytically by plating overnight with a potential of 2.6 V onto a platinum wire anode. Silver was then dissolved from the electrode with a few mL of (1 + 1) HNO_3 and the resulting solution was evaporated to dyrness. The residue was taken up in (1 + 49) HNO_3 and the 107/109 ratio was determined by surface ionization mass spectrometry. The amount of silver as µmol of Ag was then calculated and added to the silver from the gravimetric determination to yield the total silver in the sample.

This method of assaying silver solutions was first tested on solutions containing known amounts of silver. The material was SRM 748 which is 99.999+ percent silver. The solutions were prepared in the approximate concentration range of 0.10 to 0.16 mmol AgI. Four portions containing from 0.9 to 1.1

mmol Ag were withdrawn from each solution and silver was determined as described above. Nine sets of four samples were analyzed in this manner.

Statistical analysis of the data from these nine sets showed that the standard error of the mean of four values was 5.2×10^{-6} mmol/g and that the coefficient of variation (standard error/mean) for a set of four was 0.003 percent. Comparison of calculated to measured concentrations showed a small positive bias of about 0.005 percent, but this would have a negligible effect on the ratio of two assays.

The results of the assay of the separated isotope solutions are shown in table 2. Pooling the results of the assay of the separated isotope solutions and the nine sets of natural solutions, yields a value of 0.0000009 mmol Ag/g for the standard error of the average concentration of each separated isotope solution. The uncertainty associated with this value at the 95 percent confidence level is 1.8×10^{-6} mmol Ag/g.

2.5 Isotopic Analyses of the Separated Isotope Solutions

Each of the separated isotope solutions was analyzed eight times on each of two mass spectrometers, MS #1 and MS #5, by Operators 1 and 2, respectively. The mass spectrometer sources were cleaned between the analyses of the ^{107}Ag and ^{109}Ag as a precaution against the possibility of contamination from source parts, although back-to-back analyses of the two separated isotopes on the same source failed to yield any evidence of contamination. The corrected isotopic compositions of the two isotopes are shown in table 3.

The measured imprecision of the ratio measurements for each of the separated isotope solutions was less than 3 percent (2 standard deviations), however an uncertainty of 7 percent was placed on the ratios for the separated isotopes to cover possible measurement system nonlinearities and isobaric interferences. Several possible species have a potential for producing isobaric interferences in the silver mass region. These include $CaPO_2^+$, $SiPO_3^+$, and AsS^+. Any interferences from $CaPO_2^+$ can be detected by monitoring mass 103 which results from the major isotope of calcium (^{40}Ca : 96.9 percent abundant). However, the major isotopes of silicon and sulfur would produce $SiPO_3^+$ and AsS^+ interferences at mass 107. Thus extremely small quantities of either of these species could not be detected by examining other masses in the silver spectral region. A $SiPO_3^+$ species has not been reported or observed in a thermal ionization-silica gel analysis but is included herein because of the nearly infinite supply of these elements on the filament, and the fact that an intense and

TABLE 2. *Concentration of silver separated isotope solution*

Solution	Sample No.	Weight Ag I (g)	Ag from Ag I (mmol)	Ag from Filtrate (mmol)	Total Ag (mmol)	Weight Sample (g)	Concentration (mmol Ag/g)
"Ag 107"	1	0.223445	0.955668	0.002583	0.958251	34.00905	0.0281760
	2	0.231677	0.990876	0.002358	0.993234	35.25092	0.0281761
	3	0.234055	1.001047	0.001176	1.002223	35.56893	0.0281769
	4	0.228265	0.976283	0.001211	0.977493	34.69020	0.0281778
						Average	0.0281767
"Ag 109"	1	0.229478	0.973153	0.003956	0.977109	35.78498	0.0273050
	2	0.228795	0.970257	0.002316	0.972573	35.62279	0.0273020
	3	0.226974	0.962534	0.001029	0.963563	35.28974	0.0273044
	4	0.231376	0.981202	0.001614	0.982814	35.99738	0.0273024
						Average	0.0273034

TABLE 3. *Isotopic composition of the silver separated isotopes*

Separated Isotope		*Isotopic Composition* (atom percent)
"Ag 107"	^{107}Ag	99.97136 ± 0.00206[a]
	^{109}Ag	0.02863 ± 0.00206
"Ag 109"	^{107}Ag	0.02865 ± 0.00203[a]
	^{109}Ag	99.97135 ± 0.00203

[a] The uncertainty of the ratio determination is taken to be 7 percent, which is much larger than the 95 percent confidence limit, to take into account possible biases and non-linear instrumental behavior for ratios as large as these.

sustained PO_3^- ion beam is observed in the negative ion spectra under the analytical conditions reported in this paper.

Kelly and Wasserburg [25] reported the existence of AsS^+ which, if present, would enhance the measured 107/109 ratio. The measurement of the ^{109}Ag separated isotope is particularly sensitive to this interference because AsS^+ would produce a 107/109 ratio of 22. Efforts to generate as AsS^+ ion beam by loading As_2S_3 and H_2SO_4 onto the filament using the procedure described in section 2 were unsuccessful. Since the silver was loaded in (1 + 9) HNO_3, it is felt that the nitric acid allowed oxidation of the sulfur to sulfate, thus reducing any AsS^+ to an undetectable level.

Analysis under identical conditions of ^{109}Ag samples ranging in amount from 4 μg to 0.005 μg indicated a possible mass-107 isobaric interference at a level of $1\ (\pm 1) \times 10^{-16}$ A. This level of interference is below the direct measurement detection limit of the instrument but can be inferred from the slight decrease in the observed 109/107 ratio with decreasing sample size. This suspected interference would produce a bias of approximately 2 percent to the observed ratios of the ^{109}Ag separated isotope and a bias of approximately 0.002 percent to the observed ratios of the reference sample and the calibration mixes.

2.6 Preparation of Calibration Samples

Eight calibration mixes were prepared by blending weighed portions of the "Ag-107" and the "Ag-109" solutions to produce 107/109 ratios ranging from 0.5 to 2.0. Six of the calibration mixes were within a few percent of the natural ratio 107/109 of 1.076.

The portions for mixing were withdrawn from the flasks and weighed in the manner previously described for the assay of each solution. Each portion wieghed about 10 g and was weighed to ±0.05 mg. It is therefore estimated that the weighing error for each mix should not exceed one part in 10^5. To minimize any significant possibility of change in concentration of the isotope solutions with time, the portions for the calibration mixes were withdrawn from the flasks between the samples taken for assay, over a period of about 6 h.

Each calibration mix was thoroughly mixed, the sides of the beaker were washed with water, and the solution was evaporated to dryness. The residue was taken up with dilute (1 + 9) HNO_3 and the mixing, washing, and evaporation were repeated. The calibration mixes were then taken up in (1 + 9) HNO_3 to a concentration of 0.8 mg Ag/mL and transferred to small Teflon bottles. The isotopic compositions of the calibration mixes are given in table 4.

2.7 Isotopic Analyses of the Calibration Mixes and the Standard Sample

Two complete sets of analyses of the calibration mixes and reference sample were performed by Operator 1 on MS #1 and Operator 2 on MS #5. Operator 1 performed three anal-

TABLE 4. *Isotopic composition of calibration mixes*

Solution No.	Isotope Solution	Weight Solution (g)	Ag from Solution (mmol)	^{107}Ag from Solution (mmol)	^{109}Ag from Solution (mmol)	Total ^{107}Ag (mmol)	Total ^{109}Ag (mmol)	Ratio 107/109
1	"Ag 107"	10.02135	0.282369	0.282288	0.000081	0.282364	0.263439	0.07184
	"Ag 109"	9.64836	0.263433	0.000076	0.263358			
2	"Ag 107"	8.95832	0.252416	0.252344	0.000072	0.252412	0.237226	1.06402
	"Ag 109"	8.68835	0.237222	0.000068	0.237154			
3	"Ag 107"	9.24458	0.260482	0.260407	0.000075	0.260477	0.243049	1.07171
	"Ag 109"	8.90162	0.243044	0.000070	0.242974			
4	"Ag 107"	9.75636	0.274902	0.274823	0.000079	0.274897	0.257233	0.06867
	"Ag 109"	9.42110	0.257228	0.000074	0.257154			
5	"Ag 107"	9.59357	0.270315	0.270237	0.000078	0.270309	0.248906	1.08599
	"Ag 109"	9.11603	0.248899	0.000071	0.248828			
6	"Ag 107"	10.11275	0.284944	0.284862	0.000082	0.284940	0.269592	1.05693
	"Ag 109"	9.87378	0.269588	0.000078	0.269511			
7	"Ag 107"	10.01824	0.282281	0.282200	0.000081	0.282348	0.517264	0.545849
	"Ag 109"	18.94756	0.517333	0.000148	0.517185			
8	"Ag 107"	18.47706	0.520623	0.520474	0.000149	0.520552	0.270624	1.92353
	"Ag 109"	9.90908	0.270552	0.000078	0.270475			

yses of each calibration mix and 24 analyses of the reference sample. Operator 2 performed four analyses of each calibration mix and 32 analyses of the reference sample. The samples were run in a pattern alternating randomly selected mixes with the reference sample.

3. Results and Discussion

The results of the measurement of the eight calibration mixes are shown in table 5. The calibration factors for each analyst varied over a range of 0.011 percent for Operator 1 and 0.019 percent for Operator 2. In addition, the calibration factors for those mixes with a 107/109 = 2 and 107/109 = 0.5 were indistinguishable from those bracketing the reference sample, indicating an insignificant degree of nonlinearity over the measured range.

Table 6 summarizes the observed and corrected $^{107}Ag/^{109}Ag$ values for the SRM for Operators 1 and 2 as well as the absolute isotopic abundance ratio for silver and its uncertainty.

Table 7 gives summary calculations of the reference sample. The atomic weight is calculated from the absolute isotopic abundance by summing the product of the nuclidic masses obtained from Wapstra and Bos [17] and the corresponding atom fractions.

The atomic weight of silver, 107.868 ± 0.001, as recommended by the IUPAC Commission on Atomic Weights, is based on the isotopic abundance measurements reported by Shields, Garner, and Dibeler [1] in 1962. This ratio is, in fact, the average of their isotopic abundance measurements published in 1960 [26] and the results of the mineral survey described in the 1962 publication.

A reanalysis of sample C-140, a nugget of native silver from Ontario, Canada, which had yielded anomalous data in the 1962 study [1], was undertaken. One section of this nugget, referred to as position 1, had yielded a ratio of $^{107}Ag/^{109}Ag$ which was 0.3 percent higher than three other sections, designated positions 2, 3, and 4; all of which mirrored the $^{107}Ag/^{109}Ag$ ratio of the natural silver reference sample (SRM 948). At the time, little or no chemistry was performed to purify the samples prior to analysis, and it was believed that the "anomaly" might be due to natural differences or chemical impurities. A reanalysis of the four sections of C-140, confirmed that position 1 yielded a $^{107}Ag/^{109}Ag$ ratio that was higher than the ratios for positions 2, 3, and 4, which again yielded ratios indistinguishable from the natural silver reference standard. Subsequent analyses indicated the major impurity at position 1 was mercury. Following purification by electrodeposition and ion exchange chromatography, the $^{107}Ag/^{109}Ag$ ratio for position 1 matched the ratios obtained for

TABLE 5. *Determination of correction factors*

Calibration Sample No.	Isotopic Ratio, $^{107}Ag/^{109}Ag$			Correction Factors	
	Calculated	Operator 1	Operator 2	Operator 1	Operator 2
1	1.071840	1.081566	1.081420	0.99100367	0.99114511
2	1.064016	1.073622	1.073579	0.99104924	0.99109658
3	1.071705	1.081345	1.081214	0.99108175	0.99120936
4	1.068671	1.078304	1.078236	0.99106309	0.99113312
5	1.085996	1.095850	1.095681	0.99100465	0.99116505
6	1.056931	1.066516	1.066407	0.99100878	0.99111770
7	0.5458493	0.5507895	0.5506505	0.99102567	0.99128580
8	1.923532	1.941050	1.940748	0.99097136	0.99113254
Mean Values of Calibration Factors				0.9910261	0.9911607

TABLE 6. *Determination of corrected isotopic ratios*

	Observed $^{107}Ag/^{109}Ag$	Correction Factor	Corrected $^{107}Ag/^{109}Ag$
Operator 1	1.0861089	0.9910261	1.076362
Operator 2	1.0859898	0.9911607	1.076390
		MEAN	1.07638
		±	0.00022

Uncertainty components:

95 percent confidence limits in ratio determination	±0.00003
Bounds due to possible systematic error in composition of separated isotopes	±0.00009
Bounds due to possible systematic error in chemical analysis.	±0.00010

TABLE 7. *Summary calculations of the atomic weight of silver*

Values	Uncertainty Components			
	Overall Limit of Error[a]	Mass Spectrometric Analytical Error	Possible Systematic Error in Composition of Separated Isotopes	Possible Systematic Error in Chemical Analysis
Atomic Weight = 107.86815	±0.00011[b]	±0.0000137	±0.0000408	±0.0000466
Nuclidic Masses (^{12}C = 12)				
^{107}Ag = 106.905095	±0.000012			
^{109}Ag = 108.904754	±0.000010			
Atom Percent				
^{107}Ag = 51.839170	±0.005057	±0.000687	±0.00204	±0.00233
^{109}Ag = 48.160830	±0.005057	±0.000687	±0.00204	±0.00233
Isotopic Ratio ^{107}Ag/^{109}Ag = 1.07638	±0.00022	±0.0000296	±0.0000880	±0.0001006

[a] The overall limit of error is the sum of the 95 percent confidence limits and the terms covering effects of known sources of possible systematic error.

[b] Includes a component for uncertainty in nuclidic masses.

positions 2, 3, and 4, and thus, the reference standard. This experiment clearly documents the critical importance of chemical purification in the determination of accurate isotopic ratios and in establishing the limits for natural isotopic variations. The data reflecting the effects of purification on sample C-140 are shown in table 8.

TABLE 8. *Effect of Purification on C-140*

Chemical Procedure	Corrected[a] ^{107}Ag/^{109}Ag
Initial Electrodeposition	1.07686
Additional Electrodeposition	1.07676
Anion Exchange, Precipitation and Electrodeposition	1.07634

[a] All ratios have been placed on an absolute basis using the absolute isotopic abundance ratio for silver from Table 7 (^{107}Ag/^{109}Ag = 1.07638).

In an attempt to establish the limits of isotopic variability among commercially available silver, a series of analyses were performed on silver metal samples of varying purity, which had been obtained from refineries throughout the world. The results are shown in table 9. With the exception of the marz grade silver from Materials Research Corporation, all samples yielded isotopic ratios which were in excellent agreement with the natural reference standard. Although a deviation of 7.6 parts in 10^4 is small, it is not known whether the difference is real or an artifact due to a chemical impurity. Additional testing of this material will be done in conjunction with a more thorough mineral survey in the future.

The effect of this isotopic survey of commercial silver on the atomic weight is small. A recalculation of the atomic weight of silver based on the isotope ratios of these materials yields a value of 107.86816 ± 0.00032 at the 95 percent confidence level, which is nearly identical to the atomic weight of the reference sample (107.86815) as given in table 7, the major difference existing in the uncertainty.

The reference sample used in this atomic weight determination will be issued as both an isotopic and assay Standard Reference Material (SRM) and will be designated as SRM 978a. Its isotopic analysis was found to be experimentally

TABLE 9. *Isotope Ratios of Commercial Silver*

Description[a]	Corrected[b] ^{107}Ag/^{109}Ag
Needle Ag, 99.995%, Engelhard, Indiana, USA	1.07650
$AgNO_3$, 99.999%, Engelhard, Indiana, USA	1.07633
Ag, Marz Grade, Materials Research Corp., USA	1.07557
Ag, Single Crystal, Marz Grade, USA	1.07622
Refined Ag, Port Pirie, Australia	1.07634
Ag, Normal Purity, Johnson Mathey, Canada	1.07640
Ag, High Purity, Cominco, TADANAC, Canada	1.07648
Refined Ag, Mixture of Swedish Ores	1.07646
Electrolytic Ag, Boliden Mine, Sweden	1.07670
Ag, Crystalline, Mexico	1.07638
Ag, 99.99+%, Lima, Peru	1.07653
Ag, Plata Fina, Granalla, Mexico	1.07656
Average	1.07637
Standard Deviation	0.000282

[a] All descriptions are as stated by the producer. No further purity verification was made.

[b] All ratios have been placed on an absolute basis using the absolute isotopic abundance ratio for silver from table 7 (^{107}Ag/^{109}Ag = 1.07638).

identical to SRM 978, the major difference existing in the uncertainty placed on the measurement.

The redetermination of the atomic weight of silver has allowed the recalculation of the Faraday from the electrochemical equivalent of silver as determined by Bower and Davis [5], yielding a value of $96486.18 \pm 0.13\ A_{NBS}\cdot s\cdot mol^{-1}$ (one standard deviation). Further details of the calculation of this constant are given elsewhere in this issue of the Journal of Research [27].

The reduction of the overall uncertainty in the Faraday constant has allowed discernment of the discrepancies which existed between the Faraday determined from electrochemical experiments, and the Faraday calculated from other fundamental constants. It is expected that the new NBS value of the Faraday will be included in the 1981 CODATA least squares adjustment of the fundamental constants.

We are indebted to: Keith R. Eberhardt for statistical analysis; Paul J. Paulsen for analysis of the separated isotopes; Karen A. Brletic for isotope dilution analysis of many of the assay samples; William A. Bowman III for instrumental maintenance support; Joy J. Shoemaker and Theresa Sperow for skillful manuscript preparation; and our colleagues in the mass spectrometry group of the Inorganic Analytical Research Division for their helpful discussions.

We also wish to express our gratitude to Richard Davies, Executive Director of the Silver Institute for his assistance in gathering the high purity silver metal samples and to the following companies for their contributions: Boliden Metall AB, Sweden; The Broken Hill Associated Smelters Proprietory Limited, Australia; Met-Mex Peñales, S. A. de C. V., Mexico; Empresa Minera Del Centra Del Peru; Johnson-Matthey, Ltd., Canada.

We also wish to thank Oak Ridge National Laboratory for their special efforts to produce the high purity ^{107}Ag and ^{109}Ag isotopes which were so necessary for this work.

4. References

[1] Shields, W. R.; Garner, E. L.; Dibeler, V. H. Absolute Isotopic Abundance of Terrestrial Silver. J. Res. Nat. Bur. Stand. (U.S.). **66A**(1): 1–3; 1962 January–February.

[2] Shields, W. R.; Murphy, T. J.; Garner, E. L.; Dibeler, V. H. Absolute Isotopic Abundance Ratio and the Atomic Weight of Chlorine. J. Amer. Chem. Soc. **84**: 1519–1522; 1962 May 5.

[3] Shields, W. R.; Murphy, T. J.; Garner, E. L. J. Res. Nat. Bur. Stand. (U.S.). **68A**(6): 589–592; 1964 November–December.

[4] Catanzaro, E. J.; Murphy, T. J.; Garner, E. L.; Shields, W. R. Absolute Isotopic Abundances Ratio and the Atomic Weight of Bromine. J. Res. Nat. Bur. Stand. (U.S.). **68A**(6): 593–599; 1964 November–December.

[5] Shields, W. R.; Murphy, T. J.; Catanzaro, E. J.; Garner, E. L. Absolute Isotopic Abundance Ratio and the Atomic Weight of a Reference Sample of Chromium. J. Res. Nat. Bur. Stand. (U.S.). **70A**(2): 193–197; 1966 March–April.

[6] Catanzaro, E. J.; Murphy, T. J.; Garner, E. L.; Shields, W. R. Absolute Isotopic Abundance Ratios and Atomic Weight of Magnesium. J. Res. Nat. Bur. Stand. (U.S.). **70A**(6): 453–458; 1966 November–December.

[7] Catanzaro, E. J.; Murphy, T. J.; Shields, W. R.; Garner, E. L. Absolute Isotopic Abundance Ratios of Common, Equal-Atom, and Radiogenic Lead Isotopic Standards. J. Res. Nat. Bur. Stand. (U.S.). **72A**(3): 261–267; 1968 May–June.

[8] Catanzaro, E. J.; Champion, C. E.; Garner, E. L.; Marinenko, G.; Sappenfield, K. M.; Shields, W. R. Standard Reference Materials: Basic Acid; Isotopic, and Assay Standard Reference Materials. Nat Bur. Stand. (U.S.) Spec. Publ. 260-17; 1970 February. 60 p.

[9] Catanzaro, E. J.; Murphy, T. J.; Garner, E. L.; Shields, W. R. Absolute Isotopic Abundance Ratio and Atomic Weight of Terrestrial Rubidium. J. Res. Nat. Bur. Stand. (U.S.). **73A**(5): 511–516; 1969 September–October.

[10] Gramlich, J. W.; Murphy, T. J.; Garner, E. L.; Shields, W. R. Absolute Isotopic Abundance Ratio and Atomic Weight of a Reference Sample of Rhenium. J. Res. Nat. Bur. Stand. (U.S.). **77A**(6): 691–698; 1973 November–December.

[11] Barnes, I. L.; Moore, L. J.; Machlan, L. A.; Murphy, T. J.; Shields, W. R. Absolute Isotopic Abundance Ratios and the Atomic Weight of a Reference Sample of Silicon. J. Res. Nat. Bur. Stand. (U.S.). **79A**(6): 727–735; 1975 November–December.

[12] Garner, E. L.; Murphy, T. J.; Gramlich, J. W.; Paulsen, P. J.; Barnes, I. L. Absolute Isotopic Abundance Ratios and the Atomic Weight of a Reference Sample of Potassium. J. Res. Nat. Bur. Stand. (U.S.). **79A**(6): 713–725; 1975 November–December.

[13] Dunstan, L. P.; Gramlich, J. W.; Barnes, I. L.; Purdy, W. C. Absolute Isotopic Abundance and the Atomic Weight of a Reference Sample of Thallium. J. Res. Nat. Bur. Stand. (U.S.). **85**(1): 1–10; 1980 January–February.

[14] Moore, L. J.; Murphy, T. J.; Barnes, I. L. Absolute Isotopic Abundance Ratios and Atomic Weight of a Reference Sample of Strontium. J. Res. Nat. Bur. Stand. (U.S.). **87**(1): 1–8, 1982 January–February.

[15] Bower, V. E.; Davis, R. S. The Electrochemical Equivalent of Pure Silver—A Value of the Faraday. J. Res. Nat. Bur. Stand. (U.S.). **85**(3): 175–191; 1980 May–June.

[16] Garner, E. L.; Machlan, L. A.; Shields, W. R. Standard Reference Materials: Uranium Isotopic Standard Reference Materials. Nat. Bur. Stand. (U.S.) Spec. Publ. 260-27; 1971 April. 162 p.

[17] Wapstra, A. H.; Bos, K. Atomic Data and Nuclear Data Tables. **19**(3): 177–214; 1977 March.

[18] Shields, W. R., ed. Analytical Mass Spectrometry Section: Summary of Activities July 1966 to June 1967. Nat. Bur. Stand. (U.S.) Tech. Note 426; 1967 September. 53 p.

[19] Shields, W. R., ed. Analytical Mass Spectrometry Section: Instrumentation and Procedures for Isotopic Analysis. Nat. Bur. Stand. (U.S.) Tech. Note 277; 1966 July. 99 p.

[20] Garner, E. L.; Machlan, L. A.; Shields, W. R. Standard Reference Materials: Uranium Isotopic Standard Reference Materials. Nat. Bur. Stand. (U.S.) Spec. Publ. 260-27; 1971 April. 150 p.

[21] Barnes, I. L.; Murphy, T. J.; Gramlich, J. W.; Shields, W. R. Lead Separation by Anodic Deposition and Isotope Ratio Mass Spectrometry of Microgram and Smaller Samples. Anal. Chem. **45**(11): 1881–1884; 1973 September.

[22] Gramlich, J. W.; Shideler, R. W. A Programmable Sample Dryer for Thermal Ionization Mass Spectrometry. Nat. Bur. Stand. (U.S.) Tech. Note 1154, 1982.

[23] Kuehner, E. C.; Alvarez, R.; Paulsen, P. J.; Murphy, T. J. Production and Analysis of High Purity Acids Purified by Sub-boiling Distillation. Anal. Chem. **44**(12): 2050–2056; 1972 October.

[24] Longe, W. Uber die Verfärburg des Silberjodids durch wässriges Ammoniak. Z. Anarg u. Allgem. Chem. **223**: 174–176; 1935.

[25] Kelly, W. R.; Tera, F.; Wasserburg, G. J. Isotopic Determination of Silver on Picomole Quantities by Surface Ionization Mass Spectrometry. Anal. Chem. **50**(9): 1279–1286; 1978 August.

[26] Shields, W. R.; Craig, D. N.; Dibeler, V. H. Absolute Isotopic Abundance Ratio and the Atomic Weight of Silver. J. Amer. Chem. Soc. **82**: 5033–5036; 1960 October.

[27] Bower, V. E.; Davis, R. S.; Murphy, T. J.; Paulsen, P. J.; Gramlich, J. W.; Powell, L. J. A Redetermination of the Faraday Constant Due to a New Atomic Weight of Silver. J. Res. Nat. Bur. Stand. (U.S.). **87**(1): 21–23, 1982 January–February.

[28] Cohen, E. R.; Taylor, B. N. The 1973 Least Squares Adjustment of the Fundamental Constants. J. Phys. Chem. Ref. Data **2**(4): 663–734; 1973.

JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 87, No. 1, January–February 1982

Recalculation of the Faraday Constant Due to A New Value for the Atomic Weight of Silver

V. E. Bower,* R. S. Davis,* T. J. Murphy,† P. J. Paulsen,† J. W. Gramlich,† and L. J. Powell†,**

National Bureau of Standards, Washington, DC 20234

September 16, 1981

A report of the Faraday constant as determined at NBS via silver coulometry and atomic weight measurements is presented. The uncertainty of the reported result represents a five-fold improvement over measurements made at NBS 20 years ago. The result should contribute to an analysis of the self-consistency of several other fundamental constants measurements. Experimental details have been reported in other publications which are cited in the text.

Key words: Atomic weight; atomic weight of silver; coulometer; electrochemical equivalent; Faraday constant; fundamental constants; silver; silver coulometer.

The Faraday constant, F, may be derived from the following relation:

$$F = \frac{A_r(\mathrm{Ag})}{E_{\mathrm{Ag}}},$$

where E_{Ag} is the electrochemical equivalent of pure silver and $A_r(\mathrm{Ag})$ is the atomic weight of the silver used in determining E_{Ag}. In 1975, coulometric measurements of E_{Ag} undertaken at the National Bureau of Standards [1][1] achieved sufficiently high precision to warrant a careful re-evaluation of the purity of the silver used [2] as well as a more accurate redetermination of its atomic weight [3]. This last work, which is the previous paper published in this issue of the Journal of Research, completes our efforts to measure F via the silver coulometer. Our results are:

$$F_{\mathrm{NBS75}} = 96\,486.17(13)\ A_{\mathrm{NBS75}}\cdot\mathrm{s}\cdot\mathrm{mol}^{-1}\ (1.3\ \mathrm{ppm})$$
$$F_{\mathrm{BI69}} = 96\,486.06(13)\ A_{\mathrm{BI69}}\cdot\mathrm{s}\cdot\mathrm{mol}^{-1}\ (1.3\ \mathrm{ppm})$$
$$F_{\mathrm{SI}} = 96\,485.44(14)\ A_{\mathrm{SI}}\cdot\mathrm{s}\cdot\mathrm{mol}^{-1}\ (1.4\ \mathrm{ppm})$$

where the subscripts identify the electrical units used; that is,

NBS75—electrical units as maintained at the National Bureau of Standards in March of 1975, the time of the coulometric measurements.

