NEWSLETTER

OF THE NEA DATA BANK

BULLETIN

DE LA BANQUE DE DONNEES DE L'AEN

JULY 1982

No.27

JUILLET 1982

NUCLEAR ENERGY AGENCY
AGENCE POUR L'ENERGIE NUCLEAIRE

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
ORGANISATION DE COOPERATION ET DE DEVELOPPEMENT ECONOMIQUES

FOREWORD

The programs compared in this report are all designed to calculate average resonance parameters from experimentally derived resonance width data. These experimental neutron width data approximate to a Porter-Thomas distribution of s-wave reduced resonance widths; in fact they include p-wave resonances obeying different distribution functions, and are distorted by experimental errors, adjacent peaks detected as single resonances, and by missed resonances too weak for detection. The authors of the codes have attempted to eliminate the effect of p-wave resonances and experimental distortions, so as to define the Porter-Thomas distribution which best conforms to the "clean" residual distribution of reduced widths for s-wave resonances. The codes require a certain amount of subjective judgement on the part of the user.

The statistics of a Porter-Thomas distribution are well-known and are not difficult to incorporate into a computer code. The techniques for estimating the effect of distortions are not so well-defined, and often involve input parameters with only indirect physical significance, but which nevertheless must be adjusted by the user to "optimise" the results. The judgement required for this optimization is much more easily applied by the program author, who has an intimate knowledge of the coding, than by the general user, whose understanding is based on the documentation supplied with the code and his experience of running it. The present exercise was suggested by the significant differences observed between the average resonance parameters calculated by different laboratories using different codes but the same experimental data. Pseudo-experimental data sets were generated and average parameters calculated using nine different codes run by their authors.

These first results were surprisingly bad; there was disagreement as to whether the errors were caused by inadequate mathematical and numerical techniques, or by misinterpretation of input data. A second stage of the benchmark exercise was proposed, with a new set of artificially generated resonance data, and followed by a workshop held at NEA Data Bank, where the codes were discussed and some errors and inefficiencies eliminated.

A benchmark exercise can be said to be successful when all solutions obtained converge towards a real but unknown result, or as in the present case when codes give the "correct" solution to a fictitious problem which is believed to be sufficiently realistic. The subjective parameter adjustment in successive runs of many of the codes tested can only be reproduced successfully by users other than the author if the code is well documented. It is for this reason that the final "blind" runs of participating codes were made by a physicist with no previous knowledge of the problem, who was thus testing both the codes and their user documentation.

The two codes made available for general distribution by their authors both came out well in the comparison. They are BAYESZ by M. Moore of LANL, and ESTIMA by E. Fort of CEN Cadarache. They can be obtained on request from the Data Bank, as can also the two data sets used in the exercise and generated by Dr. P. Ribon of CEN Saclay. Comparison runs using these data could prove extremely useful in validating any new codes in this field.

The authors of the three articles describing successive stages of the study are Dr. M. Moore, Mr. A. Thompson and Dr. P. Ribon. Data Bank participants in the study were Dr. P. Johnston, Dr. E. Sartori, Mr. A. Thompson, Dr. N. Tubbs and Mr. J. Vanne.

LIST OF PHYSICISTS PROVIDING SOLUTIONS

First stage (Test cases 1A, 1B, 1C, 5A, 5B, 5C)

M. Caner	SOREQ, Israel
H. Derrien	CEN, Cadarache, France
E. Fort	CEN, Cadarache, France
F.H. Fröhner	KFK, Fed. Rep. of Germany
H. Gruppelaar	ECN, Petten, the Netherlands
Y.S. Gur	SOREQ, Israel
M.S. Moore	LANL, USA
G. Rohr	CBNM, Geel, Belgium
H. Weigmann	CBNM, Geel, Belgium

Second stage (test cases A1, A2, B1, B2, C1, C2)