BI69—a convenient reference set of "as maintained" electrical units defined in ref. [4]. (BI is an abbreviation for Bureau International des Poids et Mesures (BIPM), the International Bureau of Weights and Measures in Sèvres, France.) We use $A_{\mathrm{NBS75}}/A_{\mathrm{BI69}} = 1-(1.2\pm0.1) \times 10^{-6}$ [5].

SI—electrical units as defined in the Système International (i.e., absolute units). The transformation from NBS75 to SI amperes has been made by using what we believe to be the best available conversion factor, $A_{\mathrm{NBS75}}/A_{\mathrm{SI}} = 1-(7.6\pm0.5) \times 10^{-6}$ [5], although this is by no means a closed question [6,7].

All uncertainties, which have been given above in parentheses, are meant to correspond to one standard deviation. In addition, the draft recommendations of the BIPM [8] regarding the calculation of uncertainty have been followed, the most noteworthy being that all uncertainties are combined by taking the square root of the sum of their squares to produce a total uncertainty. These procedures for computing uncertainty were also adhered to in reporting the most recent measurements of the electrochemical equivalent of silver [1,2] as well as in recent compilations of the fundamental constants [4,6]. In reporting the new atomic weight of silver [3], however, a more conservative approach has been taken. That is, if the BIPM recommendations were followed, the total uncertainty (at a level of one standard deviation) in the atomic weight of silver reported in ref. [3] would shrink from 0.5 ppm to 0.3 ppm [9]. It is this latter uncertainty which we have used in the results reported above. The total uncertainty at a level of one standard deviation for F_{NBS75}, calculated

---

* Center for Absolute Physical Quantities, National Measurement Laboratory. V. E. Bower's present address: 9906 Pomona Drive, Bethesda, MD 20034.

† Center for Analytical Chemistry, National Measurement Laboratory.

** Previous publications of this author have been under the name of L. P. Dunstan.

[1] Figures in brackets indicate literature references at the end of this paper.

according to the BIPM recommendations, is shown in table 1. Thus the new measurement of the atomic weight of silver reported in ref. [3] has reduced the role of atomic weight in the uncertainty calculation of the Faraday experiment from dominance to insignificance.

TABLE 1. *Estimate of Uncertainties in F_{NBS75} from Known Sources*

Source	Uncertainty (ppm)
Electrochemical equivalent of silver	1.3ᵃ
Atomic weight of silver	0.3
Combined uncertainty	1.3

ᵃ Further details in Ref. [2].

Figure 1 shows the value of the Faraday calculated here as well as Faraday constants measured by other scientists. Point D is the present CODATA recommended value of Cohen and Taylor [4], which is calculated from other physical constants via least squares. It may be noted that not only has the overall uncertainty in the Faraday constant been reduced through this determination, but the new value of the Faraday is now essentially the same as that calculated from the proton gyromagnetic ratio determined by Kibble and Hunt [13]. No further work using silver coulometers is planned since it would be extremely difficult to reduce the uncertainty of the measurements below their present values.

References

[1] Bower, V.E.; Davis, R.S. A new determination of the Faraday by means of the silver coulometer, in *Atomic Masses and Fundamental Constants*, Vol. **5**. J.H. Sanders; A.H. Wapstra, eds. New York: Plenum Press; 1976. 578–583.

[2] Bower, V.E.; Davis, R.S. The electrochemical equivalent of pure silver—A value of the Faraday. J. Res. Nat. Bur. Stand. (U.S.). **85**(3): 175–191; 1980 May–June.

[3] Powell, L.J.; Murphy, T.J.; Gramlich, J.W. The absolute isotopic abundance and atomic weight of a reference sample of silver. J. Res. Nat. Bur. Stand. (U.S.). 87(1): 9–19, 1981 January–February.

[4] Cohen, E.R.; Taylor, B.N. The 1973 least-squares adjustment of the constants. J. Phys. Chem. Ref. Data **2**(4): 663–734; 1973.

[5] Taylor, B.N. private communication.

[6] Cohen, E.R. Status of the fundamental constants, in *Atomic Masses and Fundamental Constants*, Vol. **6**. J.A. Nolan, Jr.; W. Benenson, eds. New York: Plenum Press; 1980. 525–540. See especially p. 533.

[7] Taylor, B.N. Is the present realization of the absolute ampere in error? Metrologia **12**: 81–83; 1976.

[8] Giacomo, P.; News from the BIPM, Meeting of experts on the subject of expression of uncertainties, Draft recommendation on the statement of uncertainties. Metrologia **17**(2): 69–74; 1981.

[9] Eberhardt, Keith R. private communication.

[10] Craig, D.N.; Hoffman, J.I.; Law, C.A.; Hamer, W. J. Determination of a value of the Faraday with a silver perchloric acid coulometer. J. Res. Nat. Bur. Stand. (U.S.). **64A**(5): 381–402; 1960 September–October.

[11] Marinenko, G.; Taylor, J.K. Electrochemical Equivalents of Benzoic and Oxalic Acid. Anal. Chem. **40**(11): 1645–1651; 1968 September.

[12] Koch, W.F. The value of the Faraday via 4-aminopyridene, in *Atomic Masses and Fundamental Constants*, Vol. **6**. J.A. Nolen, Jr. and W. Benenson, eds. New York: Plenum Press; 1980, 167–172.

[13] Kibble, B.P.; Hunt, G.J. A Measurement of the Gyromagnetic Ratio of the Proton in a Strong Magnetic Field. Metrologia **15**: 5–30; 1979.

[14] Deslattes, R.D.; Henins, A.; Bowman, H.A.; Schoonover, R.M.; Carroll, C.L. and Barnes, I.L.; Machlan, L.A.; Moore, L.J.; Shields, W.R. Determination of the Avogadro Constant. Phys. Rev. Lett. **33**(8): 463–466; 1974 August. We have used the authors' most current recalculation of their result.

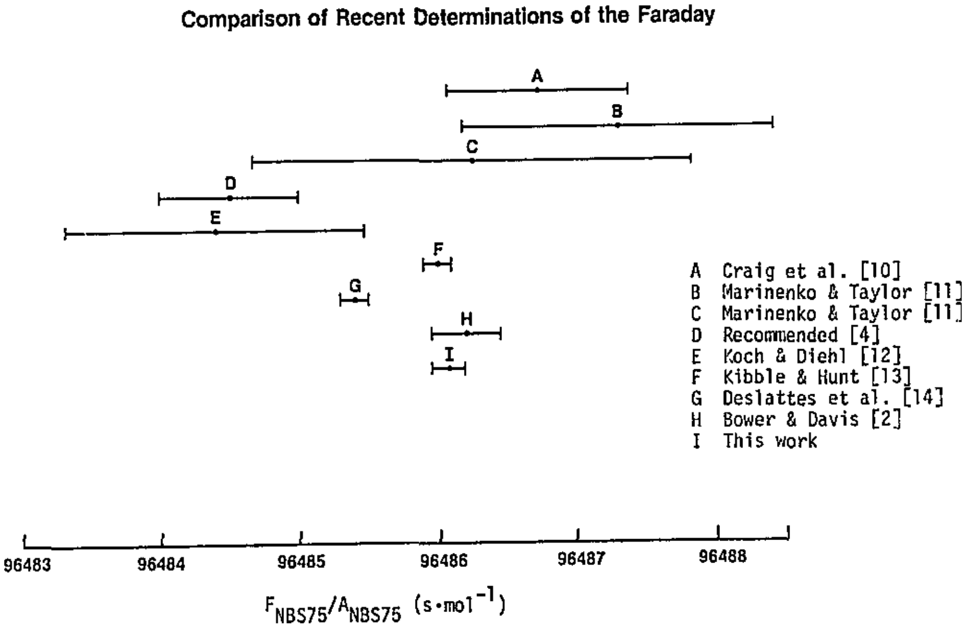

FIGURE 1.

JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 87, No. 1, January–February 1982

An Analysis of Read-Out Perturbations Seen on an Analytical Balance with a Swinging Pan

Marvin E. Cage* and Richard S. Davis*

National Bureau of Standards, Washington, DC 20234

September 8, 1981

An analysis of the dynamic behavior of a single-pan mechanical balance is presented. In particular, errors caused by a swinging pan are analyzed in detail. Results point to a large effect which, though apparently not previously appreciated, is nevertheless easily verified experimentally. It is suggested that this effect can be reduced to insignificance in a balance whose beam is servo-controlled to an angle perpendicular to the local gravitational field.

Key words: Analytical balance; balance dynamics; balance sensitivity; balance suspension; knife-edge bearings; Mathieu's equation; single-pan balance.

1. Introduction

The single-pan analytical balance has become standard equipment in laboratories performing precise mass determinations. The mechanical balances with which this paper is concerned include those having built-in dial weights, a damped beam, and an optical scale which is directly related to beam angle and which has a range equal to or greater than the increment of the smallest dial. Many commercial models of this type of balance achieve a precision of better than 0.5×10^{-6} of maximum load. Indeed, the most precise kilogram comparator—an undamped version of these balances—achieves a relative precision of 1×10^{-9} [1].[1] Equations derived below also treat the case of undamped balances.

It is a common experience of users of one-pan analytical balances that, occasionally, a badly-centered load will cause the pan to swing which, in turn, produces an oscillation of the optical-scale indication. When this occurs, one generally discards the reading and repeats the operation with more care. The question we posed, however, was: By what physical process does the pan-swing manifest itself as an oscillation of the balance beam? The answer might lead to the minimization of the unwanted effect.

The only previous attempt, to our knowledge, at addressing this question has been in a paper by Bowman and Macurdy [2]. These authors reject attempting an analytical solution to the dynamic errors associated with a swinging pan because such a solution would be difficult except under unrealistic conditions. Instead, they point out that a significant part of the dynamic swing-error is due to the centripetal acceleration of the pan. The maximum amplitude of this acceleration is $g\hat{\gamma}^2$ for small $\hat{\gamma}$,, where g is the local acceleration of gravity and $\hat{\gamma}$ is the peak angular amplitude of the pan swing [2]. Even qualitatively, however, this component may be excluded as dominant, at least on many analytical balances. The reason is that a swinging pan is seen to produce an oscillation in the optical indication at the *same frequency* as the pan oscillation, whereas the centripetal acceleration is modulated at *twice the frequency* of the pan (i.e., the centripetal acceleration passes through two maxima during one complete period of pan swing).

Our approach to obtaining a solution to the question posed above has been to generate a complete analytical solution to the equations of motion of an idealized single-pan balance. Our idealized balance operates with knives and flats acting as pivot bearings. The knives do not have perfectly sharp edges but instead have finite radii of curvature. This complexity was added to help elucidate the role of imperfect bearings in the dynamic behavior of the balance.

While our balance model, as will be seen below, requires definite restrictions (such as small-angle motions) and contains some unrealistic assumptions (such as perfectly parallel knives), it nonetheless includes many

---

* Center for Absolute Physical Quantities, National Measurement Laboratory.

[1] Figures in brackets indicate literature references at the end of this paper.

of the fundamental physical aspects of the problem and is amenable to analysis. Semiquantitative predictions of the dynamic behavior of analytical balances can be made based on solutions of the model. These predictions have been verified experimentally as described in section 5 below.

2. Equations of Motion

2.1 Method

The equations of motion for a damped, single-pan balance will be obtained by using the Principle of Virtual Work and D'Alembert's Principle [3, 4]. The suspension system, the balance beam, and the counterweight are initially represented by point masses located at their respective centers of gravity; they will later be treated as distributed masses.

The first step is to determine the coordinates of these point masses as a function of time. This information is used in the Principle of Virtual Work to obtain the generalized torques—and thus the balance beam stiffness and sensitivity expressions. The kinetic energy of the system is also obtained from the equations describing the locations of the point masses. The kinetic energy, the generalized torque, and a velocity-dependent Rayleigh dissipation function [3] are then used in accordance with D'Alembert's Principle to generate the equations of motion for the system.

2.2 Coordinate System

Figure 1 shows the coordinate system for the balance. Note that the z axis, which is defined to be parallel to the local gravitational field, arbitrarily points downward. The y axis is thus parallel to the direction that the plane, defined by the two knife edge contact points (*P*) and (*C*), would have if this plane were at gravitational

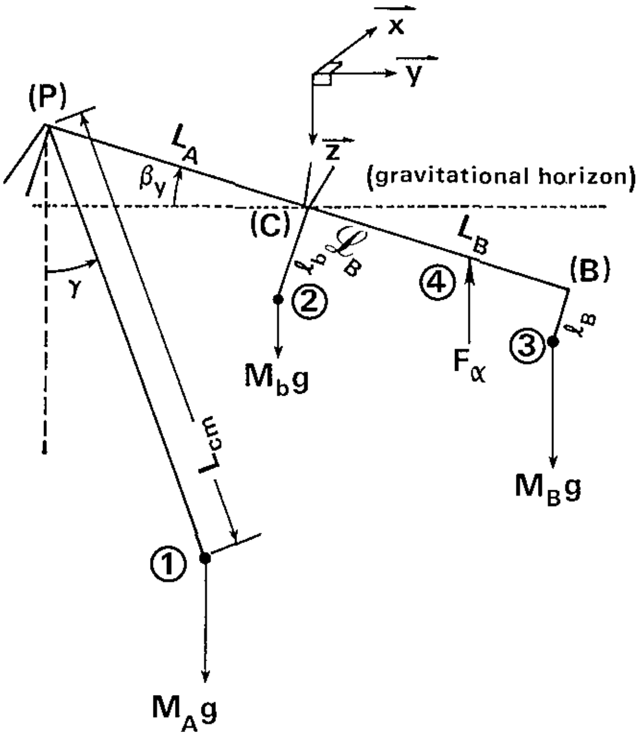

FIGURE 1. The coordinate system and quantities used in the equations of motion. The balance beam rotates in the *y-z* plane, while the suspension system-load mass combination, represented by a simple pendulum, swings either in the *x-z* or the *y-z* plane via crossed knives at pivot point (P). The origin is fixed at point (C), which is the center knife contact point for $\beta_y = 0$.

horizon; i.e., if $\beta_y = 0$. The x axis is directed into the figure so that the coordinate system will be right-handed, and the origin of the balance coordinate system is fixed at point (*C*), which is the center knife contact point for $\beta_y = 0$. Therefore, the balance beam swings in the *y-z* plane about gravitational horizon. The time development of $\beta_y(t)$ describes this motion.

Points 1, 2, and 3 of figure 1 represent the centers of mass of the suspension system—load mass combination (M_A), the balance beam mass (M_b), and the counterweight mass (M_B), respectively. A velocity-dependent damping force, F_α, is applied at point 4. The knife shown at point (*P*) is actually assumed to be two knives crossed in the same plane (see figure 5 of reference [1] or figure 3 of reference [5]). These crossed knives consist of a load knife, *L*, with a symmetry axis in the *x* direction and an intermediate knife, *I*, pointing in the *y* direction. Therefore, the suspension system is free to pivot in any direction about point (*P*) with an instantaneous angle γ. If the knives do not cross in the same plane, then the center of mass length, L_{cm}, can be replaced by $(L_{\mathrm{cm}})_x$ and $(L_{\mathrm{cm}})_y$. (The purpose of the intermediate knife in balance design is to compensate for non-parallelism of the two main knives [5]. In our model the main knives are *assumed* parallel, however. Thus, in our model the only effect of the intermediate knife is to allow a second degree of freedom for a swinging pan.)

Normally, the suspension system would include a gimbal-like device to minimize the positional effects of weights on the scale pan [2]. This would involve a double pendulum on a balance beam—making the problem analytically intractable, and also masking the physical effects. (Actually, the gimbal often has enough friction to quickly dampen the suspension system motion to that of a single pendulum, so the results of this paper still apply to many single-pan balances.)

We will assume the motion of the suspension system to be that of a simple pendulum. In order to gain the maximum physical insight, we consider its motion either to be entirely in the *x-z* plane or entirely in the *y-z* plane.

SI units will be used in the derivations, with masses in kilograms, lengths in meters, forces in Newtons, torques in Newton-meters, and angles in radians. Note that the quantities γ_x, γ_y, β_y, ℓ_b, ℓ_B, and $\mathscr{L}_B$ have signs associated with them—and that they are arbitrarily given positive signs in the instantaneous configuration at time *t* shown in figure 1. Small angle approximations: $\cos\gamma \approx 1$, $\cos\beta_y \approx 1$, $\sin\gamma \approx \gamma$, and $\sin\beta_y \approx \beta_y$ will be made at the appropriate places. The γ approximations are made because precision balances cannot function with large suspension system amplitudes, while the β_y approximations are made in order to obtain analytical solutions.

2.3 Coordinates of the Balance

2.3.1 γ_x and β_y Motions

Assume that the center, load and intermediate knife edges are symmetrically honed cylinders such that they have radii of curvature ρ_C, ρ_L and ρ_I, respectively. Then, with the aid of figures 2-4, the coordinates of points 1, 2, and 3 of figure 1 are:

$$
\begin{aligned}
x_1 &= L_{\mathrm{cm}}\sin\gamma_x - \rho_I\gamma_x \\
y_1 &= -L_A\cos\beta_y - \rho_L\sin\beta_y + \rho_C\beta_y \\
z_1 &= -L_A\sin\beta_y - \rho_L(1-\cos\beta_y) - \rho_C(1-\cos\beta_y) - \rho_I(1-\cos\gamma_x) + L_{\mathrm{cm}}\cos\gamma_x \\
x_2 &= 0 \\
y_2 &= -\ell_b\sin\beta_y + \rho_C\beta_y \\
z_2 &= \ell_b\cos\beta_y - \rho_C(1-\cos\beta_y) \\
x_3 &= 0 \\
y_3 &= L_B\cos\beta_y - \ell_B\sin\beta_y + \rho_C\beta_y \\
z_3 &= L_B\sin\beta_y + \ell_B\cos\beta_y - \rho_C(1-\cos\beta_y),
\end{aligned}
$$

where γ_x and β_y are functions of time.

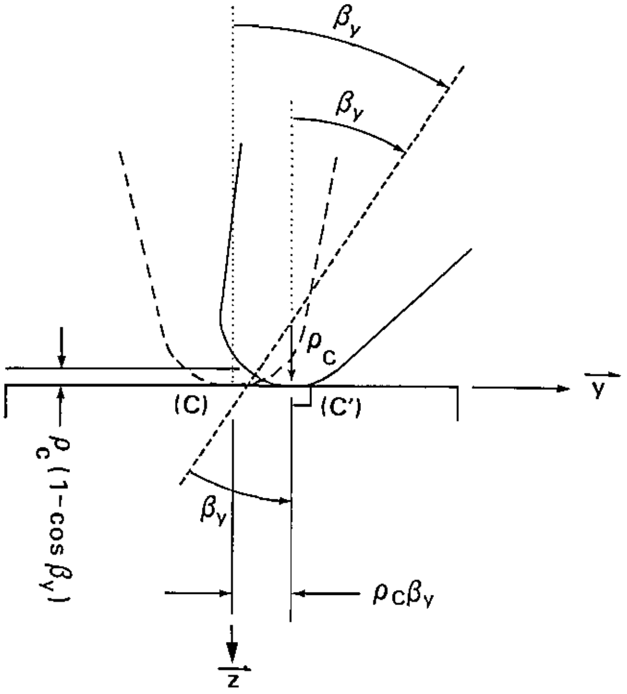

FIGURE 2. Effect on balance beam of center knife rolling on its flat. The origin is fixed at point (C), with the positive x axis directed into the figure.

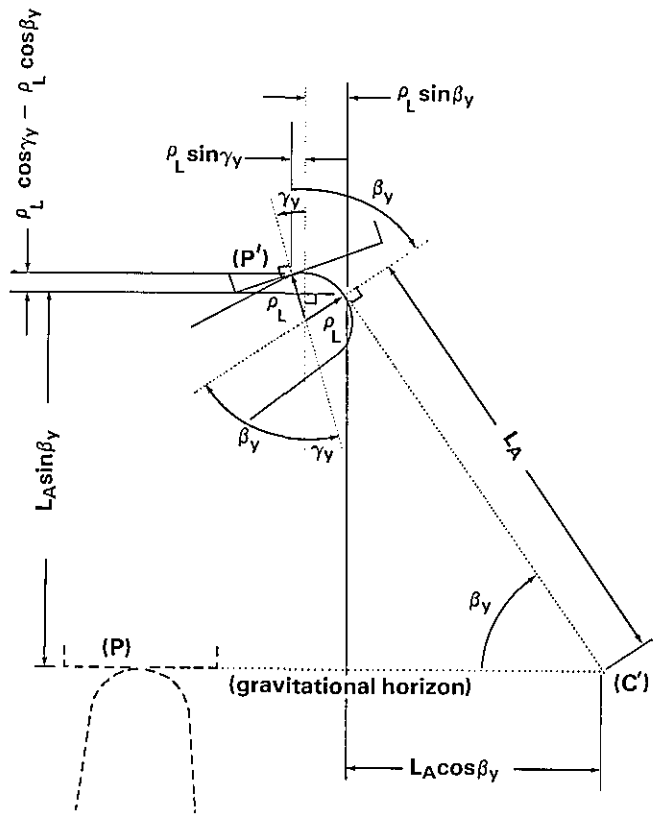

FIGURE 3. Effect on balance beam of the load knife flat rolling on the load knife edge.

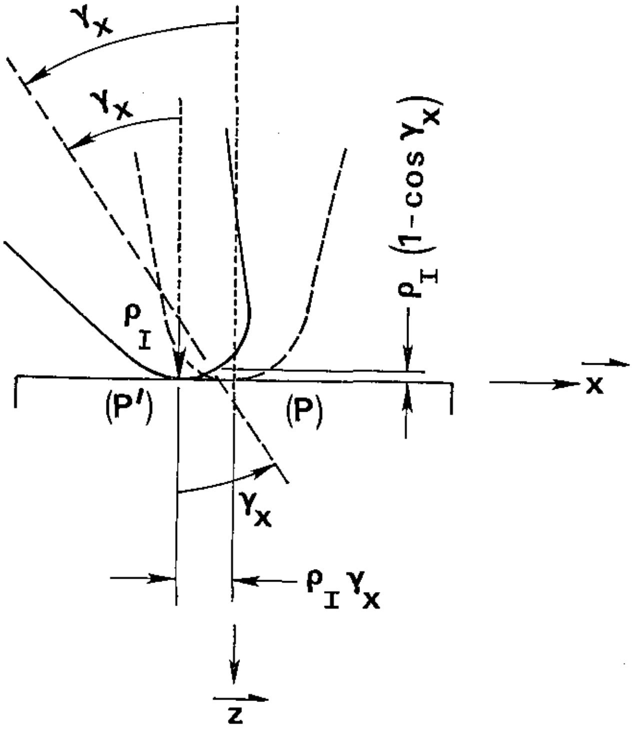

FIGURE 4. Effect of the intermediate knife rolling on a flat. The positive y direction, pointing towards the center knife, is out of the figure. For simplicity, the edges of the crossed knives are assumed to lie in a plane.

2.3.2 γ_y and β_y Motions

The coordinates of points 1, 2, and 3 of figure 1 are:

$$
\begin{aligned}
x_1 &= 0\\
y_1 &= -L_A\cos\beta_y - \rho_L\sin\beta_y - \rho_L\sin\gamma_y + \rho_C\beta_y + L_{cm}\sin\gamma_y\\
z_1 &= -L_A\sin\beta_y - \rho_L\cos\gamma_y + \rho_L\cos\beta_y - \rho_C(1 - \cos\beta_y) + L_{cm}\cos\gamma_y\\
x_2 &= 0\\
y_2 &= -\ell_b\sin\beta_y + \rho_C\beta_y\\
z_2 &= \ell_b\cos\beta_y - \rho_C(1 - \cos\beta_y)\\
x_3 &= 0\\
y_3 &= L_B\cos\beta_y - \ell_B\sin\beta_y + \rho_C\beta_y\\
z_3 &= L_B\sin\beta_y + \ell_B\cos\beta_y - \rho_C(1 - \cos\beta_y).
\end{aligned}
$$

2.4 Principle of Virtual Work

The Principle of Virtual Work [3,4] states that

$$\delta W = \sum_{i=1}^{3} \mathbf{F}_i \cdot \mathbf{d}\boldsymbol{\ell}_i = \sum_{i=1}^{3} [(F_i)_x\delta x_i + (F_i)_y\delta y_i + (F_i)_z\delta z_i]$$

where $x_i = x_i(\gamma,\beta_y)$, $y_i = y_i(\gamma,\beta_y)$ and $z_i = z_i(\gamma,\beta_y)$, and i represents any of the three points shown in figure 1.

Since $(F_i)_x = (F_i)_y = 0$ for this problem, the expression for the virtual work reduces to

$$\begin{aligned}\delta W &= \sum_{i=1}^{3} (F_i)_z \delta z_i \\ &= \sum_{i=1}^{3} \left[(F_i)_z \left(\frac{\partial z_i}{\partial \gamma_x} \right) \delta\gamma_x + (F_i)_z \left(\frac{\partial z_i}{\partial \gamma_y} \right) \delta\gamma_y + (F_i)_z \left(\frac{\partial z_i}{\partial \beta_y} \right) \delta\beta_y \right] \\ &= Q_{\gamma_x}\delta\gamma_x + Q_{\gamma_y}\delta\gamma_y + Q_{\beta_y}\delta\beta_y\end{aligned} \tag{1}$$

where Q_{γ_x}, Q_{γ_y} and Q_{β_y} are the generalized torques for the γ_x, γ_y, and β_y motions. It follows from eq (1) and figure 1 that

$$Q_{\beta_y} = M_A g \frac{\partial z_1}{\partial \beta_y} + M_b g \frac{\partial z_2}{\partial \beta_y} + M_B g \frac{\partial z_3}{\partial \beta_y}. \tag{2}$$

Analogous expressions are readily obtained for Q_{γ_x} and Q_{γ_y}, but we will only be concerned with the consequences of torques about the center knife, rather than about the load or intermediate knives, because we are interested in investigating causes of read-out fluctuations of balances.

2.5 Generalized Torques

Using the coordinates listed in section 2.3.1 for γ_x and β_y motions, and small angle approximations, eq (2) becomes

$$Q_{\beta_y} = [M_B g L_B - M_A g L_A] - [M_A g(\rho_L + \rho_C) + M_b g(\ell_b + \rho_C) + M_B g(\ell_B + \rho_C)]\beta_y. \tag{3}$$

This expression for Q_{β_y} also applies exactly for γ_y and β_y motions.

2.6 Balance Beam Stiffness and Sensitivity

Assume the balance beam is in static equilibrium; therefore, $Q_{\beta_y} = 0$, and (3) becomes

$$[M_B g L_B - M_A g L_A] = [M_A g(\rho_L + \rho_C) + M_b g(\ell_b + \rho_C) + M_B g(\ell_B + \rho_C)]\hat{\beta}_y,$$

which is of the form $\hat{\Gamma}_{\beta_y} = \kappa\hat{\beta}_y$, where $\hat{\Gamma}_{\beta_y}$ is the torque about the center knife, the symbol "^" represents equilibrium values, and κ is the balance-beam stiffness:

$$\kappa = [M_A g(\rho_L + \rho_C) + M_b g(\ell_b + \rho_C) + M_B g(\ell_B + \rho_C)]. \tag{4}$$

This result is the single-pan analogue to the stiffness of a two-pan balance with knives of finite radius [6, 7].

The balance sensitivity, S, to changes in torque is

$$S = \frac{\Delta\hat{\beta}_y}{\Delta\hat{\Gamma}_{\beta_y}} = \frac{1}{\kappa}, \tag{5}$$

and has the dimensions of rad/N·m.[2]

---

[2] Note that if the balance has ideal bearings ($\rho_L = \rho_C = 0$) and if the position of the counter-poise is adjusted so that $\ell_B = 0$, the sensitivity is completely independent of M_A and M_B. This result seems to be in contradiction to eq (7) of [7] and to one statement made in that paper. The formulation of [7] can be reconciled with our own if one realizes that the *coordinate system* of [7] is itself a function of M_B. Hence, for balances whose construction is close to ideal (i.e., $\ell_B << (M_b/M_B)\ell_b$), the *static* properties (such as sensitivity) do not depend strongly on load. The dynamic properties (such as period of the balance) do depend strongly on load: If a 1-kg single-pan mass comparator is converted to a 2-kg comparator by doubling the mass of the counter-poise, the sensitivity of the balance will not change appreciably provided the beam and knives do not deform under this modification. The period of the undamped balance will increase markedly, however. (A clear and succinct discussion of this point can be found in section 3.A of reference [12]).

Note that the balance sensitivity depends upon both ℓ_b and ℓ_B—and that in general neither their individual magnitudes nor signs are known. This does not matter, however, since the sensitivity is a directly measurable quantity.

Equation (3) can be rewritten as $Q_{\beta_y} = [M_B g L_B - M_A g L_A] - \kappa\beta_y$. Therefore, the effective changes in lever arms L_A and L_B—due to the knife edges rolling on their flats—enters into the quantity $\kappa\beta_y$; also, κ and S will not be constant if the effective values of ρ_C and ρ_L vary with β_y.

2.7 Kinetic Energies

2.7.1 γ_x and β_y Motions

The kinetic energy of the system is

$$T = T_1 + T_2 + T_3$$

$$T = \frac{1}{2} M_A(\dot{x}_1^2 + \dot{y}_1^2 + \dot{z}_1^2) + \frac{1}{2} M_b(\dot{x}_2^2 + \dot{y}_2^2 + \dot{z}_2^2) + \frac{1}{2} M_B(\dot{x}_3^2 + \dot{y}_3^2 + \dot{z}_3^2). \tag{6}$$

Using the coordinates listed in section 2.3.1 in eq (6), employing small angle approximations, and keeping only the significant terms, yields the result

$$\begin{aligned} T = {} & \frac{1}{2} M_A[L_A^2(\dot{\beta}_y)^2 + L_{\mathrm{cm}}^2(\dot{\gamma}_x)^2 - 2L_{\mathrm{cm}}\rho_I(\dot{\gamma}_x)^2 \\ & + 2L_A L_{\mathrm{cm}}\gamma_x\dot{\gamma}_x\dot{\beta}_y] + \frac{1}{2} M_b\ell_b^2(\dot{\beta}_y)^2 \\ & + \frac{1}{2} M_B[L_B^2 + \ell_B^2]\,(\dot{\beta}_y)^2. \end{aligned} \tag{7}$$

2.7.2 γ_y and β_y Motions

When the coordinates listed in section 2.3.2 are used in eq (6), the kinetic energy expression becomes

$$\begin{aligned} T = {} & \frac{1}{2} M_A[L_A^2(\dot{\beta}_y)^2 + L_{\mathrm{cm}}^2(\dot{\gamma}_y)^2 - 2L_{\mathrm{cm}}\rho_L(\dot{\gamma}_y)^2 \\ & + 2L_A L_{\mathrm{cm}}(\gamma_y + \beta_y)\dot{\gamma}_y\dot{\beta}_y + 2L_{\mathrm{cm}}(\rho_C - \rho_L)\dot{\gamma}_y\dot{\beta}_y] \\ & + \frac{1}{2} M_b\ell_b^2(\dot{\beta}_y)^2 + \frac{1}{2} M_B[L_B^2 + \ell_B^2](\dot{\beta}_y)^2\,. \end{aligned} \tag{8}$$

2.8 Rayleigh's Dissipation Function

We will assume that if the balance beam is mechanically or electronically damped, that the damping force is proportional to the velocity. Let the damping occur at point 4 of figure 1. Then for small angles, the Rayleigh dissipation function [3], F, is

$$F = \frac{1}{2}\alpha\dot{z}_4^2 = \frac{1}{2}\alpha\mathscr{L}_B^2(\dot{\beta}_y)^2\,, \tag{9}$$

and the damping force, F_α, is

$$F_\alpha = -\frac{\partial F}{\partial v_{z_4}} = -\alpha\mathscr{L}_B\dot{\beta}_y. \tag{10}$$

2.9 D'Alembert's Principle

2.9.1 γ_x and β_y Motions

D'Alemberts' Principle [3,4] for the equation of motion of the balance beam about the center knife may be written:

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{\beta}_y}\right) - \frac{\partial T}{\partial \beta_y} + \frac{\partial F}{\partial \dot{\beta}_y} = Q_{\beta_y}. \tag{11}$$

Substituting the expressions for Q_{β_y}, T and F given by eqs (3), (4), (7), and (9) into (11), and replacing the point masses of figure 1 with distributed masses, yields the result

$$[I_C]\ddot{\beta}_y + [\alpha_B^2]\dot{\beta}_y + \kappa\beta_y - [M_B g L_B - M_A g L_A] = -[M_A L_A L_{\mathrm{cm}}](\dot{\gamma}_x)^2 - [M_A L_A L_{\mathrm{cm}}]\gamma_x\ddot{\gamma}_x. \tag{12}$$

I_C is the moment of inertia of the balance about point C of figure 1 when $\beta_y = 0$; i.e.,

$$\begin{aligned} I_C &= I_1 + I_2 + I_3 \\ I_C &= [M_A L_A^2] + [(I_b)_{\mathrm{cm}} + M_b \ell_b^2] + [(I_B)_{\mathrm{cm}} + M_B(L_B^2 + \ell_B^2)] , \end{aligned} \tag{13}$$

where $(I_b)_{\mathrm{cm}}$ and $(I_B)_{\mathrm{cm}}$ are the respective moments of inertia of the balance beam and the counterweight about the x axis through the center of mass, points 2 and 3 of figure 1. The $M_b\ell_b^2$ and $M_B(L_B^2 + \ell_B^2)$ terms result from the parallel axis theorem. The combination of suspension system plus load acts as a point mass—as far as the center knife is concerned—because of the crossed knives at point P.

Equation (12) has the form

$$b\ddot{\beta}_y + c\dot{\beta}_y + \kappa\beta_y - d = -e(\dot{\gamma}_x)^2 - e\gamma_x\ddot{\gamma}_x. \tag{14}$$

Appendix A lists all algebraic substitutions used in the derivations, such as those for the coefficients b, c, κ, d and e of (14).

The r.h.s. terms of (12) and (14) can be explained by considering the torques about the center knife due to the radial and tangential accelerations of the suspension system-load mass combination about point (P') as shown in figure 5. The torque Γ_C' is

$$\begin{aligned} \Gamma_C' = {} & (M_A a_R \cos\gamma_x)(-L_A\cos\beta_y - \rho_L\sin\beta_y - \rho_C\beta_y) \\ & + (M_A a_T \sin\gamma_x)(-L_A\cos\beta_y - \rho_L\sin\beta_y - \rho_C\beta_y) , \end{aligned} \tag{15}$$

where $a_R = L_{\mathrm{cm}}(\dot{\gamma}_x)^2$ and $a_T = L_{\mathrm{cm}}\ddot{\gamma}_x$. Note that the torque Γ_C' is about the pivot point (0) of the center knife rather than about the origin (C), or about the contact point (C'); (see fig. 2). For small angles, eq (15) reduces to

$$\Gamma_C' = -[M_A L_A L_{\mathrm{cm}}](\dot{\gamma}_x)^2 - [M_A L_A L_{\mathrm{cm}}]\gamma_x\ddot{\gamma}_x = -e(\dot{\gamma}_x)^2 - e\gamma_x\ddot{\gamma}_x,$$

thus accounting for the r.h.s. terms of (12) and (14) by using Newtonian mechanics.

2.9.2 γ_y and β_y Motions

Substituting the expressions for Q_{β_y}, T and F given by eqs (3), (4), (8) and (9) into (11), and using distributed masses for the balance, yields the result

$$[I_C]\ddot{\beta}_y + [\alpha_B^2]\dot{\beta}_y + \kappa\beta_y - [M_B g L_B - M_A g L_A]$$
$$= -[M_A L_A L_{\mathrm{cm}}](\dot{\gamma}_y)^2 - [M_A L_A L_{\mathrm{cm}}](\gamma_y + \beta_y)\ddot{\gamma}_y - [M_A L_{\mathrm{cm}}(\rho_C - \rho_L)]\ddot{\gamma}_y \,, \qquad (16)$$

which has the form

$$b\ddot{\beta}_y + c\dot{\beta}_y + \kappa\beta_y - d = -e(\dot{\gamma}_y)^2 - e(\gamma_y + \beta_y)\ddot{\gamma}_y - f\ddot{\gamma}_y \,. \qquad (17)$$

Referring to figure 5, the torques about the pivot point (0) of the knife due to the radial and tangential accelerations of the suspension system-load mass combination are

$$\begin{aligned}\Gamma_C = {} & (M_A a_R \cos\gamma_y)(-L_A\cos\beta_y - \rho_L\sin\beta_y - \rho_L\sin\gamma_y - \rho_C\beta_y) \\ & + (M_A a_R \sin\gamma_y)[-L_A\sin\beta_y - \rho_L\cos\gamma_y + \rho_L\cos\beta_y - \rho_C(1 - \cos\beta_y) - \rho_C] \\ & + (M_A a_T \cos\gamma_y)[-L_A\sin\beta_y - \rho_L\cos\gamma_y + \rho_L\cos\beta_y - \rho_C(1 - \cos\beta_y) - \rho_C] \\ & + (M_A a_T \sin\gamma_y)(-L_A\cos\beta_y - \rho_L\sin\beta_y - \rho_L\sin\gamma_y - \rho_C\beta_y) - (M_A a_T)(-\rho_L) \,,\end{aligned} \qquad (18)$$

where $a_R = L_{\mathrm{cm}}(\dot{\gamma}_y)^2$ and $a_T = L_{\mathrm{cm}}\ddot{\gamma}_y$. The last term in (18) results from a torque about the center of curvature of the load knife due to an action-reaction force $M_A a_T$ applied to the non-slipping contact point (P'); this torque is transmitted to point (0) because the balance beam is rigid.

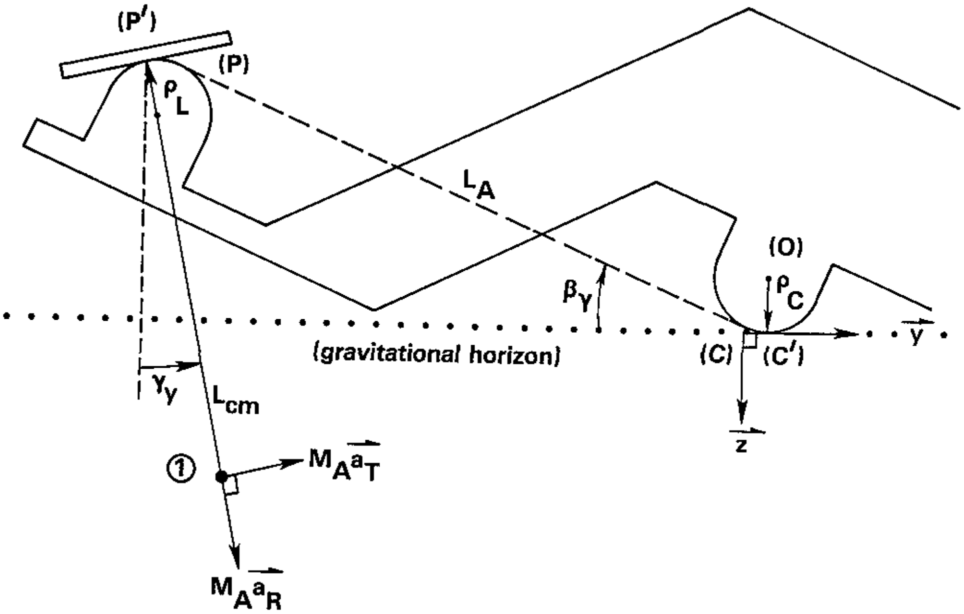

FIGURE 5. Dynamically induced forces on the swinging pendulum. These forces produce torques about the pivot axis of the center knife. This axis, which is in the x direction, passes through the center of curvature of the center knife at point (0).

For small angles, eq (18) becomes

$$\begin{aligned}\Gamma_C &= -[M_A L_A L_{\mathrm{cm}}](\dot{\gamma}_y)^2 - [M_A L_A L_{\mathrm{cm}}](\gamma_y + \beta_y)\ddot{\gamma}_y - [M_A L_{\mathrm{cm}}(\rho_C - \rho_L)]\ddot{\gamma}_y \\ &= -e(\dot{\gamma}_y)^2 - e(\gamma_y + \beta_y)\ddot{\gamma}_y - f\ddot{\gamma}_y,\end{aligned}$$

thus accounting for the r.h.s. terms of (17).

The $-[M_A L_A L_{cm}]\beta_y\ddot{\gamma}_y = -e\beta_y\ddot{\gamma}_y$ term of (16) (which is due to M_A times the horizontal component of the tangential acceleration times the change in height from gravitational horizon of the load end of the balance beam) will later be seen to have a special significance. We will refer to this quantity (divided by I_C) as the Mathieu term. (There is also a Mathieu term contribution due to the horizontal component of the radial acceleration; but it is negligible for small angles.) The quantity $[M_A L_{cm}(\rho_C - \rho_L)/I_C]\ddot{\gamma}_y$ will later be referred to as the knife-edges term.

2.10 Physical Pendulum

2.10.1 γ_x and β_y Motions

We assume that the motion of the suspension system-load combination is that of a physical pendulum, and arbitrarily choose γ_x to be a minimum at time $t = 0$ and to have an amplitude $\tilde{\gamma}_x$; i.e.,

$$\gamma_x = -\tilde{\gamma}_x \cos\omega_{\gamma_x} t \,, \tag{19}$$

where

$$\omega^2_{\gamma_x} = \frac{[M_A g L_{cm}]}{I'_P}, \tag{20}$$

$$I'_P = (I'_S)_{cm} + M_A L^2_{cm}, \tag{21}$$

and

$$\omega_{\gamma_x} = \frac{2\pi}{P_{\gamma_x}}. \tag{22}$$

I'_P is the moment of inertia of the physical pendulum about the y axis at point (P) of figure 1, $(I'_S)_{cm}$ is the moment of inertia of the suspension system-load mass combination about the y axis through the center of mass point 1 of figure 1, and $P\gamma_x$ is the pendulum period.

The higher order modulation terms have been ignored in (19) because they do not significantly affect the balance beam motion. They could be readily obtained by considering the equations of motion about point (P) of figure 1.

2.10.2 γ_y and β_y Motions

Analogous to section 2.10.1:

$$\gamma_y = -\tilde{\gamma}_y \cos\omega_{\gamma_y} t, \tag{23}$$

$$\omega^2_{\gamma_y} = \frac{[M_A g L_{cm}]}{I_P}, \tag{24}$$

$$I_P = (I_S)_{cm} + M_A L^2_{cm}, \tag{25}$$

and

$$\omega_{\gamma_y} = \frac{2\pi}{P_{\gamma_y}}. \tag{26}$$

I_P and $(I_S)_{cm}$ are the moments of inertia about the x axis through points (P) and 1, respectively of figure 1.

3. Solution for γ_x and β_y Motions

The equation of motion can now be expressed in final form by substituting eq (19) into (14), and by using the identity $\cos^2\psi-\sin^2\psi = \cos 2\psi$; it is

$$\ddot{\beta}_y + \frac{c}{b}\dot{\beta}_y + \frac{\kappa}{b}\beta_y - \frac{d}{b} = \frac{e}{b}\tilde{\gamma}_x^2\omega_{\gamma_x}^2\cos 2\omega_{\gamma_x}t. \qquad (27)$$

The general solution of (27) will consist of the solutions to the homogeneous equation (where the r.h.s. of (27) is zero), plus a particular solution.

The homogeneous solution to this equation is easily obtained by standard techniques. The particular solution can be found by writing (27) in the form

$$\ddot{\beta}_y + \bar{u}\dot{\beta}_y + v^2\beta_y - \frac{d}{b} = w\cos kt$$

and by assuming that the solution is

$$\beta_y = D\cos kt + E\sin kt + \hat{\beta}_y;$$

which yields the results

$$D = \frac{\bar{u}wk}{[(v^2 - k^2)^2 + \bar{u}^2k^2]}$$

and

$$E = \frac{w(v^2 - k^2)}{[(v^2 - k^2)^2 + \bar{u}^2k^2]}.$$

The general solution of (27) is therefore

$$\begin{aligned}\beta_y = \{&e^{-\frac{ct}{2b}}\left[A\exp\sqrt{\omega_{\beta_y}'^2}t + B\exp-\sqrt{-\omega_{\beta_y}'^2}t\right]\\ &\text{or } (A + Bt)e^{-\frac{ct}{2b}} \text{ or}\\ &e^{-\frac{ct}{2b}}\left[A\cos\omega_{\beta_y}' t + B\sin\omega_{\beta_y}' t\right]\}\\ &+ \hat{\beta}_y + \frac{e}{b}\tilde{\gamma}_x^2\omega_{\gamma_x}^2 \frac{(\omega_{\beta_y}^2 - 4\omega_{\gamma_x}^2)\cos 2\omega_{\gamma_x}t}{\left[(\omega_{\beta_y}^2 - 4\omega_{\gamma_x}^2)^2 + 4\frac{c^2}{b^2}\omega_{\gamma_x}^2\right]}\\ &+ 2\frac{ce}{b^2}\tilde{\gamma}_x^2\omega_{\gamma_x}^3 \frac{\sin 2\omega_{\gamma_x}t}{\left[(\omega_{\beta_y}^2 - 4\omega_{\gamma_x}^2)^2 + 4\frac{c^2}{b^2}\omega_{\gamma_x}^2\right]},\end{aligned} \qquad (28)$$

where

$$\omega_{\beta_y}^2 = \kappa/b, \qquad (29)$$

$$\omega_{\beta_y}'^2 = \frac{\kappa}{b} - \frac{c^2}{4b^2}, \qquad (30)$$

and

$$\hat{\beta}_y = \frac{d}{\kappa} = \frac{[M_B g L_B - M_A g L_A]}{[M_A g(\rho_L + \rho_C) + M_b g(\ell_b + \rho_C) + M_B g(\ell_B + \rho_C)]} \tag{31}$$

The choice of terms in eq (28) arises because the balance may be overdamped, critically damped, or underdamped depending on whether $c^2/4b$ is greater than, equal to, or less than κ. For all damping conditions the motion of the balance beam about the equilibrium angle $\hat{\beta}_y$ is modulated at an angular frequency which is twice that of the natural angular frequency ω_{γ_y} of the physical pendulum. This modulation effect is due to the radial and tangential accelerations of the suspension system-load mass combination.

4. Solution for γ_y and β_y Motions

4.1 Equation of Motion

The equation of motion is obtained by substituting eq (23) into (17), and by using the identity $\cos^2\psi - \sin^2\psi = \cos 2\psi$:

$$\frac{d^2\beta_y}{dt^2} + \frac{c}{b}\frac{d\beta_y}{dt} + \left(\frac{\kappa}{b} + \frac{e}{b}\hat{\gamma}_y\omega^2_{\gamma_y}\cos\omega_{\gamma_y}t\right)\beta_y = \frac{d}{b} + \frac{e}{b}\hat{\gamma}^2_y\omega^2_{\gamma_y}\cos 2\omega_{\gamma_y}t - \frac{f}{b}\hat{\gamma}_y\omega^2_{\gamma_y}\cos\omega_{\gamma_y}t. \tag{32}$$

The $\left(\frac{e}{b}\hat{\gamma}_y\omega^2_{\gamma_y}\cos\omega_{\gamma_y}t\right)\beta_y$ term of eq (32)—which is the Mathieu term referred to in section 2.9.2—complicates the problem; yet it cannot, in general, be neglected when considering the dynamical effects of precision balances. Therefore, (32) will be converted to a more tractable form.

First, we change from the time variable, t, to the dimensionless variable z by making the substitution

$$\omega_{\gamma_y}t = 2z. \tag{33}$$

Equation (32) thus becomes

$$\frac{d^2\beta_y}{dz^2} + 2\left[\frac{c}{b\omega_{\gamma_y}}\right]\frac{d\beta_y}{dz} + \left(\left[4\frac{\omega^2_{\beta_y}}{\omega^2_{\gamma_y}}\right] - 2\left[\frac{-2e}{b}\hat{\gamma}_y\right]\cos 2z\right)\beta_y$$
$$= \left[\frac{4}{\omega^2_{\gamma_y}}\frac{d}{b}\right] + \left[4\frac{e}{b}\hat{\gamma}^2_y\right]\cos 2\omega_{\gamma_y}t - \left[4\frac{f}{b}\hat{\gamma}_y\right]\cos\omega_{\gamma_y}t\,, \tag{34}$$

or

$$\frac{d^2\beta_y}{dz^2} + 2\bar{k}\frac{d\beta_y}{dz} + (\bar{a} - 2q\cos 2z)\beta_y = h + \ell\cos 4z - m\cos 2z = \bar{f}(z). \tag{35}$$

We then replace β_y with the quantity [10]

$$\beta_y = e^{-\bar{k}z}u(z)\,; \tag{36}$$

thus obtaining the equation of motion in final form:

$$\frac{d^2u}{dz^2} + (a - 2q\cos 2z)u = e^{\bar{k}z}\bar{f}(z), \tag{37}$$

where

$$a = \bar{a} - \bar{k}^2 = \frac{4}{\omega^2_{\gamma_y}}\left[\frac{\kappa}{b} - \frac{c^2}{4b^2}\right]. \tag{38}$$

The quantity a is negative if the balance beam is overdamped, zero if critically damped, and positive if underdamped; whereas q is always a negative number. The quantities z, $\bar{f}(z)$, $u(z)$, $\bar{k}$, $\bar{a}$, a, q, h, ℓ, and m are all dimensionless. The general solution of (37) consists of the solutions to the homogeneous equation, plus a particular solution.

4.2 Homogeneous Solutions

The homogeneous equation

$$u'' + (a - 2q\cos 2z)u = 0 \tag{39}$$

has the canonical form of Mathieu's equation [9,10]. The reader is referred to reference [10] for an excellent account of how to obtain solutions to this equation.

Figure 6 shows an isostability diagram [10] of ν for possible solutions to eq (39). We are only interested in underdamped or critically damped solutions because laboratory analytical balances are seldom designed to operate in the overdamped mode; therefore $1 >> a \geqslant 0$. Also, the quantity $q = -2M_A L_A L_{cm} \tilde{\gamma}_y / I_C$ is always a very small negative real number for precision balances; so $-1 << q < 0$. The region of interest for these values of a and q is indicated by an arrow in figure 6. Clearly, the solutions of (39) cannot be "simple" Mathieu functions of the sine or cosine types $s(e_1)$, $c(e_0)$ or $c(e_1)$ [9,10], but instead are more complicated functions.

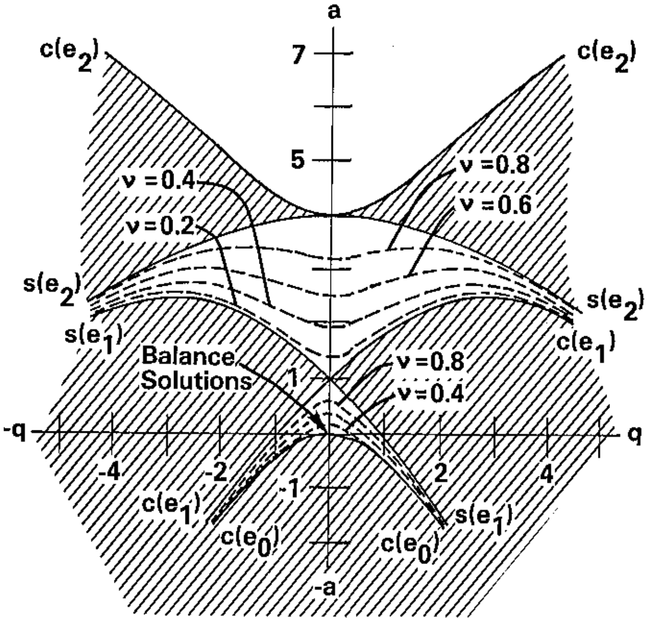

FIGURE 6. An isostability diagram of ν for possible solutions of Mathieu's equation given by (39). The arrow points to the region of interest for precision balances; i.e. $1 >> a \geqslant 0$ and $-1 << q < 0$. Solutions in the shaded region are unstable. This does not mean that an overdamped balance ($a < 0$) is mechanically unstable because $u(z)$ is multiplied by the factor $e^{-\bar{k}z}$ in the solution for β_y.