E. Fort	CEN, Cadarache, France
F.H. Fröhner	KFK, Fed. Rep. of Germany
H. Gruppelaar	ECN, Petten, the Netherlands
Y.S. Gur	SOREQ, Israel
M.S. Moore	LANL, USA
G. Rohr	CBNM, Geel, Belgium
H. Weigmann	CBNM, Geel, Belgium

Third stage (Test cases B1, C1, C2)

The codes of the following authors were used:

E. Fort, CEN Cadarache : ESTIMA

F. Fröhner, KFK Karlsruhe : STARA

H. Gruppelaar, ECN Petten : CAVECN
(the ECN version of the ENEA Bologna code CAVE by M. Stefanon)

M. Moore, LANL, USA : BAYESX
(BAYESX was replaced by BAYESZ for the third stage of the exercise)

TABLE OF CONTENTS

- - - - - -

- - - - - -

NEA DATA BANK
BANQUE DE DONNES de L'AEN
F 91191 Gif-sur-Yvette CEDEX
France
Tel: 941 39 65/908 49 12
Telex NEADATA 690920 F
Telegr. NUCLAGENCE GIF - 91191 Gif-sur-Yvette

WORKSHOP ON THE DERIVATION OF AVERAGE RESONANCE PARAMETERS

15th and 16th October, 1981

OECD Nuclear Energy Agency Data Bank, Saclay, France

Dr. M. Moore, Los Alamos National Laboratory, USA

I. INTRODUCTION

At the 1977 Specialists' Meeting on Fission Product Nuclear Data in Petten, the Netherlands (IAEA-213 (1978)), E. Fort pointed out a significant discrepancy in reported average resonance parameters. He concluded that the procedures used by certain evaluators must be of questionable validity, because the derived average parameters, all using the same set of input data, differed by a factor of two in some cases. Fort suggested that the discrepancies in methods and procedures he had observed might best be resolved by an international exercise using a set of input data that simulated actual experimental data, but in which the true average values were known.

Such a discrepancy in average resonance parameters is not purely of academic interest. The evaluation of average capture cross sections for fast reactors relies on optical and statistical model calculations; these use average parameters from the resolved resonance range as basic input. An error of 10 percent (or a factor of 2) in the s-wave average spacing is often propagated linearly into the calculated average capture cross section at higher energies. It is thus of considerable practical interest that the methods used to obtain average parameters give reliable results.

At the 1979 Specialists' Meeting on Fission Product Nuclear Data in Bologna, Italy (NEANDC(E)209"L"), Fort extended his comparison of methods to several additional nuclides, showing that the discrepancies among these methods seemed to be fairly general. At the same meeting, Ribon proposed an intercomparison of methods to determine average parameters from sets of resonance parameters. The parameter sets would be generated, Doppler and resolution broadened, and the resulting parameter set would be obtained from the calculated cross sections, which would include a statistical fluctuation to simulate actual data. Ribon volunteered to organise and evaluate the results of the exercise. Ribon's proposal was accepted, and the parameter sets he provided were distributed through the four data centres in September 1980.

II. THE FIRST EXERCISE

The first exercise consisted of six data sets, three for each of two representative fission-product nuclei. The two nuclei were to be characteristic of a spin 5/2 target in the mass region A=150 and a spin 1/2 target

near A=100. In the first case, one might expect somewhat better agreement: the s-wave neutron strength is rather large near A=150, so the input data would be reasonably free from contamination by p-wave resonance effects. The second case is perhaps representative of the most difficult problem to be encountered; the s-wave strength is low, the p-wave strength is large, and the low target spin suggests that the energy range to be considered is large enough that at the high-energy end of the range the s- and p-wave input data will be thoroughly mixed. Three cases for each nucleus were supplied to permit systematic and random errors to be differentiated.