Since q is small, we guess that the solution is of the Floquet form [9]

$$u = e^{i\nu q} - \frac{qe^{i(\nu+2)z}}{4(\nu+1)} + \frac{qe^{i(\nu-2)z}}{4(\nu-1)}, \tag{40}$$

where ν is *not* an integer; (i.e., $u(z)$ is not a Mathieu function). For small values of a, ν is defined by [9]

$$\cos\nu\pi = \left(1 - \frac{a\pi^2}{2} + \frac{a^2\pi^4}{24} + \cdots\right) - \frac{q^2\pi^2}{4}\left[1 + a\left(1 - \frac{\pi^2}{6}\right) + \cdots\right]$$
$$+ q^4\left(\frac{\pi^4}{96} - \frac{25\pi^2}{256} + \cdots\right) + \cdots .$$

We then use the identity $e^{i\theta} = \cos\theta + i\sin\theta$ in (40) to obtain the two independent solutions of eq (39); they are

$$u_1 = \cos\nu z - \bar{q}\cos\nu z\cos 2z - \nu\bar{q}\sin\nu z\sin 2z \tag{41}$$

and

$$u_2 = \sin\nu z - \bar{q}\sin\nu z\cos 2z + \nu\bar{q}\cos\nu z\sin 2z, \tag{42}$$

where terms of order $\bar{q}^2$ have been neglected, and

$$\bar{q} = \frac{q}{2(1-\nu^2)}, \tag{43}$$

$$\nu \approx \sqrt{a} \quad (\text{for } a > 0), \tag{44}$$

or

$$\nu \approx \frac{q}{\sqrt{2}} \quad (\text{for } a = 0).$$

The homogeneous solution of (39) is therefore

$$Au_1(z) + Bu_2(z). \tag{45}$$

4.3 General Solution

4.3.1 Particular Solution

We must next find a particular solution, $u_p(z)$, of

$$u'' + (a - 2q\cos 2z)u = e^{\bar{k}\bar{z}}\bar{f}(z), \tag{46}$$

which has the form of Hill's equation [10]. The solution is obtained from the expression

$$u_p(z) = -\frac{1}{\bar{c}^2}\left[u_1(z)\int_0^z u_2(z)e^{\bar{k}\bar{z}}\bar{f}(z)dz - u_2(z)\int_0^z u_1(z)e^{\bar{k}\bar{z}}\bar{f}(z)dz\right] \tag{47}$$

where u_1 and u_2 are given by eqs (41) and (42), $\bar{f}(z) = h + \ell\cos 4z - m\cos 2z$, and

$$\bar{c}^2 = u_1(z)u_2'(z) - u_2(z)u_1'(z) = \nu. \tag{48}$$

The evaluation of eq (47) is straightforward, but very lengthy; the result is included in the next subsection.

4.3.2 Complete Solution

The general solution to eq (35), using (36), (45), (47), and (48), has the form

$$\beta_y = e^{-\bar{k}z}u(z) = e^{-\bar{k}z}[Au_1(z) + Bu_2(z)] - \frac{e^{-\bar{k}z}}{\nu}\left[u_1(z)\int_0^z u_2(z)e^{\bar{k}z}\bar{f}(z)dz - u_2(z)\int_0^z u_1(z)e^{\bar{k}z}\bar{f}(z)dz\right], \tag{49}$$

which, when using (41) and (42) is

$$\begin{aligned}
\beta_y = {} & Ae^{-\bar{k}z}[\cos\nu z - \bar{q}\cos\nu z\cos 2z - \nu\bar{q}\sin\nu z\sin 2z] \\
& + Be^{-\bar{k}z}[\sin\nu z - \bar{q}\sin\nu z\cos 2z + \nu\bar{q}\cos\nu z\sin 2z] \\
& - \frac{1}{\nu[\bar{k}^2+\nu^2]}\Biggl\{\biggl(h + \frac{\bar{q}}{2}m\biggr)[-\nu + e^{-\bar{k}z}(\nu\cos\nu z + \bar{k}\sin\nu z)] \\
& - (\bar{q}h)[-\nu\cos 2z + e^{-\bar{k}z}(\nu\cos\nu z\cos 2z + \bar{k}\sin\nu z\cos 2z)] \\
& - (\nu\bar{q}h)[\bar{k}\sin 2z + e^{-\bar{k}z}(\nu\sin\nu z\sin 2z + \bar{k}\cos\nu z\sin 2z)]\Biggr\} \\
& - \frac{1}{2\nu[\bar{k}^2+(\nu-2)^2]}\Biggl\{\biggl(-m - \bar{q}h - \frac{\bar{q}}{2}\ell - \nu\bar{q}h + \frac{\nu\bar{q}}{2}\ell\biggr)[-\bar{k}\sin 2z \\
& \qquad - (\nu-2)\cos 2z + e^{-\bar{k}z}((\nu-2)\cos\nu z + \bar{k}\sin\nu z)] \\
& + (\bar{q}m)[-\bar{k}\sin 2z\cos 2z - (\nu-2)\cos^2 2z + e^{-\bar{k}z}((\nu-2)\cos\nu z\cos 2z + \bar{k}\sin\nu z\cos 2z)] \\
& + (\nu\bar{q}m)[-(\nu-2)\sin^2 2z + \bar{k}\sin 2z\cos 2z + e^{-\bar{k}z}((\nu-2)\sin\nu z\sin 2z - \bar{k}\cos\nu z\sin 2z)]\Biggr\} \\
& - \frac{1}{2\nu[\bar{k}^2 + (\nu+2)^2]}\Biggl\{\biggl(-m - \bar{q}h - \frac{\bar{q}}{2}\ell + \nu\bar{q}h - \nu\frac{\bar{q}}{2}\ell\biggr)[\bar{k}\sin 2z \\
& \qquad - (\nu+2)\cos 2z + e^{-\bar{k}z}((\nu+2)\cos\nu z + \bar{k}\sin\nu z)] \\
& + (\bar{q}m)[\bar{k}\sin 2z\cos 2z - (\nu+2)\cos^2 2z + e^{-\bar{k}z}((\nu+2)\cos\nu z\cos 2z + \bar{k}\sin\nu z\cos 2z)] \\
& + (\nu\bar{q}m)[\bar{k}\sin 2z\cos 2z + (\nu+2)\sin^2 2z + e^{-\bar{k}z}((\nu+2)\sin\nu z\sin 2z - \bar{k}\cos\nu z\sin 2z)]\Biggr\} \\
& - \frac{1}{2\nu(\bar{k}^2 + (\nu-4)^2]}\Biggl\{\biggl(\ell + \frac{\bar{q}}{2}m + \nu\frac{\bar{q}}{2}m\biggr)[-\bar{k}\sin 4z - (\nu-4)\cos 4z \\
& \qquad + e^{-\bar{k}z}((\nu-4)\cos\nu z + \bar{k}\sin\nu z)] \\
& - (\bar{q}\ell)[-\bar{k}\cos 2z\sin 4z - (\nu-4)\cos 2z\cos 4z + e^{-\bar{k}z}((\nu-4)\cos\nu z\cos 2z + \bar{k}\sin\nu z\cos 2z)] \\
& - (\nu\bar{q}\ell)[\bar{k}\sin 2z\cos 4z - (\nu-4)\sin 2z\sin 4z + e^{-\bar{k}z}((\nu-4)\sin\nu z\sin 2z - \bar{k}\cos\nu z\sin 2z)]\Biggr\} \\
& - \frac{1}{2\nu[\bar{k}^2 + (\nu+4)^2]}\Biggl\{\biggl(\ell + \frac{\bar{q}}{2}m - \nu\frac{\bar{q}}{2}m\biggr)[\bar{k}\sin 4z - (\nu+4)\cos 4z \\
& \qquad + e^{-\bar{k}z}((\nu+4)\cos\nu z + \bar{k}\sin\nu z)]
\end{aligned}$$

$$- (\bar{q}\ell)[\bar{k}\cos 2z \sin 4z - (\nu+4)\cos 2z \cos 4z + e^{-\bar{k}z}((\nu+4)\cos \nu z \cos 2z + \bar{k}\sin \nu z \cos 2z)]$$

$$\left. - (\nu\bar{q}\ell)[\bar{k}\sin 2z \cos 4z + (\nu+4)\sin 2z \sin 4z + e^{-\bar{k}z}((\nu+4)\sin \nu z \sin 2z - \bar{k}\cos \nu z \sin 2z)] \right\}$$

$$- \frac{1}{2\nu[\bar{k}^2 + (\nu-6)^2]} \left\{ \left(- \frac{\bar{q}}{2}\ell - \nu\frac{\bar{q}}{2}\ell \right) [-\bar{k}\sin 6z - (\nu-6)\cos 6z \right.$$

$$\left. + e^{-\bar{k}z}((\nu-6)\cos \nu z + \bar{k}\sin \nu z)] \right\}$$

$$- \frac{1}{2\nu[\bar{k}^2 + (\nu+6)^2]} \left\{ \left(- \frac{\bar{q}}{2}\ell + \nu\frac{\bar{q}}{2}\ell \right) [\bar{k}\sin 6z - (\nu+6)\cos 6z \right.$$

$$\left. + e^{-\bar{k}z}((\nu+6)\cos \nu z + \bar{k}\sin \nu z)] \right\}. \tag{50}$$

This is the *exact* solution (in reduced notation) for small angle γ_y and β_y motions. It is obviously much more complicated than the solution for small angle γ_x and β_y motions given by eq (28).

It requires too many printed pages to provide an expanded form of eq (50) written in the notation of (28). (There are 74 terms in that notation.) However, the reader can readily transform any term of (50) into that notation with the aid of Appendices A and B. Equation (50) is provided so that readers seeking solutions to problems similar to our own may be spared the considerable tedium of its derivation and so that the terms neglected in the sequel may be recovered by the interested reader.

5. Special Cases

5.1 A Freely-Swinging Balance

In a completely undamped balance, $\bar{k} = c = 0$, and $\omega'_{\beta_y} = \omega_{\beta_y}$; so eq (28) becomes

$$\beta_y = A\cos\omega_{\beta_y}t + B\sin\omega_{\beta_y}t + \hat{\beta}_y + \frac{e}{b}\tilde{\gamma}_x^2 \frac{\omega_{\gamma_x}^2}{(\omega_{\beta_y}^2 - 4\omega_{\gamma_x}^2)}\cos 2\omega_{\gamma_x}t. \tag{51}$$

Using Appendices A and B, and keeping only the largest terms, (50) becomes

$$\beta_y = A\cos\omega_{\beta_y}t + B\sin\omega_{\beta_y}t + \left[\hat{\beta}_y - \frac{ef}{2b^2}\tilde{\gamma}_y^2 \frac{\omega_{\gamma_y}^4}{\omega_{\beta_y}^2(\omega_{\gamma_y}^2 - 4\omega_{\beta_y}^2)}\right]$$

$$+ \frac{e}{b}\hat{\beta}_y\tilde{\gamma}_y \frac{\omega_{\gamma_y}^2}{(\omega_{\gamma_y}^2 - 4\omega_{\beta_y}^2)}\cos\omega_{\gamma_y}t - \frac{f}{b}\tilde{\gamma}_y \frac{\omega_{\gamma_y}^2}{(\omega_{\beta_y}^2 - \omega_{\gamma_y}^2)}\cos\omega_{\gamma_y}t$$

$$+ \frac{e}{b}\tilde{\gamma}_y^2 \frac{\omega_{\gamma_y}^2}{(\omega_{\beta_y}^2 - 4\omega_{\gamma_y}^2)}\cos 2\omega_{\gamma_y}t + \frac{e^2}{2b^2}\tilde{\gamma}_y^3 \frac{\omega_{\gamma_y}^4}{(\omega_{\gamma_y}^2 - 4\omega_{\beta_y}^2)(\omega_{\beta_y}^2 - 9\omega_{\gamma_y}^2)}\cos 3\omega_{\gamma_y}t, \tag{52}$$

where $b = I_C$, $e = M_A L_A L_{cm}$ and $f = M_A L_{cm}(\rho_C - \rho_L)$. Only cosine modulations of the $A\cos\omega_{\beta_y}t + B\sin\omega_{\beta_y}t$ motion appear in (51) and (52) because of the choice of the phases of the γ_x and γ_y motions given by eqs (19) and (23), respectively.

The $\cos 3\omega_{\gamma_y}t$ term of (52) is much smaller than the $\cos 2\omega_{\gamma_y}t$ term, and is therefore negligible. There is a systematic error term in the equilibrium angle: $-ef\tilde{\gamma}_y^2\omega_{\gamma_y}^4/[2b^2\omega_{\beta_y}^2(\omega_{\gamma_y}^2 - 4\omega_{\beta_y}^2)]$. This term is probably small for most precision balances and may in fact be an artifact of our initial neglect of terms in $\tilde{\gamma}_y^2$.

The size of the $\cos\omega_{\gamma_y}t$ modulation depends upon the knife edge radii ρ_C and ρ_L, and also upon the value of $\hat{\beta}_y$. The $\cos\omega_{\gamma_y}t$ modulations will be at a minimum for a given $\tilde{\gamma}_y$, when $\hat{\beta}_y$ is chosen (usually via an appropriate adjustment of M_A) such that the Mathieu term contribution is nearly equal and opposite to the knife-edges contribution. These two terms then nearly cancel, and one should observe pure $\cos 2\omega_{\gamma_y} t$ modulations. In most realistic cases, the knife-edges term can be safely ignored. (The radius of curvature of a good knife is usually less than 1 μm [12].) Then the Mathieu term vanishes when the equilibrium position of the beam is parallel to the gravitational horizon ($\hat{\beta}_y = 0$). Note that a test for the absence of a systematic error contribution to $\hat{\beta}_y$ does not *a fortiori* assure a negligible knife-edges contribution to the $\cos\omega_{\gamma_y} t$ modulations because the error term goes as $\tilde{\gamma}_y^2$, whereas the modulation term depends upon $\tilde{\gamma}_y$—and is therefore much larger.

5.2 A Damped Balance

5.2.1 Theoretical Predictions

We shall now investigate the motion of a damped balance at a time t which is long enough so that quantities involving the factor $e^{-\bar{k}z} = e^{-ct/2b}$ become negligible; (i.e., a time such that, in the absence of γ motion, the balance beam would be at rest at the equilibrium angle $\hat{\beta}_y$). Equation (31) thus becomes

$$\beta_y = \hat{\beta}_y + \frac{e}{b}\tilde{\gamma}_x^2 \frac{\omega_{\gamma_x}^2(\omega_{\beta_y}^2 - 4\omega_{\gamma_x}^2)}{\left[(\omega_{\beta_y}^2 - 4\omega_{\gamma_x}^2)^2 + 4\dfrac{c^2}{b^2}\omega_{\gamma_x}^2\right]}\cos 2\omega_{\gamma_x} t$$
$$+ \frac{2ce}{b^2}\tilde{\gamma}_x^2 \frac{\omega_{\gamma_x}^3}{\left[(\omega_{\beta_y}^2 - 4\omega_{\gamma_x}^2)^2 + \dfrac{4c^2}{b^2}\omega_{\gamma_x}^2\right]}\sin 2\omega_{\gamma_x} t, \tag{53}$$

and (50) becomes

$$\beta_y = \left[\hat{\beta}_y - \frac{ef}{2b^2}\tilde{\gamma}_y^2 \frac{\omega_{\gamma_y}^4}{\omega_{\beta_y}^2(\omega_{\gamma_y}^2 - 4\omega_{\beta_y}'^2)}\right]$$
$$+ \frac{e}{b}\hat{\beta}_y \tilde{\gamma}_y \frac{\omega_{\gamma_y}^2}{(\omega_{\gamma_y}^2 - 4\omega_{\beta_y}'^2)}\cos\omega_{\gamma_y} t$$
$$- \frac{f}{b}\tilde{\gamma}_y \frac{\omega_{\gamma_y}^2(\omega_{\beta_y}^2 - \omega_{\gamma_y}^2)}{\left[(\omega_{\beta_y}^2 - \omega_{\gamma_y}^2)^2 + \dfrac{c^2}{b^2}\omega_{\gamma_y}^2\right]}\cos\omega_{\gamma_y} t$$
$$- \frac{ce}{b^2}\hat{\beta}_y \tilde{\gamma}_y \frac{\omega_{\gamma_y}}{(\omega_{\gamma_y}^2 - 4\omega_{\beta_y}'^2)}\sin\omega_{\gamma_y} t$$
$$- \frac{cf}{b^2}\tilde{\gamma}_y \frac{\omega_{\gamma_y}^3}{\left[(\omega_{\beta_y}^2 - \omega_{\gamma_y}^2)^2 + \dfrac{c^2}{b^2}\omega_{\gamma_y}^2\right]}\sin\omega_{\gamma_y} t$$
$$+ \frac{e}{b}\tilde{\gamma}_y^2 \frac{\omega_{\gamma_y}^2(\omega_{\beta_y}^2 - 4\omega_{\gamma_y}^2)}{\left[(\omega_{\beta_y}^2 - 4\omega_{\gamma_y}^2)^2 + 4\dfrac{c^2}{b^2}\omega_{\gamma_y}^2\right]}\cos 2\omega_{\gamma_y} t$$
$$+ \frac{2ce}{b^2}\tilde{\gamma}_y^2 \frac{\omega_{\gamma_y}^3}{\left[(\omega_{\beta_y}^2 - 4\omega_{\gamma_y}^2)^2 + 4\dfrac{c^2}{b^2}\omega_{\gamma_y}^2\right]}\sin 2\omega_{\gamma_y} t$$

$$+\frac{e^2}{2b^2}\tilde{\gamma}_y^3\frac{\omega_{\gamma_y}^4(\omega_{\beta_y}^2-9\omega_{\gamma_y}^2)}{(\omega_{\gamma_y}^2-4\omega_{\beta_y}'^2)\left[(\omega_{\beta_y}^2-9\omega_{\gamma_y}^2)^2+9\frac{c^2}{b^2}\omega_{\gamma_y}^2\right]}\cos 3\omega_{\gamma_y}t$$

$$+\frac{3ce^2}{2b^3}\tilde{\gamma}_y^3\frac{\omega_{\gamma_y}^5}{(\omega_{\gamma_y}^2-4\omega_{\beta_y}'^2)\left[(\omega_{\beta_y}^2-9\omega_{\gamma_y}^2)^2+9\frac{c^2}{b^2}\omega_{\gamma_y}^2\right]}\sin 3\omega_{\gamma_y}t, \tag{54}$$

where $b = I_C$, $c = \alpha\mathscr{L}_B^2$, $e = M_AL_AL_{\mathrm{cm}}$, $f = M_AL_{\mathrm{cm}}(\rho_C-\rho_L)$, and $\omega_{\beta_y}' = 0$ at critical damping.

Sine modulation terms now appear in (53) and (54) because the damping effects introduce phase shifts into the balance beam responses to the pure cosine γ_x and γ_y motions. The $\sin\frac{n}{2}\omega_\gamma t$ terms are smaller than their equivalent $\cos\frac{n}{2}\omega_\gamma t$ terms by the ratio $\frac{-2c}{nb\omega_{\gamma_y}}$, where $n=2$, 4, or 6.

The $3\omega_{\gamma_y}t$ modulation terms are negligibly small compared with the $2\omega_{\gamma_y}t$ terms—as was the case for a freely-swinging, undamped balance. The presence of a systematic error term in the equilibrium angle can, in principle, be tested by observing the apparent $\hat{\beta}_y$ values for pure γ_x and γ_y motions. In practice, however, the over-simplifications of our model may cause problems in interpreting such a test. The modulations of frequency $\frac{\omega_{\gamma_y}}{2\pi}$ can again be minimized by choosing a nominal value of $\hat{\beta}_y$ such that terms involving the quantity $M_AL_AL_{\mathrm{cm}}\hat{\beta}_y\tilde{\gamma}_y$ are nearly equal in magnitude and opposite in sign to the knife-edges terms which include the factor $M_AL_{\mathrm{cm}}(\rho_C-\rho_L)\tilde{\gamma}_y$.

5.2.2 Experimental Verification

We have tested some of the predictions of section 5.2.1 by using a Mettler M5 balance,[3] which has a velocity-dependent damping produced by an air-dashpot. This balance has a maximum load of 20g, an on-scale range of 20 mg, and a readability of 2 μg. An analysis of some important features of this balance can be found in ref. [7]. The weighing pan is attached to the lower part of the suspension system by a hook and ring, so that the initial motion of the suspension system is that of a double pendulum. This motion quickly dampens to that of a single pendulum—which is a necessary condition for our model. The presence of the hook and ring, however, makes it difficult to generate pure γ_x or pure γ_y motion. We therefore temporarily locked the hook and ring together to provide a rigid suspension system for these tests. The tests described below were performed by gently starting the pan swinging at an amplitude of 4 ± 1 mrad about equilibrium.

What can be inferred from eq (54) about the dynamic behavior of this balance? First, eq (54) can be simplified because $\omega_{\beta_y}' \approx 0$ and $\omega_{\beta_y}^2 << \omega_{\gamma_y}^2$:

$$\beta_y = \text{const.}$$

$$+\frac{e}{b}\hat{\beta}_y\tilde{\gamma}_y\left[1+\left(\frac{c}{\mathrm{b}\omega_{\gamma_y}}\right)^2\right]^{1/2}\cos\left(\omega_{\gamma_y}t+\mathrm{Arctan}\frac{c}{b\omega_{\gamma_y}}\right) \tag{A}$$

$$+\frac{f}{b}\tilde{\gamma}_y\left[1+\left(\frac{c}{b\omega_{\gamma_y}}\right)^2\right]^{1/2}\cos\left(\omega_{\gamma_y}t+\mathrm{Arctan}\frac{c}{b\omega_{\gamma_y}}\right) \tag{B}$$

$$-\frac{e}{b}\frac{\tilde{\gamma}_y^2}{4}\left[1+\left(\frac{c}{2b\omega_{\gamma_y}}\right)^2\right]^{1/2}\cos\left(2\omega_{\gamma_y}t+\mathrm{Arctan}\frac{c}{2b\omega_{\gamma_y}}\right) \tag{C}$$

$$+\text{ higher order terms.} \tag{55}$$

---

[3] Brand names are used only for purposes of identification. Such use implies neither endorsement by the National Bureau of Standards nor assurance that the equipment is the best available.

Term A, the Mathieu term, arises from the tangential acceleration of the swinging pan. Term B is a consequence of having knife edges of finite radius. The last term considered, C, has its origin in the centripetal acceleration of the pan. The latter is the only term of importance in eq (52).

We have made the following estimates for the balance being considered:

$$
\begin{aligned}
e &\sim 5 \times 10^3 \text{ g cm}^2 \\
|f| &\leqslant 2.5 \times 10^{-2} \text{ g cm}^2 \\
b &\sim 3.5 \times 10^3 \text{ g cm}^2 \\
c &\sim 5.5 \times 10^3 \text{ g cm}^2 \text{ s}^{-1} .
\end{aligned}
$$

Also, ω_{γ_y} was measured to be 7.8 rad/s.

On the basis of these numbers, one may observe that term B always has much smaller amplitude than term C and is, therefore, always masked. We are left only with terms A and C. Equation (55) predicts that term A will dominate when $|\hat{\beta}_y| >> \hat{\gamma}_y$ but that term C will dominate when the inequality is reversed. Thus, under the condition that $|\hat{\beta}_y| >> \hat{\gamma}_y$ we would expect an oscillation in the beam at the same frequency as the pan oscillation whereas the oscillation frequency of the beam should double when $\hat{\beta}_y = 0$. The ratio of the amplitudes in the two regimes is predicted to be

$$4.4\ \hat{\beta}/\hat{\gamma}_y . \tag{56}$$

Note that this result is independent of the estimate of e given above and only weakly dependent on c and b. These predictions were tested in the following way.

With the balance arrested, a load of nearly 20g was placed on the pan. When the balance was released, the optical scale reading was found to be near zero (i.e., $\hat{\beta}_y$ was near its maximum value of $+22$ mrad [7]). Pure γ_x motion produced a small modulation of the balance beam read-out angle, β_y, at a frequency twice that of the suspension system—as predicted by eq (53)—at an amplitude of 2 ± 1 μg. Pure γ_y motion yielded a modulation having an amplitude of 32 ± 2 μg—with a frequency which was the same as that of the suspension system. The phase shift due to damping effects was too small to observe, so that this modulation was a nearly pure $\cos\omega_{\gamma_y} t$ function. Our measurements could have detected a phase shift as small as $\pi/6$ but eq (55) predicts a phase shift of about $\pi/16$.

We then changed the sign of $\hat{\beta}_y$ by adding 20 mg of tare weights, thereby moving the equilibrium of the beam near to the maximum on-scale reading (i.e., $\hat{\beta}_y$ near its minimum value of -22 mrad [7]). The result of pure γ_y motion at this position produced a 180° phase shift in the $\cos\omega_{\gamma_y} t$ modulation of β_y as predicted by eq (55).

The magnitude of M_A was then reduced by removing 10 mg so that the equilibrium position of the beam was at the middle of the optical scale. At this position, where the beam is nearly horizontal, the Mathieu-term contribution of the $\cos\omega_{\gamma_y} t$ modulation nearly vanishes. The β_y modulation now became that of the small $\cos 2\omega_{\gamma_y} t$ contribution (in agreement with the case of γ_x motion.)

These tests clearly verify several important predictions of eqs (53) and (55). The observed ratio of amplitudes in the two frequency regimes is 16, which agrees satisfactorily with the value 24 predicted by (56).

6. Conclusions

In summation, it will be useful to recall some of the important simplifications in our mathematical model. All knives have been assumed to be symmetrically honed and to roll on perfect flats. The two main knives of the balance are assumed to contact their respective flats in straight lines which are parallel to each other and to the x-y plane. Many authors have dealt with static errors which result from a violation of those conditions (e.g., [5, 11, 12]). An analysis of the corresponding dynamic errors is clearly beyond the scope of this paper. We have also chosen to look at pan oscillations in the context of the small-angle approximation of an idealized balance. While we believe this is appropriate, it should be mentioned that small variations in balance sensitivity as a function of $\hat{\beta}_y^2$ can be observed in many analytical balances having a large on-scale range [7]. Such behavior cannot be predicted from our equations, which ignore angular terms higher than first order.

The above considerations serve as a caveat to the quantitative application of our results to real balances. Nevertheless, the solutions found have proved valuable. The question posed in the introduction has been answered: Two processes are important in analyzing the effect of a swinging pan on balance read-out. One, which has been mentioned in the past [2], is due to centripetal acceleration of the pan. The second—and often predominant—process is a more subtle effect by which the tangential acceleration of the swinging pan leads to a torque on the beam if the latter is not at the gravitational horizon and if the former has a component of motion in the *y-z* plane. Even a crude approximation of the tangential acceleration (e.g., a_T(max.) ~ $g\hat{\gamma}$) suggests that its influence will exceed that of the centripetal acceleration if $\hat{\beta} > \bar{\gamma}$. A convincing argument, however, can only be advanced through the rigorous mathematical solution to the equations of motion as presented in section 4.

The disappearance of this second, or "Mathieu" process when the rest-angle of the beam is near gravitational horizon may be used to minimize the effect. Alternatively, it may be possible to use the effect to adjust the beam angle in a class of experiments [13, 14, 15] where it is advantageous that the beam be parallel to gravitational horizon. Here the radii of the knives must be determined (under load) in a separate experiment [12].

7. References

[1] Almer, H.E. One kilogram balance NBS No. 2, J. Res. Nat. Bur. Stand. (U.S.). **76C**(1 and 2): 1–10; 1972 January–June. The balance described in this reference is now housed at the Bureau International des Poids et Mesures where it achieves a standard deviation of 1 μg in 1 kg.

[2] Bowman, H.A.; Macurdy, L.B. Gimbal device to minimize the effects of off-center loading on balance pans. J. Res. Nat. Bur. Stand. (U.S.). **64C**(4): 277–279; 1960 October–December.

[3] Goldstein, H. *Classical mechanics*. Massachusetts: Addison–Wesley Publishing Co., Inc.; 1950. 399 p. See especially pp 14–18.

[4] Wells, D.A. *Theory and problems of Lagrangian dynamics*. New York: Schaum Publishing Co.; 1967. See especially pp 58–62.

[5] Bowman, H.A.; Almer, H.E. Minimization of the arrestment error in one-pan, two-knife systems. J. Res. Nat. Bur. Stand. (U.S.). **67C**(3): 227–235; 1963 July–September.

[6] Waldersee, J. The effect of the finite radii of the knife edges on the performance of an equi-arm balance. Australian J. Appl. Sc.. **6**: 158–166; 1955.

[7] Bowman, Horace A.; Schoonover, Randall M.; [with Appendix by Jones, Mildred W.]. Procedure for high precision density determinations by hydrostatic weighing. J. Res. Nat. Bur. Stand. (U.S.). **71C**(3): 179–198; 1967 July–August.

[8] Stephenson, Reginald J. *Mechanics and properties of matter*. New York: John Wiley and Sons, Inc.; 1952. 371 p. See especially pp 122–129.

[9] Abramowitz, M.; Segun, I.A. *Handbook of mathematical functions*. New York: Dover Publications, Inc.; 1968. 1,046 p. See especially pp 722–746.

[10] McLachlan, N.W. *Theory and application of Mathieu functions*. England: Clarendon Press; 1947. 401 p.

[11] Thiesen, J. Etudes sur la balance. Trav. et Mem. Bureau International des Poids et Mesures 5: 3–40; 1886.

[12] Corwin, A.H. Weighing in technique of organic chemistry, part 1 in *Physical methods of organic chemistry*, vol. 1. Arnold Weissberger, ed. New York: Interscience Publishers, Inc.; 1959. 71–129.

[13] Cage, M.E. An analysis of the effects of dynamic and static forces present in the NBS SI volt experiment. Nat. Bur. Stand. (U.S.) NBSIR 80-2143; 1980 September. 218 p.

[14] Olsen, P.T.; Phillips, W.D.; Williams, E.R. A proposed coil system for the improved realization of the absolute ampere. J. Res. Nat. Bur. Stand. (U.S.) **85**(4): 257–272; 1980 July–August.

[15] Olsen, P.T.; Cage, M.E.; Phillips, W.D.; Williams, E.R. The realization of the ampere at NBS. IEEE Trans. on Instr. and Meas. **29**: 234–237; 1980.

Appendix A.
Algebraic substitutions and changes of variables

This appendix lists all of the algebraic substitutions and changes of variables used in the derivations. Therefore, any equation given in the main text can be readily expressed in terms of the basic quantities γ_x, γ_y, β_y, L_{cm}, L_A, ℓ_b, L_B, $\mathscr{L}_B$, ℓ_B, M_A, M_b, M_B, g, α, t, ρ_C, ρ_L, and ρ_l—as defined in figures 1–4.

$$\kappa = [M_A\, g(\rho_L + \rho_C) + M_b\, g(\ell_b + \rho_C) + M_B\, g(\ell_B + \rho_C)]$$

$$S = \frac{1}{\kappa}$$

$$F_\alpha = -\alpha \mathcal{L}_B\, \dot{\beta}_y$$

$$I_C = M_A L_A^2 + [(I_b)_{\mathrm{cm}} + M_b \ell_{\mathrm{cm}}^2] + [(I_B)_{\mathrm{cm}} + M_B(L_B^2 + \ell_B^2)]$$

$$I_P = (I_S)_{\mathrm{cm}} + M_A L_{\mathrm{cm}}^2$$

$$I'_P = (I'_S)_{\mathrm{cm}} + M_A L_{\mathrm{cm}}^2$$

$$\omega_{\gamma_x}^2 = \frac{[M_A g L_{\mathrm{cm}}]}{I'_P}$$

$$\omega_{\gamma_y}^2 = \frac{[M_A g L_{\mathrm{cm}}]}{I_P}$$

$$\omega_{\gamma_x} = \frac{2\pi}{P_{\gamma_x}}$$

$$\omega_{\gamma_y} = \frac{2\pi}{P_{\gamma_y}}$$

$$\gamma_x = -\hat{\gamma}_x \cos \omega_{\gamma_x} t$$

$$\gamma_y = -\hat{\gamma}_y \cos \omega_{\gamma_y} t$$

$$b = I_C$$

$$c = \alpha \mathcal{L}_B^2$$

$$d = [M_B g L_B - M_A g L_A]$$

$$e = M_A L_A L_{\mathrm{cm}}$$

$$f = M_A L_{\mathrm{cm}}(\rho_C - \rho_L)$$

$$\hat{\beta}_y = \frac{d}{\kappa} = \frac{[M_B g L_B - M_A g L_A]}{[M_A g(\rho_L + \rho_C) + M_b g(\ell_b + \rho_C) + M_B g(\ell_B + \rho_C)]}$$

$$\omega'_{\beta_y} = \sqrt{\frac{\kappa}{b} - \frac{c^2}{4b^2}} = \frac{2\pi}{P'_{\beta_y}}$$

$$\omega_{\beta_y} = \sqrt{\frac{\kappa}{b}} = \frac{2\pi}{P_{\beta_y}}$$

$$\bar{u} = \frac{c}{b} = \frac{\alpha \mathcal{L}_B^2}{I_C}$$

$$v = \sqrt{\frac{\kappa}{b}} = \omega_{\beta_y}$$

$$w = \frac{e}{b}\, \hat{\gamma}_x^2\, \omega_{\gamma_x}^2 = \frac{M_A L_A L_{\mathrm{cm}}}{I_C}\, \hat{\gamma}_x^2\, \omega_{\gamma_x}^2$$

$$k = 2\omega_{\gamma_x}$$

$$D = \frac{\bar{u} w k}{[(v^2 - k^2)^2 + \bar{u}^2 k^2]}$$

$$E = \frac{w(v^2 - k^2)}{[(v^2 - k^2)^2 + \bar{u}^2 k^2]}$$

$$z = \frac{\omega_{\gamma_y}}{2} t$$

$$\bar{k} = \frac{c}{b\omega_{\gamma_y}} = \frac{\alpha \mathscr{L}_B^2}{I_C \omega_{\gamma_y}}$$

$$\bar{a} = \frac{4\omega_{\beta_y}^2}{\omega_{\gamma_y}^2}$$

$$q = -2 \frac{e}{b} \tilde{\gamma}_y = -2 \frac{M_A L_A L_{\mathrm{cm}}}{I_C} \tilde{\gamma}_y$$

$$h = \frac{4d}{\omega_{\gamma_y}^2 b} = 4 \frac{[M_B g L_B - M_A g L_A]}{I_C \omega_{\gamma_y}^2}$$

$$\ell = 4 \frac{e}{b} \tilde{\gamma}_y^2 = 4 \frac{M_A L_A L_{\mathrm{cm}}}{I_C} \tilde{\gamma}_y^2$$

$$m = 4 \frac{f}{b} \tilde{\gamma}_y = 4 \frac{M_A L_{\mathrm{cm}}}{I_C} (\rho_C - \rho_L) \tilde{\gamma}_y$$

$$\bar{f}(z) = h + \ell \cos 4z - m \cos 2z$$

$$\beta_y = e^{-\bar{k}z} u(z)$$

$$a = \bar{a} - \bar{k}^2 = \frac{4}{\omega_{\gamma_y}^2} \left[\frac{\kappa}{b} - \frac{c^2}{4b^2} \right] = 4 \frac{\omega'^2_{\beta_y}}{\omega_{\gamma_y}^2}$$

$$\omega'_{\beta_y} = 0 \quad (\text{if } a = 0)$$

$$\bar{q} = \frac{q}{2(1 - \nu^2)} \approx -\frac{e}{b} \frac{\omega_{\gamma_y}^2 \tilde{\gamma}_y}{(\omega_{\gamma_y}^2 - 4\omega'^2_{\beta_y})}$$

$$\nu \approx \sqrt{a} \approx 2 \frac{\omega'_{\beta_y}}{\omega_{\gamma_y}} \quad (\text{if } a > 0)$$

$$\nu \approx \frac{q}{\sqrt{2}} \approx \sqrt{2}\bar{q} \approx -\sqrt{2}\frac{e}{b} \tilde{\gamma}_y \quad (\text{if } a = 0)$$

$$u_1 = \cos \nu z - \bar{q} \cos \nu z \cos 2z - \nu \bar{q} \sin \nu z \sin 2z$$

$$u_2 = \sin \nu z - \bar{q} \sin \nu z \cos 2z + \nu \bar{q} \cos \nu z \sin 2z$$

$$\bar{c}^2 = u_1 u_2' - u_2 u_1' = \nu$$

$$\nu z = \omega'_{\beta_y} t$$

$$2z = \omega_{\gamma_y} t$$

$$\bar{k}z = \frac{ct}{2b}$$

$$\nu \bar{q} = -2 \frac{e}{b} \frac{\omega_{\gamma_y} \omega'_{\beta_y} \tilde{\gamma}_y}{(\omega_{\gamma_y}^2 - 4\omega'^2_{\beta_y})}$$

Appendix B

An aid to identifying terms in eq (50)

This appendix provides some relationships that can aid the reader in transforming the terms in eq. (50) for γ_y and β_y motions into the notation used in (28) for γ_x and β_y motions. The combination

$$\frac{X}{[\bar{k}^2 + (\nu + n)^2]} + \frac{Y}{[\bar{k}^2 + (\nu - n)^2]} \tag{B1}$$

often occurs in (50), where $n = 0, 2, 4$ or 6. By using the common denominator

$$[(\bar{a} - n^2)^2 - 4n^2a] = \frac{16}{\omega_{\gamma_y}^4}\left[\left(\omega_{\beta_y}^2 - \frac{n^2}{4}\,\omega_{\gamma_y}^2\right)^2 + \frac{n^2}{4}\frac{c^2}{b^2}\,\omega_{\gamma_y}^2\right],$$

(B1) can be written as

$$-\frac{n\nu\bar{k}\omega_{\gamma_y}^4}{4\left[\left(\omega_{\beta_y}^2 - \frac{n^2}{4}\,\omega_{\gamma_y}^2\right)^2 + \frac{n^2c^2}{4\,b^2}\,\omega_{\gamma_y}^2\right]} \quad (\text{if } X = \bar{k} \text{ and } Y = -\bar{k}); \tag{B2}$$

as

$$\frac{\bar{k}(4\omega_{\beta_y}^2 + n^2\omega_{\gamma_y}^2)\omega_{\gamma_y}^2}{8\left[\left(\omega_{\beta_y}^2 - \frac{n^2}{4}\,\omega_{\gamma_y}^2\right)^2 + \frac{n^2}{4}\frac{c^2}{b^2}\,\omega_{\gamma_y}^2\right]} \quad (\text{if } X = \bar{k} \text{ and } Y = \bar{k}); \tag{B3}$$

as

$$\frac{\nu(4\omega_{\beta_y}^2 - n^2\omega_{\gamma_y}^2)\omega_{\gamma_y}^2}{8\left[\left(\omega_{\beta_y}^2 - \frac{n^2}{4}\,\omega_{\gamma_y}^2\right)^2 + \frac{n^2}{4}\frac{c^2}{b^2}\,\omega_{\gamma_y}^2\right]} \quad (\text{if } X = (\nu + n) \text{ and } Y = (\nu - n)); \tag{B4}$$

or as

$$\frac{n}{8}\frac{(4\omega_{\beta_y}^2 + n^2\omega_{\gamma_y}^2 - 8\omega'^2_{\beta_y})\omega_{\gamma_y}^2}{\left[\left(\omega_{\beta_y}^2 - \frac{n^2}{4}\,\omega_{\gamma_y}^2\right)^2 + \frac{n^2}{4}\frac{c^2}{b^2}\,\omega_{\gamma_y}^2\right]} \quad (\text{if } X = (\nu + n) \text{ and } Y = -(\nu - n)). \tag{B5}$$

With the help of eqs (B1)–(B5) and Appendix A, the terms of (50) can be expressed in the notation of (28). For example:

$$\frac{m\bar{k}}{2\nu}\frac{\sin 2z}{[\bar{k}^2 + (\nu + 2)^2]} - \frac{m\bar{k}\sin 2z}{2\nu[\bar{k}^2 + (\nu - 2)^2]} = -\frac{cf}{b^2}\,\hat{\gamma}_y\,\frac{\omega_{\gamma_y}^3}{\left[(\omega_{\beta_y}^2 - \omega_{\gamma_y}^2)^2 + \frac{c^2}{b^2}\,\omega_{\gamma_y}^2\right]}\sin\omega_{\gamma_y}t,$$

and

$$-\frac{\nu\bar{q}h\bar{k}\sin 2z}{2\nu[\bar{k}^2 + (\nu + 2)^2]} - \frac{\nu\bar{q}h\bar{k}\sin 2z}{2\nu[\bar{k}^2 + (\nu - 2)^2]} = \frac{ce}{b^2}\,\hat{\gamma}_y\,\frac{\hat{\beta}_y\omega_{\beta_y}^2\omega_{\gamma_y}(\omega_{\beta_y}^2 + \omega_{\gamma_y}^2)}{(\omega_{\gamma_y}^2 - 4\omega'^2_{\beta_y})\left[(\omega_{\beta_y}^2 - \omega_{\gamma_y}^2)^2 + \frac{c^2}{b^2}\,\omega_{\gamma_y}^2\right]}\sin\omega_{\gamma_y}t\,.$$

JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 87, No. 1, January–February 1982

A 30 kg Capacity High Precision Load Cell Mass Comparator

Randall M. Schoonover*

National Bureau of Standards, Washington, DC 20234

April 15, 1981

Described here are simple means to fabricate a 30 kg mass comparator based on an ordinary direct reading load cell. The mass comparator performs with a precision of 1 ppm.

Key Words: Constant loading; high precision; load cell; mass comparator; substitution weighing; weighing.

1. Introduction

A paper [1][1] describing a high precision load cell mass comparator was published in 1979. The principle of that device was the maintenance of a spring force on the active load cell element nearly equal to the gravitational forces of the weights being compared, even during the period when weights were exchanged. Doing so caused an ordinary load cell which had a precision of 1 part in 10,000 as a direct reading instrument to perform with a precision of a few parts per million (ppm) as a mass comparator. This work was with loads of 225 kg, and it was speculated that scaling the method to 30 kg would be quite useful in small-mass metrology. Personal communication with several members of the measurement community, however, revealed their unanimous opinion that such scaling would result in serious loss of precision.

The work reported here results from the successful fabrication of a 30 kg mass comparator based on the above principle that has a precision of about 1 ppm.

2. The Instrument

The comparator differs significantly in design from the 225 kg version in several ways. The instrument is self-supporting and is provided with a built-in weight exchanger as shown in figure 1. This feature not only loads the cell without shock, but also aligns the weight in the center of the weighing pan. The cell itself is supported by four parallel springs in tension rather than a single spring in compression as before. Finally, the flexure universals above and below the cell are replaced with gimbaled joints fabricated from ball bearing assemblies. The load cell incorporates a solid state bridge and has a capacity of 45 kg (100 lb.). A schematic view of the load cell and spring assembly is shown in figure 2.

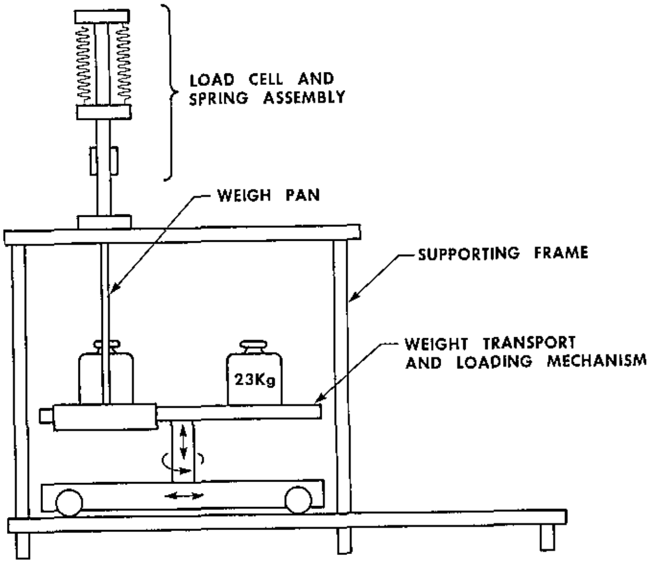

FIGURE 1. Cross-sectional view showing the essential components of the constant loading mechanism.

It is noteworthy to report one major operational change in the use of the comparator. Unlike the previous instrument the springs are used to support the load cell during the weighing mode, whereas before a mechanical shunt intervened and supported the cell. This method of supporting the cell appears to improve the isolation from ambient vibration as has been observed independently by others [2]. In addition nearly 100 percent of the load is maintained on the cell at all times.

3. Test Results

The test weighings were made in the usual manner, that is, six double substitutions between two 23 kg weights (50 lb.) comprised one test. The data was reduced and the standard deviation was calculated in the usual way.

---

* Center for Absolute Physical Quantities, National Measurement Laboratory.

[1] Figures in baackets indicate literature references at the end of this paper.

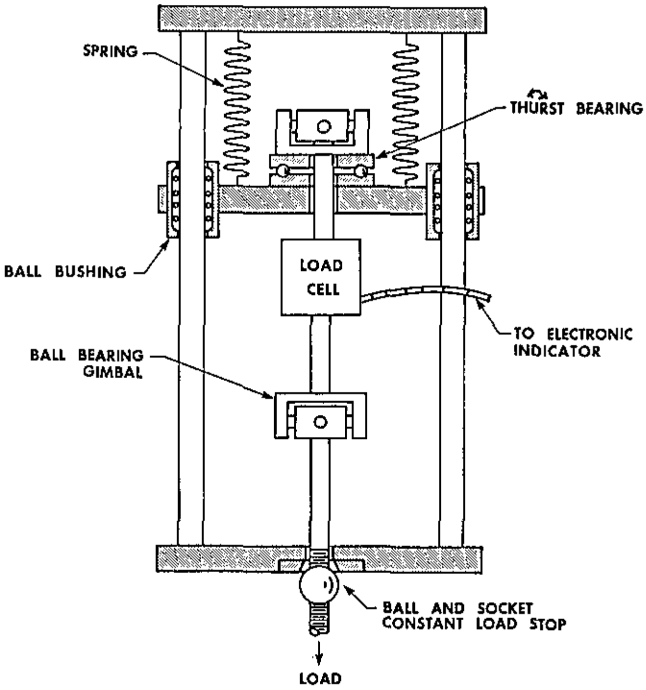

FIGURE 2. A schematic view of the complete weighing system.

Three different electronic load cell voltage measuring instruments were used in collecting the data and the pooled standard deviation, SD_p, for each group of three tests is given in the following table.

Current Bridge Used in Ref. 1	DJ101 [3]	Cutkosky & Davis [4] Bridge
SD_p 58 mg	24 mg	30 mg

The Cutkosky and Davis circuit was especially adapted for this application. That data was supplied by R. S. Davis (personal communication).

During the above testing the mass comparator was uncased and exposed to the detrimental effects, if any, of air circulation in the laboratory.

4. Discussion

A standard deviation of 24 mg for a 30 kg comparator is about 1 ppm of the applied load. A review of commercial mechanical instruments of this capacity that contain knife edges or flexure bearings indicate that the units from a given manufacturer have a standard deviation, when used as a mass comparator, which varies between 1 and 20 mg. It appears to the author, that if more effort is given to the load cell and associated electronic indicator design and careful attention is paid to ambient vibration isolation, this device would outperform the conventional mass comparator and operate with a precision a few parts in 10^7 or better. This instrument would be more rugged, would reduce the measurement time, provide much more on-scale range, and, because of its mechanical simplicity be of substantially lower cost.

The author wishes to thank Albert Tholen of the National Bureau of Standards Office of Weights and Measures for financial support for this work.

5. References

[1] Schoonover, Randall M. A high precision load cell mass comparator. J. Res. Nat. Bur. Stand. (U.S.). **84**(5): 347–351; 1979 September-October.

[2] Page, Dennis. White Sands Missile Range, personal communication.

[3] The DJ101 is a commercial device manufactured by DJ Instruments, Inc., and was on loan for this test from NBS Division 737. Certain trade names and company products are identified in order to adequately specify the experimental procedure. In no case does such identification imply *recommendation or endorsement by the National Bureau of Standards* nor does it imply that the products are necessarily the best available for the purpose.

[4] Cutkosky, R.D.; Davis, R.S. A simple control circuit for temperature regulation and other bridge applications. Accepted for publication in Rev. Sci. Instr.

JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 87, No. 1, January–February 1982

Electron Impact Ionization of Lithium

S. M. Younger*†

National Bureau of Standards, Washington, DC 20234

August 18, 1981

The electron impact ionization cross section of the neutral lithium atom has been calculated in a distorted wave exchange approximation. The total cross section is in good agreement with available experimental data at incident electron energies above 10 eV. Analytic fits are provided for the 1*s* and 2*s* subshell partial cross sections.

Key Words: Electron-atom scattering; electron impact ionization; lithium atom.

1. Introduction

Cross sections for the electron impact ionization of lithium atoms are required for studies of neutral beam injection in thermonuclear plasma machines. Experimental electron impact ionization cross sections for neutral lithium have been reported by Jalin et al. [1] for the incident electron energy range 100 to 2000 eV and by Zapesochnyi and Aleksakhin [2] for the threshold to 30 eV range. Earlier experiments include the relative measurements of Brink [3] and the absolute cross sections of McFarland and Kinney [4], both restricted to energies well above threshold. Theoretical calculations [5, 6, 7], mainly in the plane wave Born approximation, support the high energy data of Jalin et al. [1], which is more than a factor of two below the McFarland and Kinney [4] results.

The present work describes the application of the distorted wave exchange approximation to the calculation of the neutral lithium electron ionization cross section. This method, which has been described in detail in previous publications [8, 9, 10], is an improvement over previous theoretical work on lithium in that: (1) it utilizes a more accurate ground state target wavefunction, (2) it allows for potential distortion effects for all three continuum electrons involved in the ionization process, (3) it considers the effect of scattering exchange, and (4) it includes inner shell ionization.

2. Technique and Results

The distorted wave exchange approximation used in the present work has been described in detail in previous publications [8–10]. The target was described by the Hartree-Fock ground state wavefunction given by Clementi and Roetti [11]. The ionization energies for the ejection of a 2*s* and a 1*s* electron were $I_{2S} = 5.393$ eV and $I_{1S} = 64.9$ eV. The incident and final scattered partial waves in the direct matrix element were computed in the static local potential of the ground state plus a semiclassical energy dependent exchange potential. The ejected partial wave was computed in the potential of the ion with a 2*s* or 1*s* vacancy depending on whether outer or inner shell ionization was being considered. For the exchange matrix elements the final state partial wave potentials were reversed, and the phase of the matrix element was chosen so as to maximize the effect of exchange and hence minimize the total cross section [8]. The maximum partial wave orbital angular momenta considered were 14 for the incident and scattered waves and 10 for the ejected waves, sufficient to obtain convergence of the partial wave series.

The total cross sections for outer and inner shell ionization of Li I are given in table 1.

Rather than calculate individual cross sections at high incident electron energy, which would be costly due to the large number of partial waves involved at high energies, we have used a fitting procedure [12] to obtain analytic expressions for the cross sections based on low energy distorted wave data and a Bethe slope derived from the photoionization calculations of McDowell and Chang [13] for the 2*s* subshell

TABLE 1. *Electron impact ionization cross sections for Lithium* (10^{-16} *cm*^{2})

1*s* subshell			2*s* subshell		
u_{1s}	Exchange	No-Exchange	u_{2s}	Exchange	No-exchange
1.25	0.0155	0.0184	1.25	2.95	2.50
1.50	.0292	.0356	1.50	3.91	3.53
2.25	.0523	.0632	2.25	4.51	4.35
3.50	.0627	.0711	3.50	4.01	4.35
4.00	.0629	.0729	4.00	3.76	4.16
5.00	.0611	.0696	5.00	3.35	3.76

---

* Center for Radiation Research, National Measurement Laboratory.

†Supported in part by the Department of Energy, Office of Fusion Energy.

and of Manson [14] for the 1*s* subshell. The resulting expressions describing the electron impact ionization of lithium are

$$EQ_{2s} = 251.\left(1-\frac{1}{u_{2s}}\right) - 121.\left(1-\frac{1}{u_{2s}}\right)^2 + 11.1\,\ell n u_{2s} - \frac{157.}{u_{2s}}\,\ell n u_{2s} \qquad (1)$$

and

$$EQ_{1s} = 26.9\left(1-\frac{1}{u_{1s}}\right) - 12.5\left(1-\frac{1}{u_{1s}}\right)^2 + 10.7\,\ell n u_{1s} - \frac{33.7}{u_{1s}}\,\ell n u_{1s} \qquad (2)$$

where $u_{1s} = E/I_{1s}$ and $u_{2s} = E/I_{2s}$ with the incident electron energy, *E*, in eV. *Q* is the cross section in 10^{-16} cm^2.

Figure 1 compares the present results with the available experimental data and other theoretical calculations. The distorted wave data are in good agreement with the more recent experiments, except for incident electron energies below about 10 eV ($u_{2s} < 1.85$) where a more complex interaction between the target and scattering wave is present. The Born approximation calculations of McGuire [5] and of Peach [6] are 10–50 percent lower than the present results, and are in slightly better agreement with experiments at very low and very high incident electron energies. McGuire's calculation includes the contribution from 1*s* ionization; Peach's results are for 2*s* ionization only.

The distorted wave method is essentially a model potential scattering theory which assumes that the target atomic structure is unaffected by the presence of the scattered electrons. While numerous attempts have been made to include polarization effects, electron correlation and even relativistic corrections in a distorted wave model, it is not clear to what extent such modifications are true improvements in such a simple theory versus attempts to obtain a better fit to numerical data. The present choice of a direct local potential with a semiclassical exchange potential has been found to yield partial wave phase shifts in remarkable agreement with frozen core Hartree-Fock calculations, and to produce reasonably accurate total ionization cross sections for a wide variety of atoms and ions [8–10, 12]. Attempts to improve the distorted wave exchange ionization theory, specifically by

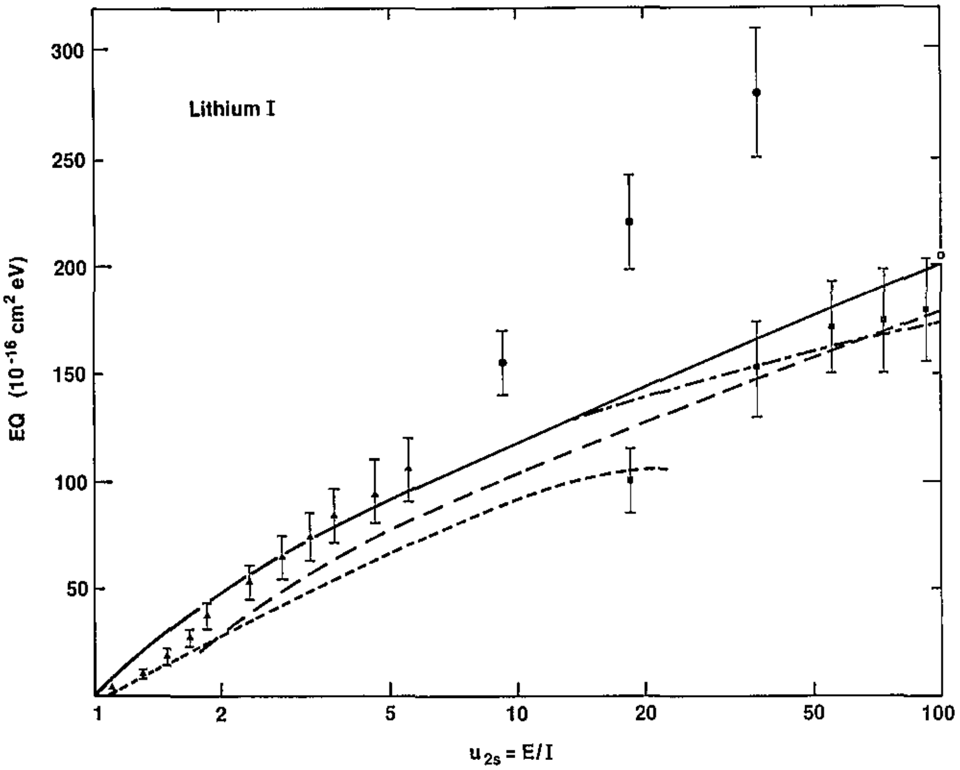

FIGURE 1. Fano plot of the electron impact ionization cross section of lithium. ———total ionization cross section computed in the distorted wave exchange approximation (present); —·—·—2*s* ionization cross section, distorted wave exchange approximation (present); ———Plane wave Born, Ref. [5]; ----Plane wave Born, Ref. 6; ○ Bethe theory, Ref. 7; ● Crossed-beam experiment, Ref. [4]; ▲ Crossed-beam experiment, Ref. 2; Crossed-beam experiment, Ref. [1]

means of polarization potentials, have met with mixed success, improving agreement with experiment in some cases and causing greater disagreement in others. The present results are felt to represent a realistic compromise between the simplicity of the plane wave Born approximation and the rigors of a full many body approach.