Nine physicists participated in the first exericse: M. Caner, SOREQ, Israel; H. Derrien, CEN, Cadarache, France; E. Fort, CEN, Cadarache, France; F.H. Fröhner, KFK, Karlsruhe, F.R. Germany; H. Gruppelaar, ECN, Petten, the Netherlands; Y.S. Gur, SOREQ, Israel; M.S. Moore, Los Alamos National Laboratory, USA; G. Rohr, CBNM, Geel, Belgium; and H. Weigmann, CBNM, Geel, Belgium. The solutions were identified by the letters A to I, assigned randomly, and are compared with one another and with the true values in Tables 1 and 2 and figures 1 to 6. The results of this exercise, circulated to participants in March 1981, came as a surprise to all who were involved. For the difficult exercise (Table 1), the average parameters reported differed systematically by up to a factor of two, and the discrepancies were generally well outside the uncertainties assigned by the participants. For the easier exercise (Table 2), the results were perhaps even more disturbing: the reported average spacings agreed fairly well with one another, within the errors assigned, but were systematically wrong. Almost invariably, the reported values were too large by an amount that exceeded the assigned uncertainty.

The systematic bias introduced into the results is caused by one of the following:

either (i) the physicist's misinterpretation of the output of the codes,

or (ii) the physics and mathematics approximations incorporated into the coding.

Ribon argues (NEANDC(E)213-AL) that (i) is the correct conclusion - the discrepancies were of the same order of magnitude for a variety of different algorithms. Certain of the participants disagreed with this. For example, Moore felt that the problem was related to correcting the truncated Porter-Thomas distribution for overlapping resonances. Two large, closely spaced resonances are likely to be counted as one extra large resonance, especially at high energies.

Fröhner also recognised that the systematic discrepancy for the spin 5/2, A=150 nucleus in the exercise was due to resonance overlap and devised a method of correcting the Porter-Thomas distribution for this effect before applying the method of maximum likelihood in his code STARA. He reported the modifications at the IAEA Specialists' Meeting on Resonance Parameters of Uranium and Plutonium Isotopes, 28th September to 2nd October 1981.

The following comments can also be made:

a) Case 5 (Table 2 and Figures 4 to 6)

This is the simplest case of the first exercise, simulating a spin 5/2 target with A=150; the energy range spanned was 0 to 320 eV. The results were reasonably consistent; the s-wave strength function was calculated to within 1-3 percent and showed no systematic deviation; the spacings were calculated with a dispersion of 1-4 percent, and were systematically about 7 percent too large. The systematic error was found to be due to levels missed because of resonance overlap (Reference 1). If the various codes were to be modified to take this effect into account (as was done by both Fröhner and Moore, and presumably could be done by any of the participants), we expect that the solutions obtained by any method will give satisfactory agreement within the limitations of the method. In other words, this problem can be considered to be solved. Probably 80-90 percent of the cases encountered in the analysis of real data are no more complex than this.

b) Case 1 (Table 1 and Figures 1 to 3)

Perhaps this is representative of the most difficult problem one might expect to encounter in the analysis of actual data. From Table 1, it is seen that the s-wave strength function discrepancies are clearly related to the energy range considered by the evaluator. If the whole range is considered, the solutions are found to be reasonably close to the true values, but if the range is truncated, the solutions obtained are lower. As it turned out, this is a real effect; the three sets of data generated by Ribon were not statistically independent. However, it is much more difficult to explain the discrepancies in the s-wave average spacings. We had thought that the discrepant results might be attributable to different ways of applying Bayes' theorem in distinguishing s- and p-wave contributions (the classic reference is not correct for this problem). But this seems to be not the case; the application of Bayes' theorem appears to have been done in an appropriate manner by the participants using it. There are, of course, strong correlations between s- and p-wave strengths and spacings, and it is extremely difficult to trace the discrepancies to a single cause.