3. References

[1] Jalin, R., Hagemann, R., and Botter, R., J. Chem. Phys. **59**, 952 (1973).

[2] Zapesochyni, I.P., and Aleksakhin, I.S., Zh. Eksp. Teor. Fiz **55**, 76 (1968) [JETP **28**, 41 (1969)].

[3] Brink, G.O., Phys. Rev. **134**, A345 (1964).

[4] McFarland, R.H., and Kinney, J.R., Phys. Rev. **137**, 1058 (1965).

[5] McGuire, E.J., Phys. Rev. A **3**, 267 (1971).

[6] Peach, G., Proc. Phys. Soc. **87**, 381 (1966).

[7] Kim, Y.K., and Inokuti, M., Phys. Rev. A **3**, 665 (1971).

[8] Younger, S.M., Phys. Rev. A **22**, 111 (1980).

[9] Younger, S.M., Phys. Rev. A **22**, 1425 (1980).

[10] Younger, S.M., Phys. Rev. A **23**, 1138 (1981).

[11] Clementi, E., and Roetti, C., At. Data Nuc. Data Tables **14**, 177 (1974).

[12] Younger, S.M., J. Quant. Spectrosc. Rad. Transfer, (1981).

[13] McDowell, M.R.C., and Chang, E.S., Mon. Not. R. Astron. Soc. **142**, 465 (1969).

[14] Manson, S., private communication.

JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 87, No. 1, January–February 1982

An Approach to Peak Area Estimation

John Rice

Department of Mathematics, University of California at San Diego, LaJolla, CA 92093

and

National Bureau of Standards, Washington, DC 20234*

May 27, 1981

We consider the problem, arising in nuclear spectroscopy, of estimating peak areas in the presence of a baseline of unknown shape. We analyze a procedure that chooses the baseline to be as smooth as is consistent with the data and note that the estimates have a certain minimax optimality. Expressions are developed for the systematic and random errors of the estimate, and some large sample approximations are derived. Procedures for choosing a smoothing parameter are developed and illustrated by simulations.

Key words: linear models; minimax; peak area; smoothing; spectroscopy; splines.

1. Introduction

The estimation of peak area in the presence of a baseline of unknown shape is a common problem in nuclear and other spectroscopies. In this paper we analyze some of the properties of a generalization of a procedure proposed by Currie [2][1] and note that the procedure has a certain minimax optimality.

We first introduce the problem and some notation. We suppose that counts are accumulated in n channels over a length of time T, and that the total number of counts has mean $\mu = \nu T$, where ν = mean counting rate per unit time. We let y_j denote the proportional count in the j^{th} channel, i.e. the total count in the j^{th} channel divided by μ, and we assume that

$$y_j = \beta_0\gamma_j + \beta_j + \varepsilon_j, j = 1, \ldots, n$$

Here, $\Gamma = (\gamma_1, \ldots, \gamma_n)^T$ is a vector representing a peak shape, which is assumed known (Γ might be known from theory or from measurement of pure specimens, for example), β_0 is its unknown amplitude, which we wish to determine, and β_j is the unknown baseline mean in the j^{th} channel. The ε_j's are random counting errors with mean zero and nonsingular covariance matrix $\mu^{-1}W^{-1}$ where W is a matrix which is assumed to be known. (In applications, W is typically estimated rather than known. An application of the δ-method [7] to the perturbation thus introduced shows that the asymptotic means and variances are unchanged.) In vector notation the model can be written

$$\begin{aligned} Y &= [\Gamma{:}I]\beta + \varepsilon \\ &= A\beta + \varepsilon \end{aligned}$$

where $Y = (y_1, \ldots, y_n)^T$, $\beta = (\beta_0, \beta_1, \ldots, \beta_n)^T$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$. We note that this model is underdetermined, and that even in the limit, with no counting error, there is no unique solution for β_0.

Currie [2] proposed estimating β by forcing the baseline to be as smooth as is consistent with the data (in a sense explained below), taking as measures of smoothness

---

* Statistical Engineering Division, Center for Applied Mathematics.

[1] Figures in brackets refer to literature references at the end of this paper.

$$S_1 = \sum_{j=1}^{n-1} (\beta_i - \beta_{i+i})^2$$

or

$$S_2 = \sum_{j=1}^{n-2} (\beta_i - 2\beta_{i+1} + \beta_{i+2})^2$$

or generally

$$S_k = \sum_{j=1}^{n-k} (\Delta^k \beta)^2$$

where Δ is a differencing operator. The estimate $\hat{\beta}$ is formed by minimizing S_k subject to the constraint

$$(Y - A\beta)^T W (Y - A\beta) = c$$

where the constraint c is obtained from the χ^2 distribution. Using the technique of Lagrange multipliers, the solution is found to be

$$\beta = (A^T WA + \lambda U^T U)^{-1} A^T WY$$

when λ is chosen to force β to satisfy the constraint and S_k is expressed as

$$S_k = \|U\beta\|^2.$$

By considering numerical examples, Currie reached some empirical conclusions about the statistical behavior of the method, with special attention to the bias, or systematic error, of the method.

Techniques of this kind have been used in solving ill-posed problems such as integral equations of the first kind [1] and in smoothing data via smoothing splines [8, 11]. Motivated by such problems, Kuks and Olman [5] and Speckman [9] have considered the problem of estimating a linear functional $h^T\beta$ by linear functionals of the data, $\ell^T Y$. Their result is the following: Consider the linear model

$$Y = A\beta + \varepsilon$$

where ε has a nonsingular covariance matrix $\sigma^2 W^{-1}$, and assume that $\|U\beta\|^2 \leq \alpha^2$ for some matrix U such that $N(U) \cap N(A) = \phi$ ($N(A)$ = null space of A). Then the estimate $\ell_0^T Y$ for which

$$E(\ell_0^T Y - h^T\beta)^2 = \min_{\ell} \max_{\|U\beta\|^2 \leq \alpha^2} E\,(\ell_0^T Y - h^T\beta)^2$$

is unique and is given by

$$\ell_0^T Y = h^T (A^T WA + (\sigma^2/\alpha^2) U^T U)^{-1} A^T WY.$$

Identifying λ with σ^2/α^2 this solution is seen to be formally the same as the estimate proposed by Currie for estimating the peak amplitude $\beta_0 = (1, 0, \ldots, 0)\beta$. An operational difference is that the minimax theorem assumes the smoothness parameter α^2 to be known, whereas Currie implicitly estimates it from the data. It should be noted that the estimate is minimax for estimating any single linear functional but is not generally minimax for estimating several linear functionals simultaneously [10].

In the next section we will consider the more general problem of several peaks of known shape and unknown amplitudes, superposed on an unknown baseline (Currie considered only the single peak case). We will develop expressions for the bias and variance of the amplitude estimates and limiting approximations as the expected total count $\mu \rightarrow \infty$ which give some insight into the properties of the method. In section 4 a procedure for choosing λ from the data is discussed and is illustrated by some simulations.

2. Bias and variance

In this section we will assume the following, multi-peak model:

$$\begin{aligned} Y &= \beta_{11}\Gamma_1 + \cdots + \beta_{1p}\Gamma_p + \beta_2 + \varepsilon \\ &= [\Gamma{:}I]\,\beta + \varepsilon \\ &= A\beta + \varepsilon \end{aligned}$$

where Y is an n-vector, $\Gamma = [\,\Gamma_1, \Gamma_2, \ldots, \Gamma_p]$, $\beta_2 = (\beta_{21}, \ldots, \beta_{2n})^T$ is the vector of mean background counts, $\beta^T = (\beta_1^T, \beta_2^T)$, and ε is a vector of random errors with nonsingular covariance matrix $\mu^{-1}W^{-1}$. We will derive expressions for the bias and variance of the estimate

$$\hat{\beta} = (A^TWA + \lambda U^TU)^{-1}A^TWY$$

when U is of the form

$$\underset{(n+p-k)\,x\,(n+p)}{U} = \begin{bmatrix} \underset{pxp}{0} & \underset{pxn}{0} \\ \underset{(n-k)xp}{0} & \underset{(n-k)xn}{U_1} \end{bmatrix},$$

and thus U^TU is of the form

$$\underset{(n+p)\,x\,(n+p)}{\mathrm{U^TU}} = \begin{bmatrix} \underset{pxp}{0} & \underset{nxn}{0} \\ \underset{pxp}{0} & \underset{nxn}{D} \end{bmatrix},$$

where

$$D = U_1^TU_1$$

(D is not diagonal) and $\lambda = 1/\mu\alpha^2$ is given. If λ is estimated from the data these expressions are conditional on λ. The unconditional bias and variance are different.

We will focus attention on the estimate β_1 of the vector of peak amplitudes, which is of primary interest. It is thus useful to partition the matrix $(A^TWA+\lambda U^TU)^{-1}$:

$$\begin{aligned} (A^TWA+\lambda U^TU)^{-1} &= \begin{bmatrix} \Gamma^TW\Gamma & \Gamma^TW^T \\ W\Gamma & W+\lambda D \end{bmatrix}^{-1} \\ &= \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \end{aligned}$$

From an identity for the inverse of a partitioned matrix [7],

$$\begin{aligned} B_{11} &= (\Gamma^TW\Gamma)^{-1}+(\Gamma^TW\Gamma)^{-1}\Gamma^TW^T[W+\lambda D-W\Gamma(\Gamma^TW\Gamma)^{-1}\Gamma^TW^T]^{-1}W\Gamma(\Gamma^TW\Gamma)^{-1} \\ &= G^{-1}+G^{-1}\Theta^TR^{-1}\theta G^{-1} \end{aligned}$$

where $G = (\Gamma^T W \Gamma)$, $\theta = W\Gamma$, and R is the matrix given in square brackets. With this notation,

$$B_{12} = -G^{-1}\theta^T R^{-1}.$$

$B_{12} = B_{21}^T$; we will not need B_{22}. Now,

$$E\hat{\beta} = (A^T W A + \lambda U^T U)^{-1} A^T W A \beta$$

and

$$A^T W A \beta = \begin{bmatrix} \Gamma^T W \Gamma \beta_1 + \Gamma^T W \beta_2 \\ W \Gamma \beta_1 + W \beta_2 \end{bmatrix}$$

$$= \begin{bmatrix} G\beta_1 + \theta^T \beta_2 \\ \theta \beta_1 + W \beta_2 \end{bmatrix},$$

so that

$$E\hat{\beta}_1 = (G^{-1} + G^{-1}\theta^T R^{-1} \theta G^{-1})(G\beta_1 + \theta^T \beta_2) - G^{-1}\theta^T R^{-1} (\theta\beta_1 + W\beta_2).$$

We thus have, after simplification, an expression for the bias of $\hat{\beta}_1$:

$$\beta_1 - E\hat{\beta}_1 = -G^{-1}\theta^T [I - R^{-1}(W - \Theta G^{-1} \Theta^T)]\beta_2. \tag{1}$$

Note that the bias does not involve β_1 and that the derivation of the bias expression has not assumed that $\mu^{-1} W^{-1}$ is the true covariance matrix of the random errors. In the appendix it is shown that the bias is zero if $U_1 \beta_2 = 0$.

A simple bound for the bias may be obtained as follows: from the expression above, the squared bias for a particular component β_{1k}, say, may be written in the form

$$|\beta_{1k} - E\hat{\beta}_{1k}|^2 = |r^T \beta_2|^2$$

Let $P = U_1^T (U_1 U_1^T)^{-1} U_1$ be the matrix which projects onto $N(U_1)$, let $Q = I - P$ project onto $N(U_1)$, and express $\beta_2 = P\beta_2 + Q\beta_2$. Noting from above that $r^T Q \beta_2 = 0$, we may write

$$\begin{aligned} |r^T \beta_2|^2 &= |r^T U_1^T (U_1 U_1^T)^{-1} U_1 \beta_2|^2 \\ &\leqslant \sup_{\{\beta_2:\ \|U_1\beta_2\|^2 \leqslant \alpha^2\}} |r^T U_1^T (U_1 U_1^T)^{-1} U_1 \beta_2|^2 \\ &= \alpha^2 |r^T U_1^T (U_1 U_1^T)^{-2} U_1 r|^2 \end{aligned}$$

We now consider the variance of the estimate. Under the assumption that the covariance matrix of the errors is $\mu^{-1} W^{-1}$, it is immediate that the covariance matrix of $\hat{\beta}$ is

$$\Sigma = \mu^{-1} (A^T W A + \lambda U^T U)^{-1} A^T W A (A^T W A + \lambda U^T U)^{-1}$$

In an appendix it is shown how this matrix may be partitioned and that the covariance matrix of β_1 can be expressed as

$$\Sigma_{11} = \mu^{-1} F^T F \tag{2}$$

where

$$F = W^{-1/2}[I-(W-\theta G^{-1}\theta^T)R^{-1}]\Theta G^{-1} \tag{3}$$

and $W^{1/2}$ is the symmetric square root of W.

We will now develop approximations to the bias and Σ_{11} for large samples by examining their behavior as T and thus $\mu \to \infty$ and $\lambda \to 0$. The expressions for Σ_{11} and the bias both involve the matrix

$$I - R^{-1}(W-\Theta G^{-1}\Theta T) = I - [W+\lambda D - W\Gamma(\Gamma^T W\Gamma)^{-1}\Gamma^T W]^{-1} \cdot (W - W\Gamma(\Gamma^T W\Gamma)^{-1}\Gamma^T W)$$

As $\lambda \to 0$, $R \to W - W\Gamma(\Gamma^T W\Gamma)^{-1}\Gamma^T W$, but this matrix is singular (the null space is spanned by $\Gamma_1, \ldots, \Gamma_p$). A further complication is that D will typically not be of full rank (for example, D may annihilate constant and linear functions). However, our assumption that $N(U) \cap N(A) = \phi$ guarantees that $D\Gamma_j \neq 0, j=1, \ldots, p$ and thus that the matrix R is invertible. In the appendix we prove the following:

LEMMA. *Suppose that* C *is an* nxn *non-negative definite matrix with* p *dimensional null space spanned by* $v_1, \ldots, v_p$. *Suppose that* D *is another* nxn *non-negative definite matrix and that* $N(C) \cap N(D) = \phi$. *Then as* $\lambda \to 0$

$$I - (C + \lambda D)^{-1}C = V(V^T D V)^{-1}V^T D + 0(\lambda)$$

where $V = [v_1, \ldots, v_p]$ *is an* nxp *matrix.*

Applying this lemma to the expressions for Σ_{11} and the bias of β_1, with $W - W\Gamma(\Gamma^T W\Gamma)^{-1}\Gamma^T W$ corresponding to C and Γ corresponding to V we have,

COROLLARY: *Under the assumptions of our linear model, as* $\lambda \to 0$ $(\mu \to \infty)$,

$$\beta_1 - E\hat{\beta}_1 = -(\Gamma^T D\Gamma)^{-1}\Gamma^T D\beta_2 + 0(\lambda) \tag{1}$$

$$\mu\Sigma_{22} = (\Gamma^T D\Gamma)^{-1}(\Gamma^T D W^{-1} D\Gamma)(\Gamma^T D\Gamma)^{-1} + 0(\lambda). \tag{2}$$

The expression for the bias is simpler to understand if we write it as

$$\beta_1 - E\hat{\beta}_1 \to -[(U_1\Gamma)^T(U_1\Gamma)]^{-1}(U_1\Gamma)^T(U_1\beta_2)$$

and keep in mind that U_1, is a differencing operator. The bias is determined by the relationships of the vectors $U_1\Gamma_j$, $j=1, \ldots, p$ and $U_1\beta_2$. If the baseline β_2 is quite smooth $U_1\beta_2$ will be small. If a particular peak shape Γ_j does not overlap any other peaks then the limiting $(\mu \to \infty)$ bias of the estimate of its amplitude is simply

$$\beta_{1j} - E\hat{\beta}_{1j} \cong \frac{(U_1\Gamma_j)^T(U_1\beta_2)}{\|U_1\Gamma_j\|^2} < \frac{\alpha}{\|U_1\Gamma_j\|}$$

which follows from the rule for the inverse of a partitioned matrix and the Cauchy-Schwartz inequality. The large components of $U_1\Gamma_j$ will be those near the peak center and if the true background β_2 is smooth in this region, the bias will be small.

When two peaks overlap substantially, however, the bias will typically be worse than the bias if either one of the peaks were absent, since corresponding elements of the matrix $[(U_1\Gamma)^T(U_1\Gamma)]^{-1}$ will be large.

Finally, we note that this limiting bias does not depend on the weighting matrix W and that it depends linearly on the baseline proportion.

The variance of the estimate β_{1j} of a peak amplitude can also be expressed simply in the case that the matrix W is diagonal and the peak does not overlap other peaks:

$$Var(\hat{\beta}_{1j}) \cong \frac{(U_1\Gamma_j)^T U_1 W^{-1} U_1^T (U_1\Gamma_j)}{\mu\|U_1\Gamma_j\|^2}$$

but in the case that there is considerable peak overlap the variance may be inflated considerably.

It is of some interest to consider the relative size of the bias to the standard error and to understand qualitatively how this is affected by varying the baseline amplitude. To this end we consider a single peak model with a peak shape standardized so that $\Sigma\gamma_j = 1$ and a standard baseline profile with $\sum_{j=1}^{n} \beta j = 1$. Any mixture of this peakshape and background profile with peak proportion β_0 and background proportion $1-\beta_0$ can be expressed as $\beta_0\Gamma + (1-\beta_0)\beta$, where $0 \leqslant \beta_0 \leqslant 1$. Denoting $D\Gamma$ by $V = (V_1, \ldots, V_n)^T$ and taking $W^{-1} = \operatorname{diag}(\beta_0\gamma_j + (1-\beta_0)\beta_j)$, the appropriate bias ($B$) and standard error ($\sigma$) of β_0 given by the equations above are

$$|\mathrm{B}| \simeq (1-\beta_0)\, \Sigma\, V_i\beta_i / \Sigma V_i\gamma_i$$

$$\sigma \simeq \frac{1}{\sqrt{\mu}} \left[(1-\beta_0)\, \Sigma V_i^2\beta_i + \beta_0\, \Sigma V_i^2\gamma_i\right]^{1/2} / \Sigma F_i\gamma_i$$

From these expressions we may make some observations that agree with observations made by Currie on the basis of empirical experiments: (1) The bias is proportional to the background proportion; (2) For small values of β_0 the standard error is proportional to the square root of the background proportion; (3) Since $\Sigma V_i^2\beta_i$ is typically less than $\Sigma V_i^2\gamma_i$, the standard error increases with increasing peak area proportion.

We conclude this section with a brief consideration of the problem of mis-specification of Γ. Suppose that the true peak profile is $\Gamma_0 = \Gamma + \delta\Gamma$; from calculations similar to those done above for the bias, we find that the additional bias introduced by $\delta\Gamma$ is

$$G^{-1}\Theta^T \left[I - R^{-1}(W - \Theta G^{-1}\Theta^T)\right]\delta\Gamma\beta_1$$

which, as $\mu \to \infty$, tends to

$$(\Gamma^T D\Gamma)^{-1}\, (\Gamma^T D\delta\Gamma)\beta_1 .$$

In the single peak case, the Cauchy-Schwarz inequality shows that this quantity is bounded in absolute value by $\beta_1 \|U_1\delta\Gamma\| / \|U_1\Gamma\|$. Thus a variation $\delta\Gamma$ such that $U_1\delta\Gamma$ is highly correlated with $U_1\Gamma$ will give rise to a relatively large bias proportional to the peak amplitude.

3. Choosing λ

If the parameter α^2 is known, the minimax λ is $\lambda = 1/\mu\alpha^2$. In the absence of this knowledge, λ must be chosen from the data. In this section we discuss a class of such procedures and illustrate them with examples.

Given a non-negative definite matrix B, one might attempt to choose λ to minimize

$$E(\hat{Y}(\lambda) - EY)^T B\, (\hat{Y}(\lambda) - EY) = ET_B(\lambda)$$

where

$$\begin{aligned} \hat{Y}(\lambda) &= A(A^T WA + \lambda U^T U)^{-1} A^T WY \\ &= A(\lambda)Y. \end{aligned}$$

$ET_B(\lambda)$ is a weighted mean-square error. This quantity may be estimated from the data by using

$$\begin{aligned} RSS_B(\lambda) &= (Y - \hat{Y}(\lambda))^T B\, (Y - \hat{Y}(\lambda)). \\ &= Y^T (I - A(\lambda))^T B\, (I - A(\lambda))Y \\ &= Y^T GY . \end{aligned}$$

The expectation of $RSS_B(\lambda)$ can be computed to be

$$ERSS_B(\lambda) = ET_B(\lambda) + \mu^{-1} tr\,(BW^{-1}) - 2\mu^{-1}\, tr\,(BA(\lambda)W^{-1})$$

and thus an unbiased estimate of $ET_B(\lambda)$ is

$$\hat{T}_B(\lambda) = RSS_B(\lambda) - \mu^{-1} tr\,(BW^{-1}) + 2\mu^{-1} \mathrm{tr}\,(\mathrm{BA}(\lambda)\mathrm{W}^{-1})\,.$$

We note that if Y follows a Gaussian distribution, then

$$Var\hat{T}_B(\lambda) = 2\,\mu^{-2} tr\,(CW^{-1})^2 + 4\,\mu^{-1}\,(A\beta)^T GW^{-1} A\beta$$

For a given B we propose choosing λ to minimize $\hat{T}_B(\lambda)$. (Similar procedures with $B = \mathrm{I}$ have been discussed in [3, 6].)

If it were possible, we might choose B so that $ET_B(\lambda) = E\|\beta_1 - \beta_1(\lambda)\|^2$, the total mean square error of the estimates of the peak amplitudes. However, if we write

$$EY = [\Gamma : I]\begin{bmatrix}\beta_l \\ \beta_2\end{bmatrix}$$

$ET_B(\lambda)$ may be expressed as

$$\begin{aligned} ET_B(\lambda) = {} & E(\beta_1 - \hat{\beta}_1(\lambda))^T\Gamma^T B\Gamma\,(\beta_1 - \hat{\beta}_1(\lambda)) \\ & + E(\beta_2 - \hat{\beta}_2(\lambda))^T B\,(\beta_2 - \hat{\beta}_2(\lambda)) \\ & + 2E(\beta_1 - \hat{\beta}_1(\lambda))^T\Gamma^T B\,(\beta_2 - \hat{\beta}_2(\lambda)) \end{aligned}$$

from which it is apparent that it is imposible to choose B so that the second two terms vanish and the first does not.

We have experimented with three choices of B: $B_1 = I$, $B_2 = \Gamma(\Gamma^T\Gamma)^{-1}\Gamma^T$ and $B_3 = \Gamma(\Gamma^T\Gamma)^{-2}\Gamma^T$. B_2 is the matrix which projects onto the column space of Γ; the motivation for choosing B_2 is that $\beta_2 - \hat{\beta}_2(\lambda)$ will hopefully not be highly correlated with the columns of Γ and thus the second two terms will be small and the first term will dominate. Choosing B_3 reduces the first term to $E\|\beta_1 - \hat{\beta}_1(\lambda)\|^2$ and hopefully causes the other terms to be small. A disadvantage in using B_2 or B_3 is that if there are two or more peaks with considerable overlap, the variance of $\hat{T}_B(\lambda)$ may be rather large, causing the procedure to be rather unstable.

Currie suggests choosing λ so that $RSS_W(\lambda) = n/\mu$. The motivation for this is that $\mu \cdot RSS_W$ would follow a chi-square distribution with n degrees of freedom if $E\hat{Y}(\lambda) = EY$ and no parameters were estimated from the data. In fact, however, parameters have been estimated from the data, although it is not clear how many "degrees of freedom" remain, and $E\hat{Y}(\lambda) \neq EY$. Thus the application of the x^2 distribution is questionable. The procedure outlined above with $B = W$ would choose λ to minimize

$$\hat{T}_W(\lambda) = RSS_W(\lambda) - n\mu^{-1} + 2\mu^{-1}\, tr\, A(\lambda)$$

which would cause $RSS_W(\lambda)$ to be somewhat smaller than n/μ. (In a vague sense, the "degrees of freedom" of the Chi-square statistic are reduced.)

We now briefly discuss the results of some simulations of this technique. The configurations are the following: (1) two slightly overlapping peaks on a linear baseline, (2) the same peaks on a quadratic baseline, (3) two highly overlapped peaks on a quadratic baseline, and (4) a single peak on a quadratic baseline which also contains a small "unsuspected" peak obscured by the dominant peak. All the simulations were done over a width of 20 channels with a total count $\mu = 10^5$. The sum of squared second differences was used as the smoothness measure. Computations were done on the Univac 1100 at the National Bureau of Standards. Subroutines from the IMSL library were used to generate random numbers and for matrix calculations. The most numerically sensitive calculation is the inversion of the matrix $A^T WA + \lambda U^T U$, which in theory is positive definite; however, the matrix may be for practical purposes numerically singular for very small or very large values of λ, so it is important that a good algorithm be used and that diagnostic messages be printed when instabilities arise. (An alternative to actually forming and inverting this matrix is to simultaneously diagonalize A^TWA and U^TU; having done this once, $(A^TWA + \lambda U^TU)^{-1}$ may be computed quite rapidly for various values of λ.)

1. Two peaks on a linear baseline; the peak shapes were Gaussian with locations at channels 8 and 12 and standard deviations 1.5. Each peak contained 30 percent of the total area. The baseline was $\beta_j = c(1 + j)$ where c was chosen so that the baseline area was 40 percent. For this configuration the optimal (minimum variance unbiased) method of peak area estimation is weighted linear least squares; we are interested in seeing what "price" has to be paid for the additional flexibility of the smoothing method in this null case. Table 1a shows the bias, variance, and total mean square error of the peak area estimates for various values of λ. From the table we see that *ETB* decreases as λ increases (for λ greater than 10^7 numerial problems develop). For $\lambda = 10^5$ the variance is very close to that for the linear least squares.

TABLE 1a.

λ	Bias β_{11}	Var β_{11}	Bias β_{12}	Var β_{12}	Total *MSE*	$ETB_1 \times 10^5$	$ETB_2 \times 10^5$	$ETB_3 \times 10^5$
10^0	0	0.559(−4)	0	0.593(−4)	0.115(−3)	0.664	0.180	0.985
10^1	0	.315(−4)	0	.345(−4)	.660(−4)	.404	.178	.975
10^2	0	.108(−4)	0	.126(−4)	.234(−4)	.278	.176	.963
10^3	0	.692(−5)	0	.772(−5)	.146(−4)	.237	.175	.956
10^4	0	.519(−5)	0	.596(−5)	.111(−4)	.217	.174	.952
10^5	0	.487(−5)	0	.575(−5)	.106(−4)	.214	.174	.951
least squares ($\lambda = \infty$)	0	.486(−5)	0	.575(−4)	.106(−4)			

Table 1b shows the results for one realization with random Poisson noise added. As stated above, the total count was 10^5. $\hat{T}B_1$ is minimized at $\lambda = 10^3$ and $\hat{T}B_2$ and $\hat{T}B_3$ are minimized at $\lambda = 10^5$. (In this and in the later simulations in which noise was added, the weighting matrix W was estimated from the data.)

TABLE 1b.

λ	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$TB_1 \times 10^5$	$TB_2 \times 10^5$	$TB_3 \times 10^5$
10^0	0.295	0.298	0.692	0.180	0.984
10^1	.295	.299	.515	.178	.972
10^2	.297	.302	.389	.173	.948
10^3	.299	.302	.340	.171	.942
10^4	.299	.300	.370	.171	.933
10^5	.299	.300	.381	.170	.932

2. Two peaks on a quadratic baseline—the peaks were as above and the background was $\beta_j = c(1 + j + j^2/20)$ above c was chosen so that $\Sigma\beta_j = 0.4$. This shape deviates only slightly from a linear baseline. Table 2a exhibits the biases, variance, and total mean square error for various values of λ; as λ increases the variance decreases and the bias increases. For this discretization the minimum total mean square error occurs for $\lambda = 350$ ($MSE = .17 \times 10^{-4}$). The mean square error for the least squares method is much larger, being dominated by the bias ($MSE = 0.42 \times 10^{-3}$). The minima of ETB_1, ETB_2, and ETB_3 occur at $\lambda = 250$, 450, and 550 respectively, over which range the *MSE* does not change appreciably.

Table 2b summarizes the results of a single realization with random Poisson noise. $\hat{T}B_1$, $\hat{T}B_2$, and $\hat{T}B_3$ are minimized at $\lambda = 350$, 250 (or 350), and 350, respectively. It is noteworthy that the estimates do not change substantially over the tabulated range of λ. Other realizations gave similar results.

For this example there is little difference in the results for B_1, B_2, or B_3—any choice would give satisfactory results. B_1 is somewhat easier to compute.

TABLE 2a.

λ	Bias β_{11}	Bias β_{12}	Var β_{11}	Var β_{12}	Total $MSE \times 10^4$	$ETB_1 \times 10^5$	$ETB_2 \times 10^5$	$ETB_3 \times 10^5$
50	−0.327(−4)	−0.291(−4)	0.144(−4)	0.169(−4)	0.312	0.304	0.174	0.949
150	−.649(−4)	−.185(−4)	.899(−5)	.111(−4)	.201	.272	.173	.944
250	.245(−3)	.468(−3)	.787(−5)	.976(−5)	.179	.266	.1723	.942
350	.448(−3)	.750(−3)	.740(−5)	.911(−5)	.1727	.267	.17219	.9413
450	.658(−3)	.102(−2)	.714(−5)	.868(−5)	.1728	.270	.17216	.94099
550	.867(−3)	.127(−2)	.696(−5)	.836(−5)	.177	.275	.17218	.94097
650	.107(−2)	.151(−2)	.684(−5)	.811(−5)	.184	.281	.17225	.94117
750	.128(−2)	.173(−2)	.673(−5)	.791(−5)	.193	.288	.1723	.942
850	.148(−2)	.195(−2)	.665(−5)	.774(−5)	.204	.296	.173	.942
950	.167(−2)	.215(−2)	.657(−5)	.759(−5)	.216	.304	.173	.943
least squares ($\lambda = \infty$)	.156(−1)	.128(−1)	.465(−5)	.579(−5)	4.19			

TABLE 2b.

λ	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$TB_1 \times 10^5$	$TB_2 \times 10^5$	$TB_3 \times 10^5$
50	0.300	0.305	0.256	0.173	0.945
150	.299	.303	.206	.1721	.9415
250	.299	.302	.196	.17190	.9405
350	.298	.302	.195	.17190	.9404
450	.298	.301	.199	.17196	.9406
550	.298	.301	.205	.1721	.9409
650	.298	.301	.213	.1722	.9412
750	.297	.300	.222	.1723	.9416
850	.297	.300	.232	.1724	.9421
950	.297	.300	.242	.173	.943

3. Two peaks on a quadratic baseline; the peaks were close enough together (centers 9, 11, $\sigma = 1.5$) so that there was no trough between them when they were superimposed. The peak areas were 0.3 and 0.3 again and the baseline was as in the previous example. On a grid of λ values spaced linearly by 150 the minimum *MSE* occurred at $\lambda = 800$ ($MSE = 0.20610 \times 10^{-4}$); the minimum of ETB_1 was at $\lambda = 350$ ($MSE = 0.213 \times 10^{-4}$); the minimum of ETB_2 was at $\lambda = 650$ ($MSE = 0.20611 \times 10^{-4}$); the minimum of ETB_3 was at $\lambda = 950$ ($MSE = 0.207 \times 10^{-4}$). The *MSE* for a linear least squares fit was 0.241×10^{-3}. Table 3 records the minimizing values of λ for $\hat{TB}_1$, $\hat{TB}_2$, and $\hat{TB}_3$, and the corresponding *MSE*'s for 4 realizations. The results suggest that $\hat{TB}_1$ may be a more stable criterion function in this situation, but we would not wish to make a conclusion on the basis of a sample size of 4!

TABLE 3. *Minimizing values of* λ *and corresponding MSE's for four realizations.*

	TB_1	TB_2	TB_3
1	*50(.279 × 10⁻⁴)*	*1400(.217 × 10⁻⁴)*	*1300(.217 × 10⁻⁴)*
2	$500(208 \times 10^{-4})$	$3000(288 \times 10^{-4})$	$2150(246 \times 10^{-4})$
3	$500(208 \times 10^{-4})$	$1100(210 \times 10^{-4})$	$2600(267 \times 10^{-4})$
4	$950(207 \times 10^{-4})$	$5000(407 \times 10^{-4})$	$6500(501 \times 10^{-4})$

4. A single peak (center = 10, $\sigma = 2$) on a quadratic baseline with a hidden peak centered at 12 with standard deviation 2. The peak area of the dominant peak was 0.8 and the area of the hidden peak was 0.02. In an attempt to mimic a situation in which the hidden peak is unsuspected, a single peak model was fit. The behaviors of ETB_1, ETB_2, and ETB_3 were somewhat different. ETB_1 had a minima at $\lambda = 10$ ($MSE = 0.55 \times 10^{-4}$) whereas ETB_2 and ETB_3 had minimum at $\lambda = 10^4$ ($MSE = 0.96 \times 10^{-4}$). The *MSE* was minimum at $\lambda = 10^7$ ($MSE = 0.18 \times 10^{-4}$). The *MSE* of the linear least squares procedure was 0.21×10^{-4}. The reason that ETB_1 was minimized for a smaller value of λ is that this criterion gives greater weight to fitting the baseline as well as the peak than do the other two, which concentrate more on the peak. The baseline (which includes the hidden peak) is fit well with small values of λ since it is not very smooth. Since the hidden peak has substantial correlation with the modelled peak, however, B_2 and B_3 fail to choose λ large enough.

On several realizations with random noise $\hat{T}B_1$ achieved a minimum at small values of λ and $\hat{T}B_2$ and $\hat{T}B_3$ at larger values of λ. On some occasions $\hat{T}B_2$ and $\hat{T}B_3$ also had local minima at small values of λ. Figure 1 shows the estimated baseline for $\lambda = 20$, which was the attained minimum for $\hat{T}B_1$ on a particular realization. The unsuspected peak shows quite clearly, giving valuable diagnostic information! The estimated baseline for the larger value of $\lambda = 10^4$ at which $\hat{T}B_2$ and $\hat{T}B_3$ were minimized smooths over the peak (fig. 2). We also plotted residuals on a square root scale to stabilize the variance, $y_i = \sqrt{y_i} - \sqrt{\hat{y}_i(\lambda)}$. Figure 3 shows the residual plot for $\lambda = 10^4$; there is a hint of a discrepancy near channel 12.

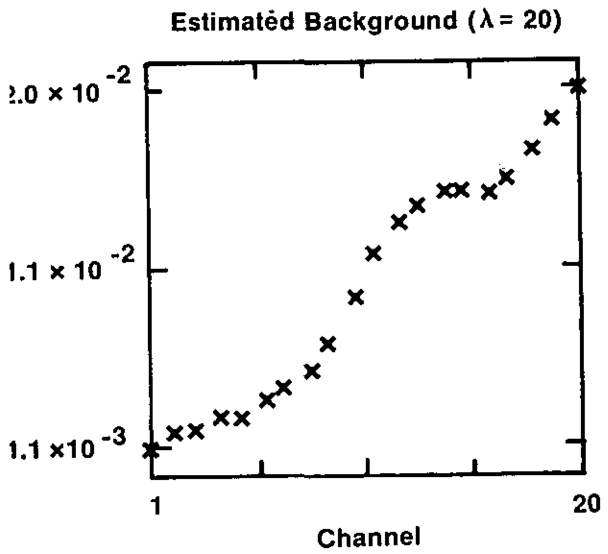

FIGURE 1.

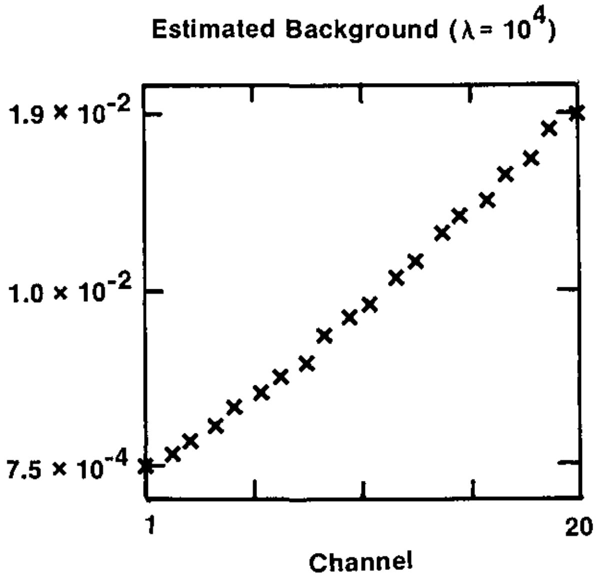

FIGURE 2.

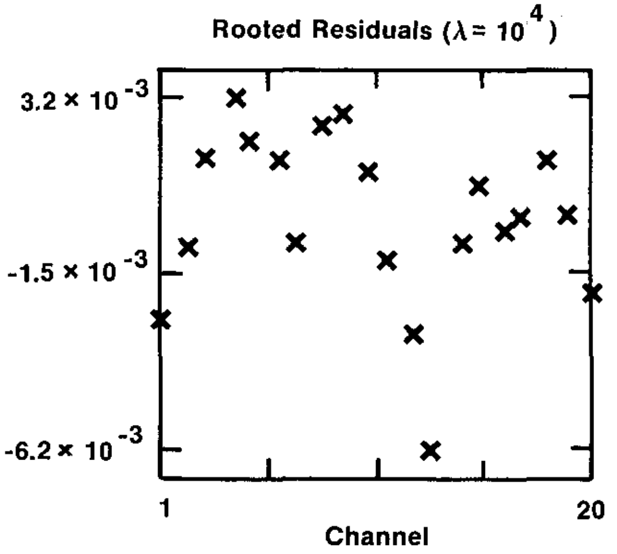

FIGURE 3.

If the hidden peak is incorporated into the model, the total *MSE*, ETB_1, ETB_2 and ETB_3 are all minimized for $\lambda \approx 10^3$. The total *MSE* is 0.36×10^{-4} and the individual *MSE*'s are 0.20×10^{-4} and 0.16×10^{-4} for the large and small peaks respectively. The bias and variance for the small peak are 0.99×10^{-3} and 0.15×10^{-4} so that the relative error in estimating this peak area is quite large. For the linear least squares method the total *MSE* is 0.13×10^{-3}; the bias and variance for the small peak are $.36 \times 10^{-2}$ and $.12 \times 10^{-4}$.

On the basis of these computations there is no clear evidence that would favor B_2 or B_3 over B_1, despite the fact that they were designed to focus more on the peak. The last example shows that focusing on the peak may hide unsuspected features of the baseline. The computations suggest that choosing λ to minimize $\widehat{TB}(\lambda)$ is reasonable, but they are not nearly extensive enough to give insight into the stochastic behavior of the minimizing λ.

There are many possibilities we have not investigated. Other choices of B are possible; for example $B = \Gamma_j((\Gamma_j^T\Gamma_j)^{-1}\Gamma_j^T$ would focus on the jth peak if there were more than one peak, $B = W^{-1}$ would weight the deviations according to the variances of the observed counts; a possible advantage of this choice is that the statistics $RSS_W(\lambda)$ might be compared with the percentiles of a χ^2 distribution (above, however, we have noted some difficulties with this procedure). Another possibility is to attempt to choose between several smoothness criteria by computing $\widehat{TB}^{(k)}(\lambda)$ for $k = 1, 2, 3, \ldots, K$ and choosing the solution corresponding to

$$\min_{k} \inf_{\lambda} \widehat{TB}^{(k)}(\lambda) .$$

4. Final Comments

The results above leave several questions unanswered and suggest problems for further research. The following is perhaps the most immediate: in many applications the peak vector is not known exactly, but is assumed to have a parametric form such as $\gamma_j = \gamma_j(\mu,\sigma) = \frac{1}{\sigma}\gamma\frac{j-\mu}{\sigma}$, where γ is a given function μ and σ are location and shape parameters and must be estimated from the data. If the peak profile Γ is estimated from other experiments, for example from pure sources, the variability of the estimate will affect subsequent analyses in which it is used. We plan to pursue the analysis of these problems in the future.

An alternative approach to the problem is to use the method of maximum likelihood with the assumption of Poisson statistics; which might be more appropriate for small counts. The likelihood function of β could be maximized subject to the constraint $\|U\beta\|^2 = \alpha^2$. Although we conjecture that the large sample properties of the estimates would be equivalent to the results above, the small sample properties would be different.

Finally, we note again that in the multi-peak situation the estimates we have considered are minimax for any single peak amplitude but are probably not jointly minimax. One might attempt to solve the simultaneous minimax problem by numerical optimization; we conjecture that the results would not be substantially different.

5. References

[1] Cullum, J. Ill-posed problems, regularization, and the singular value decomposition. IBM Tech. Rep. RC 6465; 1977.

[2] Currie, L. Model uncertainty and bias in the evaluation of nuclear spectra. J. Radioan. Chem. **39**(1–2): 223–237; 1977.

[3] Golub, G.; Heath, M.; Wahba, G. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics **21**(2): 215–223; 1979 May.

[4] Greub, W. *Linear Algebra*, 2nd Ed. New York: Springer-Verlag; 1963. 338 p.

[5] Kuks, J.; *Olman, V. A minimax estimator of regression coefficients.* (Russian). Izv. Akad. Nauk. Eston. SSR **21**: 66–72; 1972.

[6] Mallows, C. L. Some comments on Cp. Technometrics **15**(4): 661–675; 1973 November.

[7] Rao, C. R. *Linear statistical inference and its applications*. New York: Wiley; 1965. 624 p.

[8] Reinsch, C. Smoothing by spline functions. Numer. Math. **10**(3): 177–183; 1967 October, 4.

[9] Speckman, P. Minimax estimates of linear functionals in a Hilbert space. Ann. Stat., to appear. 1981.

[10] Speckman, P. On minimax estimation of linear operators in Hilbert spaces from noisy data. Manuscript. 1980.

[11] Whittaker, E. and Robinson, G. The calculus of observations [sec. 151–155]. London: Blackie and Sons; 1944.

6. Appendix

Here we derive an expression for the covariance matrix of $\hat{\beta}_1$ and prove the lemma in section 2 of the text. The covariance matrix of $\hat{\beta}$ is, with the notation of section 2,

$$\begin{aligned}\mu\Sigma &= (A^TWA+\lambda U^TU)^{-1}A^TWA(A^TWA+\lambda U^TU)^{-1}\\ &= \begin{bmatrix} B_{11} & B_{12}\\ B_{21} & B_{22}\end{bmatrix}\begin{bmatrix} G & \Theta^T\\ \Theta & W\end{bmatrix}\begin{bmatrix} B_{11} & B_{12}\\ B_{21} & B_{22}\end{bmatrix}\end{aligned}$$

We are interested in Σ_{11}. Multiplying through and noting that $B_{21} = B_{12}^T$

$$\begin{aligned}\mu\Sigma_{11} &= B_{11}GB_{11} + B_{12}\Theta B_{11} + B_{11}\Theta^TB_{12}^T + B_{12}WB_{12}^T\\ &= B_{11}\Gamma^TW\Gamma B_{11} + B_{12}W\Gamma B_{11} + B_{11}\Gamma^TWB_{12} + B_{12}^TWB_{12}\\ &= (W^{1/2}\Gamma B_{11} + W^{1/2}B_{12}^T)^T\,(W^{1/2}\Gamma B_{11} + W^{1/2}B_{12}^T)\\ &= F^TF\ .\end{aligned}$$

Now, using the expressions for B_{11} and B_{12}, and $\Gamma = W^{-1}\Theta$

$$\begin{aligned}F &= W^{1/2}\,(\Gamma G^{-1} + \Gamma G^{-1}\Theta^TR^{-1}\Theta G^{-1} - R^{-1}\Theta G^{-1})\\ &= W^{1/2}\,(W^{-1} + W^{-1}\Theta G^{-1}\Theta^TR^{-1}-R^{-1})\,\Theta G^{-1}\\ &= W^{-1/2}\,[I - (W-\Theta G^{-1}\Theta^T)R^{-1}]\,\Theta G^{-1}\ ,\end{aligned}$$

which is the expression to be derived.

We now prove the lemma. The key to the proof is the fact that under the assumptions of the lemma C and D may be simultaneously diagonalized [4]; there exists a nonsingular matrix X such that

$$\begin{aligned}X^TCX &= \Omega\\ X^TDX &= M\end{aligned}$$

where Ω and M are diagonal matrices with elements ω_i and μ_i. From this representation we note that the null space of C (resp. D) is spanned by those columns of X corresponding to zero diagonal elements of Ω (resp. M). The assumption of the lemma guarantees that the two null spaces contain no vectors in common. Now expressing C and D in terms of X, Ω, and M, and writing $I = XX^{-1}$

$$I - (C+\lambda D)^{-1}C = X\,[I-(\Omega+\lambda M)^{-1}\Omega]\,X^{-1}$$
$$= XR_\lambda X^{-1}$$

where $R_\lambda = \text{diag}\,[\lambda\mu_i/(\omega_i+\lambda\mu_i)]$.

We note that if $\beta_2 \varepsilon N(U_1) = N(D)$ this representation makes it clear that β_1 is unbiased, for if x_j is a column of X corresponding to $\mu_j = 0$, then

$$XR_\lambda X^{-1}x_j = X\frac{\lambda\mu_j}{\omega_j+\lambda\mu_j}e_j = 0$$

where e_j is the j^{th} unit vector.

The diagonal elements of R_λ corresponding to $\omega_j = 0$ are 1's, so that

$$XR_\lambda X^{-1} = X\begin{pmatrix}0 & 0\\ 0 & I\end{pmatrix}X^{-1} + \lambda X\begin{pmatrix}N_\lambda & 0\\ 0 & 0\end{pmatrix}X^{-1}$$

where $N_\lambda = \text{diag}\,[\mu_i/(\omega_i+\lambda\mu_i)]$. It is easily verified that the first matrix, call it P, on the right hand side of the expression above has the following properties: (1) it is idempotent with range $N(C)$; (2) $Pv = 0$ if $v \varepsilon N(D)$; (3) for any vector v, $(Pv)^T D(I-P)v = 0$. P is therefore a projection matrix which projects orthojonally with respect to the pseudo inner-product $\langle u,v\rangle = u^T Dv$, and may be written

$$P = V(V^T DV)^{-1}V^T D$$

where $V = (v_1, \ldots, v_p)$ spans the null space of C. Finally noting that N_λ is bounded, we have

$$XR_\lambda X^{-1} = P + 0(\lambda)$$

Finally, we note that expansions for small values of λ (corresponding to large samples) or small values of λ^{-1} (corresponding to a nearly linear background and moderate sample size) may be carried using identities of the form

$$\frac{1}{1+\varepsilon} = 1 - \varepsilon + \frac{\varepsilon^2}{(1+\varepsilon)}.$$

JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 87, No. 1, January–February 1982

A Note on the Behavior of Least Squares Regression Estimates When Both Variables Are Subject to Error

Clifford Spiegelman*

National Bureau of Standards, Washington, DC 20234

April 19, 1981

For the errors in variables model $X = U + V$, $Y = \beta f(U) + W$, sufficient conditions are given for the L.S. limiting estimate of β to satisfy $P(\hat{\beta}/\beta < 1) = 1$ or $P(\hat{\beta}/\beta > 1) = 1$ as the sample size tends to infinity.

Key words: Errors in variable; structural; functional; regression; large sample, convex.

The problem of linear regression when both variables are subject to error is known to be difficult, see Madansky [1],[1] and Moran [2]. In particular under general conditions there is no consistent estimator for β in the model (1.1), (1.2) below based upon only the first two moments of X and Y. Let

$$X = U + V \tag{1.1}$$

$$Y = \beta f(U) + W, \quad \text{where} \tag{1.2}$$

U, V, and W are unobservable independent random variables with $EV = EW = 0$. In addition β is an unknown constant and f is a given function. We suppose that EX^2 and EY^2 are finite. This is known as the structural form of the errors in variables problem. Since there is a great deal of confusion in the literature between the case when U is a random variable and when U is not (the functional case), only the structural case is dealt with directly. Parallel results for the functional case can be obtained in a straightforward manner. These results will, however, restrict the values that a sequence of constants $U_1, \ldots, U_n$ can take.

The least squares estimate of β is, of course,

$$\hat{\beta} = \sum Y_i f(X_i) \Big/ \sum (f(X_i))^2 \tag{2}$$

where the observable random pairs (X_i, Y_i) $i = 1, \ldots, n$, are independent and have the same joint distribution as X and Y, see (1.1) and (1.2). It is well known that when $\beta \neq 0$, and $f(X) = X$ that

$$P\left(\frac{\hat{\beta}}{\beta} < 1\right) \rightarrow 1 \quad \text{as} \quad n \rightarrow \infty\,. \tag{3}$$

(The least squares estimate is biased toward zero.)

It is also known (see Kendall and Stuart [3]) that for $f(X) = X^k$, $(k = 1, 2, \ldots)$, and X and Y are normal that result (3) holds. However until now general conditions under which (3) holds were not available. We give sufficient conditions under which either (3) (Theorem 1) or the opposite result (Theorem 2)

$$P\left(\frac{\hat{\beta}}{\beta} > 1\right) \rightarrow \quad \text{as} \quad n \rightarrow \infty \quad \text{holds.} \tag{4}$$

While the author has not found these two results in the literature he believes that they may be well known by somebody.

---

* Center for Applied Mathematics, National Measurement Laboratory.

[1] Figures in brackets indicate literature references at the end of this paper.

THEOREM 1. *If* f(X)^2 *is convex and not constant a.s. then* (3) *holds for* $\beta \neq 0$.

PROOF:

$$\frac{\hat{\beta}}{\beta} = \frac{\Sigma f(U_i)f(X_i)}{\Sigma f(X_i)^2} + \frac{1}{\beta}\frac{\Sigma W_i f(X_i)}{\Sigma f(X_i)^2} = \frac{\Sigma f(U_i)f(X_i)}{\Sigma f(X_i)^2} + O_p(n^{-1/2}) \tag{5}$$

Since $f^2(X)$ is convex it follows by Jensen's inequality that

$$Ef(X)^2 \geqslant E(f(E[X|U]))^2 = Ef(U)^2 \tag{6}$$

From another application of Jensen's inequality we have

$$Ef(X)^2 > E(E[f(X)|U])^2 \tag{7}$$

In addition notice that

$$\begin{aligned} Ef(U)f(X) &= E(f(U)E[f(X)|U)] \\ &\leqslant E(|f(U)||E[f(X)|U]|) \leqslant (E(f(U)^2)E(E[f(X)|U])^2)^{1/2} \end{aligned}$$

by the Cauchy-Swartz inequality. Therefore by (6) and (7)

$$(Ef(X)^2)^2 \geqslant (Ef(U)f(X))^2$$

which implies that

$$Ef(X)^2 > Ef(U)f(X)$$

The theorem now follows by applying the strong law of large numbers to the terms $\frac{\Sigma f(U_i)f(X_i)}{n}$ and $\frac{\Sigma f(X_i)^2}{n}$ in (4).

If f^2 is not convex a positive β may be overestimated. Theorem 2 provides the necessary support for this statement.

THEOREM 2. *If* f *has two continuous derivatives and satisfies*

$$\frac{-\,f(z)f''(z)}{2} > (f'(z))^2 \tag{8}$$

for z *in some interval* I *then there exists distributions for* U, V, *and* W *such that (4) holds.*

PROOF: Following statements in the proof of Theorem 1 it is sufficient to show that there exists distributions for U, V, and W such that

$$Ef(U)f(X) > Ef(X)^2.$$

Since f has two continuous derivatives it follows that

$$\begin{aligned} f(X) &= f(U) + f'(U)V + f''(\Theta_1)\frac{V^2}{2} \\ f(X)^2 &= f(U)^2 + 2f(U)f'(U)V + [f''(\Theta_2)f(\Theta_2) + (f'(\Theta_2))^2]V^2 \end{aligned}$$

where Θ_1 and Θ_2 are points between U and V.