III. THE SECOND EXERCISE (Tables 3 to 5 and Figures 7 to 10)

Most of the methods do depend to some extent on the judgement of the physicist using them, and in such cases it is not difficult to reconcile the derived parameters with the true value a posteriori. Recognising this, Ribon generated a second exercise very similar to the difficult nucleus of the first exercise (I=1/2, A=100) but in three steps that supposedly would permit the participant to trace the deficiency in his method. Step A consisted simply of a list of s-wave neutron reduced widths, the objective being to determine the average value. Step B consisted of a complete set of parameters as input data, undistorted by any kind of simulated experimental broadening, statistical spread, or overlap. Step C consisted of a set of parameters, similar to those supplied in the first exercise, that simulated what would be reported in an actual measurement. This second exercise was distributed to participants in June 1981, and formed the basis for discussion at the Workshop held from 15th to 16th October 1981, at the OECD Nuclear Energy Agency Nuclear Data Bank at Saclay, France.

Participants in the second exercise consisted of seven of the nine physicists who had submitted earlier solutions; the workshop was also attended by Dr. J.S. Story of the UK Atomic Energy Establishment at Winfrith. The submitted solutions are compared with one another and with the true values in Tables 3 to 5; the anonymity of the participants is preserved by arranging the results as in Tables 1 and 2. Numbers in parentheses in Tables 4 and 5 reflect discarded a priori solutions obtained while the participants worked out the exercise but which were not reported until after the true solutions were known. They reflect to some extent the dependence of the method on the judgement of the users.

All the methods used by the participants in the second exercise are similar in that they are based on the properties of the reduced neutron-width distribution; for s-wave neutrons this is basically a truncated Porter-Thomas distribution, which may be modified to account for resonance overlap. It seems to be widely recognised that while the Dyson-Mehta Δ_3 statistic (F.J. Dyson and M.L. Mehta, J. Math. Phys. 4, 701 (1963)) is extremely powerful for nearly perfect input data, it is not a useful tool for the determination of average parameters from realistic input.

However, even though all the methods are based on the same general approach, they differ widely in implementation. At the workshop meeting, the methods were categorised into two groups: those that require one or several minutes on a high-speed modern computer, and those that take only a few seconds. A brief description of each of the codes was given at the meeting, and is summarised in the Appendix.

Although the participants were not aware of this, the Porter-Thomas distributions were the same for cases A1, B2, and C1, and for A2, B1, and C2. The s-wave width distribution for A1-B2-C1 was a fairly improbable one; it contained more large levels and fewer average-sized ones than a typical Porter-Thomas distribution. While this should present no problem to those methods using maximum likelihood (provided the truncation threshold was not too high), the method of moments would be expected to give much too large a spacing, and for step A, too large a value of the average reduced width. This, indeed, proved to be so, and the effect persisted, as expected, through all three steps.

Ribon expected step B to be easier than step C. The distributions in step B were complete and contained no experimental effects, missing levels, or resonance overlap. They were, however, mixed s- and p-wave distributions, with no indication of any kind to permit the distinguishing of the members of the two sets. For this reason, most of the participants felt that Cases B1 and B2 were actually more difficult than C1 and C2; in step C, the fitting codes used to provide "experimental" input data were able to point out the largest p-wave levels with reasonable accuracy. Those methods that do not use such information might be expected to be somewhat less reliable.

The results of the second exercise are, unfortunately, still not completely satisfactory. While the very large discrepancies observed in the first exercise appear to have been eliminated, even the most elaborate methods occasionally seem to show unexplained disagreements with the true values of 20 to 30 percent for s-wave parameters, and up to a factor of two in p-wave estimates.

IV. RECOMMENDATIONS

The time was too short at the Saclay workshop meeting to permit any exhaustive study of the problem areas in the various codes. (Only one or two tests could be run for any of the codes during the two-day meeting, even though all the codes that had been submitted in advance were operable at the NEA Data Bank.) A number of possible deficiencies in approach were noted, however, for the participants to consider. It was agreed that two steps should be implemented: (1) each of the codes, corrected if necessary, will be submitted to the NEA Nuclear Data Bank, with a complete description and set of instructions for using it, and (2) the codes will then be exchanged among participants interested in performing comparative studies. It was noted that while the existing exercises serve as an adequate benchmark for testing the codes, the degree of subjectivity inherent in many of the methods can permit the user to get a right answer with a deficient code if he knows what that answer should be. Therefore, Ribon agreed to consider generating a third exercise that would be representative of a wide range of input data to be encountered in resonance evaluation throughout the table of nuclides.

SUMMARY OF CONCLUSIONS

i) Stage 1

a) The results calculated were much less accurate than expected. For the easier test, the average parameters computed by different codes agreed with each other within quoted errors, but were systematically wrong; for the more difficult case, the computed results disagreed by up to a factor of 2.

b) There was disagreement over whether these errors were due to deficiencies in the methods used by the codes, or in the interpretation of the numerical output.

c) The fact that the results were unexpectedly bad demonstrates the importance of undertaking a benchmark test.

ii) Stage 2

a) While the very large discrepancies observed in the first exercise appear to have been eliminated, even the most elaborate methods occasionally seem to show unexplained disagreements with the true values of 20 to 30 percent for s-wave parameters, and up to a factor of two in p-wave parameters.

b) Since all the codes have been run by their own authors, the importance of the subjective judgement required to interpret the numerical output is not clear.

Reference

1) F.H. Fröhner "Level Density Estimation with Account of Unrecognised Multiplets Applied to Uranium and Plutonium Resonance Data". Contribution to the IAEA Specialists' Meeting on Uranium and Plutonium Resonance Parameters, Vienna, 28th September to 2nd October, 1981.

Appendix

DESCRIPTION OF THE METHODS

Perhaps the most elaborate of the codes is STARA, written by F.H. Fröhner. It performs a maximum likelihood fit of the neutron width to a composite distribution consisting of a weighted sum of two Porter-Thomas distributions (for s- and p-wave), and a p-wave exponential. An energy-dependent, diffuse minimum detection threshold is also included in the likelihood function. The two parameters automatically adjusted are the mean s-wave neutron width and a quantity specifying the detection threshold. The latest modification is the taking account of levels lost due to unresolved multiplets; earlier versions of STARA were described in "Nuclear Theory for Applications", Trieste, IAEA-SMR-43 (1980), p.90, and at the Bologna Specialists' Meeting on Neutron Cross Sections of Fission Product Nuclei, RIT/FIS-LDN(80)1, NEANDC(E)209L (1980), p. 145. The modification to account for resonance overlap was described at the Vienna Specialists' Meeting on Resonance Parameters of Uranium and Plutonium Isotopes, IAEA, to be published (1981). While in principle the STARA code should be capable of determining p-wave as well as s-wave average parameters, only the s-wave parameters were reported at the Saclay workshop meeting (Tables 3 to 5).

The code CAVE, by M. Stefanon, again uses the maximum-likelihood method to determine the average value and threshold of a truncated Porter-Thomas distribution for even-even targets. The threshold is energy-dependent but sharp rather than diffuse, and no correction is made for unresolved doublets. Under the assumption that D varies as (1/2J+1), an estimate is also obtained for the p-wave strength. The code was modified slightly by H. Gruppelaar and G. Delfini at ECN, to permit it to be used for target spins other than zero; it is the Gruppelaar results, obtained with the modified code CAVECN, that are reported in Tables 3 to 5. The code was described by Dr. Stefanon at the Bologna Specialists' Meeting in 1979 (M. Stefanon, RIT/FIS-LDN(80)1, NEANDC(E)209"L"9 (1980), p. 161), and Miss Delfini described the modification and some results obtained for a large class of fission product nuclides at the same meeting (G. Delfini and H. Gruppelaar, ibid., p. 169).