Take u_0 to be in I. Let η be any point such that

$$\inf_{u-\eta \leqslant z_1, z_2 \leqslant u+\eta} \frac{f(U_0)f''(z_1)}{2} - f(z_2)f''(z_2) > (f'(z_2))^2. \tag{9}$$

Such a point η exists because f'' is continuous, and (8) holds by hypothesis. Take V to have a two point distribution

$$P(V=\eta) = P(V=-\eta) = 1/2 .$$

Then

$$\begin{aligned} &E(f(u_0)f(X) - f^2(X)) \\ &= f(u_0)\left(f(u_0) + Ef''(\Theta_1)\frac{\eta^2}{4}\right) - \left[f(u_0)^2 + \frac{\eta^2}{2}E(f''(\Theta_2)f(\Theta_2) + f'(\Theta_2)^2)\right] > 0 . \end{aligned}$$

Finally we note that, since this last inequality is strict, U may have uniform distribution in a narrow interval around the chosen point U_0.

EXAMPLE 1: Let $f(X) = X^\alpha$, $X > 0$. Then, if $\alpha \geqslant 1/2$ the conditions of Theorem 1 are satisfied. On the other hand if $\alpha < 1/3$, the conditions of Theorem 2 are satisfied. This example is important for NBS standards work for concrete strength, see [4]. (However, the functional case is appropriate.) It is also important for background characterization in x-ray spectroscopy.

EXAMPLE 2: The conditions of theorem 2 are not necessary. If $f(X)$ satisfies condition (8) for Z a rational number in the unit interval and arbitrary elsewhere, then the proof of Theorem 2 can be used to construct distributions such that (4) holds.

COMMENT 1: The conditions of Theorems 1 and 2 can be used to check parameters are estimated in the new model.

$$\begin{aligned} &X \text{ as in } 1.1 \\ &Y' = \Sigma a_j h_j(U) + W \quad (\text{for example } h_j(x) = x^{j-1}) \end{aligned} \tag{1.3}$$

by considering orthogonalized h_j's, see Ferguson [5]. Since $(X-\alpha)^2$ is always convex Theorem 1 holds for the slope in the linear case when a constant term is in the model (1.1), (1.3).

COMMENT 2: While the results given here directly relate to the property of being biased toward or away from the origin they do not relate to attenuation of slope. Attenuation requires the extra condition that

$$P\left(\frac{\hat{\beta}}{\beta} > 0\right) \to 1 \quad \text{as} \quad n \to \infty .$$

To see that this extra condition may fail in Theorem 1 take

$$\begin{aligned} &f(x) = x^2 - 2cx, \text{ where } c \text{ is a fixed positive constant,} \\ &P(0 \leqslant U \leqslant c) = 1, \text{ and } P(|V| \geqslant 4c) = 1. \end{aligned}$$

References

[1] Mandansky, A. The Fitting of Straight Lines when Both Variables are Subject to Error, 1959 December; JASA **54**(285):173–205; 1959 March; JASA CORR 54(288):812.

[2] Moran, P. Estimating Structural and Functional Relationships. J. Multivariate Anal. 1(2): 232–255; 1971 June.
[3] Kendall, M. and Stuart, A. *The Advanced Theory of Statistics*, 4th Edn., Chapter 29, volume **2**, New York: Macmillan; 1979. 748 p.
[4] Carino, N.J. and Lew, H.S. Re-Examination of the Relation Between Splitting Tensile and Compressive Strength of Normal Weight Concrete (Manuscript); 1981.
[5] Ferguson, T.S. *Mathematical Statistics, A Decision Theoretic Approach*. Chapter 5.6, New York: Academic; 1967. 396 p.

JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 87, No. 1, January–February 1982

A Univariate Inequality for Medians

Clifford Spiegelman*

National Bureau of Standards, Washington, DC 20234

September 8, 1981

An inequality is provided for medians which is an analog of a theorem due to Karamata, dealing with majorization.

Key words: concave; convex; inequality; majorization; median.

There has been a great deal of recent emphasis on majorization and related correlation inequalities, for example Marshall and Olkin [1][1] and Karlin and Rinott [2]. These inequalities have a variety of important applications. However they are based solely on mathematical expectations, for example the result below due to Karamata [3]. After stating this result we give our analog for medians.

Karamata's result (THEOREM): *Let* F *and* G *be distribution functions, on* (c, d). *Let* $\mu = F - G$. *Then* $\int \phi d\mu \geq 0$ *for all convex functions,* $\phi:(c,d)\mapsto R$, *if and only if*

$$\int_c^d d\mu = \int_c^d x d\mu = 0 \quad \textit{and}$$

$$\mu\,(c,x] \geq 0, \qquad c \leq x \leq d\ .$$

Also see Spiegelman [4], for a different presentation of the direct part.

It should be noted that the direct part of Karamata's result is a generalization of Jensen's inequality. As previously suggested, characterizations such as this hold for medians as well. We define a median M of a random variable X to satisfy

$$P(X \geq M) \geq 1/2 \leq P(X \leq M)\ .$$

It is clear that M may not be uniquely defined. In order to avoid technical difficulties we define a p-median. Suppose a and b are endpoints of the largest closed interval such that every point M

$$a \leq M \leq b \text{ is a median of } X\ .$$

Then for p, such that $0 \leq p \leq 1$, define the p-med $X = pa + (1 - p)b$.

The use of p-medians complicates the statement of the theorem below. If all the random variables in the remainder of this paper have unique medians and, in addition, the class of convex functions is reduced to the class of strictly convex functions, then we get a less complicated analog to Karamata's result. In the new statement of our result the p-median notation is replaced by the word median. However, even under these more restrictive conditions, the converse part of Karamata's theorem has no explicit analog. (This is easy to see for random variables X such that $p(X=\infty) = p(X=-\infty)>.25$, and a convex function defined on the extended real line.) Our analog to Karamata's converse requires the class of functions $\Psi = \{\psi\}$ where each $\psi: R \rightarrow \{0,1\}$ and has the form $\psi = 1-\chi_I$, where χ_I is the characteristic function of some bounded interval.

The following theorem gives an analog to the theorem of Karamata. The direct part is an extension of Tomkins' [5] version of Jensens' inequality for medians. Tomkins' inequality is for conditional medians and

---

* Center for Applied Mathematics, National Engineering Laboratory.

[1] Figures in brackets indicate literature references at the end of this paper.

a restatement of his theorem is too lengthy. His result for unconditional medians is cast in our notation at the end of the theorem's statement.

THEOREM: (Direct part) *Let* F *and* G *be distribution functions for the random variables* X *and* Y *respectively. Let* F *and* G *have a common set of medians (the interval)* [a,b].

Also assume:

$$\mathrm{F}(\mathrm{t}) \geqslant \mathrm{G}(\mathrm{t}) \qquad \mathrm{t} \leqslant \mathrm{a} \tag{1a}$$

$$\mathrm{G}(\mathrm{t}) \geqslant \mathrm{F}(\mathrm{t}) \qquad \mathrm{t} \geqslant \mathrm{b}\,. \tag{1b}$$

Then for any convex function ϕ *defined on the support of both* F *and* G *and for every* p, $0 \leqslant \mathrm{p} \leqslant 1$, *there exists a* q, $0 \leqslant \mathrm{q} \leqslant 1$ *such that*

$$\mathrm{q}\text{-}med\ \phi(\mathrm{X}) \geqslant \mathrm{p}\text{-}med\ \phi(\mathrm{Y}). \tag{2}$$

In addition, if ϕ *is monotonic over the range of all* p-*medians for* Y *and* $\mathrm{p} \leqslant .5$ *then* q *may be taken equal to* p.

A partial converse of this result holds which requires an extra condition on the class, Ψ·

$$\text{For some } p \text{ median of } X,\ M, \qquad \psi(M+) \quad \text{or} \quad \psi(M-)=0. \tag{3}$$

If for all ψ satisfying condition (3) above $\psi(X)$ is stochastically larger than $\psi(Y)$, i.e. $P(\psi(X)>t) \geqslant P(\psi(Y)>t)$, and if for all monotonic functions, ϕ, (2) holds, then (1a) and (1b) hold.

COMMENT: Tomkins' inequality when applied to unconditional medians is a special case of our direct part. Simply take Y to have unit mass at p-med X.
Proof of the direct part: 1a and 1b hold.
Case 1—ϕ is monotone. Then clearly

$$p\text{-med}\ \phi\,(X) = p\text{-med}\ \phi\,(Y).$$

Case 2—ϕ has a minimum at a point t

$$\text{i.e.,} \qquad \phi(z) \geqslant \phi\,(t) \text{ for all } z \varepsilon R.$$

Define $F^{-1}(q) = \inf\,\{x|F(x) \geqslant q\}$.

Let r-med $\phi(Y) = \phi(z_{rj})\ j=1, 2$ with $z_{r1} \leqslant z_{r2}$. It follows immediately from Tomkins' result that there exists an r such that $z_{r1} \leqslant p\text{-med } Y \leqslant z_{r2}$.

If all the p-medians of Y lie on one side of the value, t, where $\phi\,(z)$ takes its minimum and $p \leqslant .5$, it follows that $\phi(p(0\text{-med } Y) + (1-p)\,(1\text{-med } Y)) \leqslant p\ \phi\,((0\text{-med } Y) + (1-p)\ \phi\,(1\text{-med } Y))$. Since ϕ is monotonic in the interval between the 0 and 1 medians, it follows that $p\ \phi\,(0\text{-med } Y) + (1-p)\ \phi\,(1\text{-med } Y) \leqslant p\text{-med}\ \phi(Y)$. Thus in this case we may take $r=p$. Otherwise:

for $z \leqslant z_{r1}$

$$F(z) \geqslant G(z)$$

which implies

$$z \geqslant F^{-1}(G(z)).$$

Similarly for $z \geqslant z_{r2}$

$$z \leqslant F^{-1}(G(z)).$$

Note that if G is continuous and strictly monotonic $F^{-1}G(Y)$ may be taken equal to X. By definition

$$\begin{aligned} 1/2 &\leqslant P(\phi(Y) \geqslant r\text{-med } \phi(\mathrm{Y})) \\ &= P(Y \leqslant z_{r1}) + P(Y \geqslant z_{r2}) \,. \end{aligned}$$

Notice that

$$P(Y \leqslant z_{r1}) \leqslant P(F^{-1}G(Y) \leqslant z_{r1})$$

$$\text{and} \quad P(Y \geqslant z_{r2}) \leqslant P(F^{-1}G(Y) \geqslant z_{r2}) \quad \text{(by assumptions 1a and 1b).}$$

Therefore

$$P(\phi(X) \geqslant r\text{-med } \phi(Y)) \geqslant 1/2 \,.$$

If Y does not have a continuous distribution standard, approximation procedures can be applied to complete the proof.

The converse case:

Since $\phi(x) = x$ and $\phi(x) = -x$ are convex functions, if follows that p-med $X = p$-med Y. Suppose at some point to the left of the smallest median of X, $G(t) > F(t)$, then if

$$\psi_1(z) = \begin{cases} 1 \text{ if } z \leqslant t \\ 0 \text{ if } t < z \leqslant \sup_p \, (p\text{-median of } X) \\ 1 \text{ if } z > \sup_p \, (p\text{-median of } X) \end{cases}$$

$$\text{and} \quad \psi_2(z) = \begin{cases} 1 \text{ if } z \leqslant t \\ 0 \text{ if } t < z < \sup_p \, (p\text{-median of } X) \,. \\ 1 \text{ if } z \geqslant \sup_p \, (p\text{-median of } X) \end{cases}$$

Then either $\psi_1(Y)$ or $\psi_2(Y)$ is stochastically larger than $\psi_1(X)$ or $\psi_2(X)$, respectively.

Q.E.D.

COMMENT: A. Marshall in a private communication has pointed out that the proof of the direct part of the above theorem is easily expanded to include functions ϕ *s.t.* $\{x | \phi(x) < c\}$ is an interval.

Churchill Eisenhart, Editor, asked if the definition of median given by Dunham Jackson; Bulletin of the American Math Society, 1921 160–164, "Note on the Median of a Set of Numbers," also 1923, 17–20, "Note on Quartiles and Allied Measures," could be used as a unique choice of median in the preceding theorem? The answer is yes; I conjecture more is true. Let $0 < c < 1$ and let $L(X) = (1-c)\,X^p$ for $X \geqslant 0$, $L(X) = c(-X)^p$ for $X \leqslant 0$, $p \geqslant 1$. Let $A_p(L,X)$ denote the minimizer of $EL(X-a)$ with respect to a. Then for any convex function ϕ; $\lim_{p \to 1} A_p(L, \phi(X)) \geqslant \lim_{p \to 1} \phi(A_p(L,X))$. A detailed proof will appear later. Thus, Jackson's definition of quantiles also satisfies a version of Jensen's inequality. Versions of Karamata's theorem also hold.

References

[1] Marshall, A.; Olkin, I. *Inequalities: Theory of majorization and its applications.* New York: Academic Press; 1979. 569p.

[2] Karlin, S.; Rrinott, Y. Classes of orderings of measures and related correlation inequalities. 1. Multivariate Totally positive distributions. J. Mult. Anal. **10**(4): 467–498, 1980 December.

[3] Karamata, J. Sur Une inegalite relative aux fonctions convexes. Publ. Math. Univ. Belgrade **1**: 145–148 1932.

[4] Spiegelman, C. H. A univariate extension of Jensen's inequality, J. Res. Nat. Bur. Stand. (U.S.) **85**(5): 363–365; 1980 September–October.

[5] Tomkins, R. J. On conditional medians. Ann. Prob. **3**(2): 375–379; 1975 April.

JOURNAL OF RESEARCH of the National Bureau of Standards
Vol. 87, No. 1, January–February 1982

Minimum-Loop Realization of Degree Sequences*

A. J. Goldman† and R. H. Byrd†

The Johns Hopkins University, Baltimore, MD 21218

June 10, 1981

Given a finite sequence D of nonnegative integers, let $M(D)$ denote its maximum element and $S(D)$ its sum. It is known that D is realizable as a degree sequence by some graph if and only if $S(D)$ is even, and by a loopless graph if and only if the even integer $S(D) - 2M(D) \geqslant 0$. Here it is shown that if the even integer $2M(D) - S(D)$ is positive, then one-half this integer is the minimum number of loops in graphs realizing D, and that the minimum-loop realization is unique. These results are extended to a more general loop-cost minimization problem in which loops incident at different vertices can have different costs. The possible numbers of loops, in graphs realizing D, are also determined.

Key words: graph, loopless graph, degree sequence, incidence sequence, partition.

1. Introduction

This note deals with finite undirected graphs. Our usage of "graph" permits both *loops* (edges from a vertex to itself) and *multiple links* (bundles of two or more edges with the same pair of distinct endpoints). The *degree* of vertex v in graph G, denoted $d_G(v)$, is the number of incidences upon v of edges of G; here a loop is considered to be twice-incident upon its single endpoint. Any enumeration of the set $\{v_i\}_1^n$ of the vertices of G gives rise to a sequence $\{d_G(v_i)\}_1^n$ of nonnegative integers which is called a *degree sequence* of G; it is clearly unique up to permutations.

Given any sequence $D = \{d_i\}_1^n$ of nonnegative integers, we set $S(D) = \Sigma_1^n d_i$ and $M(D) = \max_i d_i$. If graph G is such that D is a degree sequence of G, we shall say that G *realizes* D. The theory of such realizations (and their analogs for directed graphs) has a considerable literature including papers [1], [5]–[7][1] on topics close to the present one; an extensive account is given, for example, in Chapter 6 of Chen [2]. Here we require only the two basic results of that theory ([3], [7]):

THEOREM A. *Sequence* D *is realized by some graph* G *if and only if* S(D) *(and hence* S(D) − 2M(D)*) is even; in that case* G *can be chosen free of multiple links.*

THEOREM B. *Sequence* D *is realized by some loopless graph if and only if* S(D) *is even and* 2M(D) ≤ S(D).

Our purpose here is to provide explicit statements and a convenient reference for some elementary results, probably largely of "folklore" nature, related to Theorem B. We shall determine the possible numbers of loops in graphs realizing a given sequence D, and also solve an associated loop-cost minimization problem in which loops incident at different vertices can have different costs.

2. Results and Analyses

Our first objective is to round out the information contained in Theorem B, by presenting the following result.

---

* AMS Subject Classification: 05C35, 05A17

† Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD 21218, and Center for Applied Mathematics, National Bureau of Standards, Washington, DC 20234.

[1] Numbers in brackets indicate literature references at the end of the paper.

THEOREM 1. *If sequence* D *has* $2M(D) - S(D) = 2L$ *where* L *is a positive integer, then* D *can be realized by a unique graph with* L *loops, but not by a graph with fewer loops. The "unique graph" has all its loops incident at the unique vertex of maximum degree.*

It will be convenient to base this theorem's proof on the following:

LEMMA. *With* D *as in* Theorem 1, *any graph which realizes* D *has at least* L *loops at its unique vertex of maximum degree.*

PROOF (of Lemma): (a) Let G be a graph which realizes D; choose the numbering so that $M(D) = d_1$. Since $2M(D) > S(D)$, v_1 will be the *only* vertex of maximum degree. Suppose G has L_1 loops at v_1.

(b) By Euler's handshaking lemma, the number of edges of G is $S(D)/2$. Of these, the $S(D)/2 - L_1$ which are not loops at v_1 each have at least one endpoint in $\{v_i\}_2^n$, and so each contributes either 1 or 2 to the sum $\Sigma_2^n d_i = S(D) - M(D)$. It follows that $S(D)/2 - L_1 \leqslant S(D) - M(D)$, yielding

$$L_1 \geqslant M(D) - S(D)/2 = L.$$

PROOF (of Theorem 1): (a) By the Lemma, no graph which realizes D can have fewer than L loops.

(b) Choose the numbering so that $M(D) = d_1$. Then a graph with L loops, which realizes D, is obtained by placing L loops at v_1 and drawing d_i edges from v_1 to v_i for $2 \leqslant i \leqslant n$; the correctness of this graph's degree at v_1 follows from

$$\sum\nolimits_2^n d_i + 2L = S(D) - M(D) + 2L = M(D) = d_1.$$

(c) Now let G be any graph which realizes D and has exactly L loops; by the Lemma, all these loops are incident at v_1. Let d_i^* be the number of edges from v_1 to v_i in G, for $2 \leqslant i \leqslant n$. Then on the one hand $\Sigma_2^n d_i^* = d_1 - 2L = \Sigma_2^n d_i$, and on the other hand $d_i^* \leqslant d_i$ for $2 \leqslant i \leqslant n$. It follows that $d_i^* = d_i$ for $2 \leqslant i \leqslant n$, so that G coincides with the graph constructed in (b). This completes the proof of Theorem 1.

We turn now to a more general problem. Suppose given a sequence $C = \{c_i\}_1^n$ of nonnegative real numbers, and interpret c_i as the "cost" per loop incident at v_i; i.e., if graph G with vertex-set $\{v_i\}_1^n$ has λ_i loops attached at v_i, then the total loop-cost of G is $\Sigma_1^n c_i \lambda_i$. We seek a graph G which realizes a given sequence D as degree-sequence, and does so at minimum total loop-cost. (Theorem 1 treated the special case in which all $c_i = 1$.)

By Theorem A, this problem has a solution if and only if $S(D) - 2M(D)$ is even. When this is the case and $2M(D) \leqslant S(D)$, it follows from Theorem B that the optimal solution is found as a loopless graph realizing D. The remaining possibility is resolved by the following theorem, which shows that the solution is essentially independent of the cost-structure C.

THEOREM 2. *Suppose* $2M(D) - S(D) = 2L$ *where* L *is a positive integer. Then an optimal solution, unique if all* $c_i > 0$, *is given by the "unique graph" of* THEOREM 1.

PROOF: This is an immediate consequence of the Lemma and Theorem 1.

The uniqueness assertions in Theorems 1 and 2 bear the same relation to uniqueness results by Hakimi [4], Owens and Trent [7] and Senior [8], as do the remaining assertions of Theorems 1 and 2 to Theorem B.

Finally, we wish to determine the possible numbers of loops in graphs which realize a given sequence D. Theorem B and Theorem 1 specify the minimum of these numbers; it remains to specify their maximum, and to ascertain which values between the two can actually arise. To this end it is convenient to define, for $D = \{d_i\}_1^n$, Odd (D) to be the cardinality of $\{i\colon d_i \text{ is odd}\}$.

THEOREM 3. *Sequence* D, *with its number of positive entries different from* 2 *and with* S(D) *even, is realized by a graph with precisely* k *loops if and only if* 2k *lies between* max (0, 2M(D) − S(D)) *and* S(D) − Odd(D) *inclusive*.

PROOF: (a) Theorems B and 1 give $\max(0, 2M(D) - S(D))$ as the minimum possible value for $2k$.

(b) From $D = \{d_i\}_1^n$ we determine a sequence $\Delta = \{\delta_i\}_1^n$ of nonnegative integers as follows: by requiring $d_i = 2\delta_i + 1$ if d_i is odd, $d_i = 2\delta_i$ if d_i is even. Then $S(D) = 2S(\Delta) + \mathrm{Odd}(D)$, so that $\mathrm{Odd}(D)$ is even. Clearly any graph with vertex-set $\{v_i\}_1^n$ that realizes D can have at most δ_i loops incident at v_i, thus at most $S(\Delta) = (S(D) - \mathrm{Odd}(D))/2$ loops in all. This upper bound is achieved by attaching δ_i loops to v_i for $1 \leq i \leq n$, pairing off in any way the members of the even-cardinality set $\{v_i\colon d_i \text{ is odd}\}$ counted by $\mathrm{Odd}(D)$, and joining the vertices in each pair by a single edge. So $S(D) - \mathrm{Odd}(D)$ is indeed the maximum value for $2k$.

(c) Beginning with the graph constructed in (b), repeat the following step as long as possible, producing a sequence of graphs each realizing D and having one fewer loop than its predecessor: if the current graph has three distinct vertices, v_i, v_p, v_q such that v_i bears a loop ℓ and some edge e joins v_p and v_q, then replace ℓ and e by a pair of edges from v_i to v_p and to v_q respectively.

Let G be the graph with which this process terminates and let λ_j be the number of loops of G at v_j $(1 \leq j \leq n)$, for a total of k loops. If $k=0$, we are done, so assume $k>0$. For any vertex v_i such that $\lambda_i>0$, it follows from the construction of G that

$$d_i - 2\lambda_i = \textstyle\sum_{j \neq i} (d_j - 2\lambda_j). \qquad (*)$$

It follows that there is either just one such vertex, say v_1, or else exactly two, say v_1 and v_2, with $d_1 - 2\lambda_1 = d_2 - 2\lambda_2$ and with $d_j = 2\lambda_j = 0$ for all $j>2$. The latter case is ruled out by the theorem's hypothesis on D. In the former case, (*) yields

$$d_1 - 2\lambda_1 = S(D) - d_1,$$

from which it readily follows that

$$2k = 2\lambda_1 = 2d_1 - S(D) = 2M(D) - S(D);$$

again we are done.

It only remains to treat the exceptional case excluded by the hypothesis of Theorem 3.

THEOREM 4. *Sequence* D, *with exactly two positive entries and with* S(D) *even, is realized by a graph with precisely* k *loops if and only if* 2k *lies between* 2M(D) − S(D) *and* S(D) − Odd(D) *inclusive and* 2k ≡ 2M(D) − S(D) (mod 4).

PROOF: (a) Number so that d_1 and d_2, with $d_1 \geq d_2$, are the two positive entries of D. Since $S(D) = d_1 + d_2$ is even, d_1 and d_2 have the same parity; $\mathrm{Odd}(D)$ is 2 or 0 according as the parity is odd or even.

(b) The arguments in (a) and (b) of Theorem 3's proof still apply, to show that $S(D) - \mathrm{Odd}(D)$ and $2M(D) - S(D)$ are respectively double the maximum and minimum numbers of loops in graphs that realize D. These extreme values of $2k$ differ by $2d_2 - \mathrm{Odd}(D)$, a multiple of 4.

(c) Define $\Delta = \{\delta_i\}_1^n$ as in the proof of Theorem 3. Form a graph realizing D which has δ_1 loops incident at v_1, δ_2 loops incident at v_2, and $\mathrm{Odd}(D)$ edges between v_1 and v_2. Then repeat the following step as long as possible, producing a sequence of graphs each realizing D and having two fewer loops (hence, a value of $2k$ less by 4) than its predecessor: if the current graph has loops at both v_1 and v_2, then delete one loop at each of these vertices and replace them by two new edges between v_1 and v_2.

The final graph in this process has a number k of loops (all at v_1) given by

$$2k = 2(\delta_1 - \delta_2) = d_1 - d_2 = 2M(D) - S(D).$$

Thus all values of $2k$ identified in Theorem 4's statement are indeed achieved.

(d) To show that no other values can be achieved, consider any graph realizing D, with λ_j loops at v_j ($j = 1,2$) for a total of $k = \lambda_1 + \lambda_2$ loops. Counting the edges from v_1 to v_2 in two different ways (by incidences on v_1 and by incidences on v_2) yields the relation $d_1 - 2\lambda_1 = d_2 - 2\lambda_2$, so that

$$2k = 2(\lambda_1 - \lambda_2) + 4\lambda_2 = (d_1 - d_2) + 4\lambda_2 = 2M(D) - S(D) + 4\lambda_2;$$

thus the residue of $2k$ (mod 4) is as stated in the theorem.

The following observation is included for completeness. Let $\lambda_G(v)$ denote the number of loops incident on vertex v in graph G with vertex-set $\{v_i\}_1^n$; then $\Lambda(G) = \{\lambda_G(v_i)\}_1^n$ is the *loop-sequence* of G corresponding to this enumeration of the vertices. Given a pair (D,Λ) of sequences $D = \{d_i\}_1^n$ and $\Lambda = \{\lambda_i\}_1^n$ of nonnegative integers, it is natural to ask whether there exists a graph G with $D = D(G)$ and $\Lambda = \Lambda(G)$. But this is the case if and only if $D - 2\Lambda = \{d_i - 2\lambda_i\}_1^n$ is the degree sequence of a *loopless* graph, and a necessary and sufficient condition for that to hold is found by applying Theorem B to $D - 2\Lambda$.

3. References

[1] Ayoub, J. N.; Frisch, I. T. Degree realization of undirected graphs in reduced form. J. Franklin Institute, **289** (4): 303–312; 1970 April.

[2] Chen, W. K. *Applied Graph Theory*. Amsterdam: North-Holland Publishing Co.; 1976. 542 p.

[3] Hakimi, S. L. On realizability of a set of integers as degrees of the vertices of a linear graph I. J. Soc. Indust. Appl. Math. **10** (3): 496–506; 1962 September.

[4] Hakimi, S. L. On realizability of a set of integers as degrees of the vertices of a linear graph II. Uniqueness. J. Soc. Indust. Appl. Math. **11** (1): 135–147; 1963 March.

[5] Kleitman, D. J. Minimal number of multiple edges in realization of an incidence sequence without loops. SIAM J. Appl. Math. **18** (1): 25–28; 1970 January.

[6] Owens, A. B. On determining the minimum number of multiple edges for an incidence sequence. SIAM J. Appl. Math. **18** (1): 238–240; 1970 January.

[7] Owens, A. B.; Trent, H. M. On determining minimal singularities for the realization of an incidence sequence. SIAM J. Appl. Math. **15** (2): 406–418; 1967 March.

[8] Senior, J. K. Partitions and their representative graphs. Amer. J. Math. **73** (3): 663–689; 1951 July.