The ESTIMA code, used at Cadarache by E. Fort and H. Derrien, again uses the maximum-likelihood method of determining the average value and total number of resonances of a truncated Porter-Thomas distribution as a function of the threshold; the solutions are generally found to exhibit a stabilisation plateau, and are judged to be "most physical" when this plateau is reached. The truncation threshold is taken high enough to exclude all p-wave levels, and the energy interval is restricted to exclude unresolved doublets. The ESTIMA maximum-likelihood code was described in several reports. (P. Ribon, E. Fort, Krebs, and Tran Quoc Thuong, CEA-N-1832 (1975); Tran Quoc Thuong, EANDC(E)-160AL; and E. Fort, Specialists' Meeting on Fission Product Nuclear Data, Petten (1977)). It should be noted that Fort's solutions to the second exercise were obtained not only with the ESTIMA maximum-likelihood method, but also with the Keyworth-Moore missing-level estimator with weighting factors of unity, and by least-squares fitting of the s-wave distributions, described below. Judgement is therefore required to decide how to weight the three solutions obtained.

The last of the codes that require a relatively large amount of computer time is BAYESX, the Los Alamos code developed by M.S. Moore. The method of moments is used to determine the average value and total number of levels of a truncated Porter-Thomas distribution. The method was first described in a paper by Keyworth, Moore, and Moses at the Specialists' Meeting on Fast Neutron Fission (ANL-76-90, NEANDC(US)-199/L (1976), p. 353), and was later extended by using Bayes' equations to modify the weights in the various sums, which permitted the determination of p-wave strengths and spacings. After the first exercise, the method was further extended to account for resonance overlap. It should be noted that while all parameters (d_o, S_o, d_1, S_1) are reported, there is a parameter (the threshold above which all levels are assumed to have known s- or p-wave character) that is in practice used by the evaluator to give the appropriate s- to p-wave spacing.

Weigmann's code performs a least-squares fit to the truncated distorted Porter-Thomas distribution. In practice, the low-energy data are used to determine the s-wave spacing distribution (this procedure is found to give the least amount of p-wave contamination), then all the data above a given threshold are fitted to give the s-wave average width. Finally, the p-wave average width is determined by least-squares fitting of the difference distribution to an appropriately (2J+1) weighted sum of Porter-Thomas and exponential distributions, under the assumption that the p-wave spacing is known (from the s-wave spacing). Judgement of the evaluator is used to select the energy ranges and threshold levels that lead to satisfactory values of chi-square in the fitting.

The MISDO code of G. Rohr uses Bayes' equations to eliminate p-wave contamination of the observed input distribution, and then uses the Fuketa-Harvey technique (Nuclear Instruments and Methods 33, 107 (1965)), with an energy-dependent threshold, to obtain the corrected average value and number of levels. Rohr's method was described in a paper by Rohr, Maisano, and Shelley at the 1979 Bologna Specialists' Meeting (RIT/FIS-LDN(80)1, NEANDC(E)209"L", p. 197 (1980)). It may be noted that the Fuketa-Harvey method uses an iterative technique to correct the observed first moment of a truncated Porter-Thomas distribution to the expected first moment of the complete distribution, but no fitting procedures or higher-moment information are used. The code of Y. Gur of SOREQ appears to be rather similar to that of Rohr.

J.S. Story reviewed two codes in use at AEEW; the level density code DBAR uses the Dyson-Mehta W statistic for just the first few levels; the code LJPROB appears again to be similar to that of Rohr et al., and Gur, in that it uses Bayes' equations on the integral truncated Porter-Thomas distribution to descriminate against p-wave levels and the Fuketa-Harvey approach to correct the average value for missing levels. Story had not completed the second exercise. His studies on the first exercise were done a posteriori, and the results not included in Tables 1 and 2.

Table 1. Results for spin 1/2 nuclei (A=100) in the first exercise

	ENERGY RANGE (eV)	CASE 1A Do (eV)	D_1 (eV)	So ($\cdot 10^{-4}$)	S_1 ($\cdot 10^{-4}$)	$\Gamma\gamma o$ (eV)	CASE 1B Do (eV)	D_1 (eV)	So ($\cdot 10^{-4}$)	S_1 ($\cdot 10^{-4}$)	$\Gamma\gamma o$ (eV)	CASE 1C Do (eV)	D_1 (eV)	So ($\cdot 10^{-4}$)	S_1 ($\cdot 10^{-4}$)	$\Gamma\gamma o$ (eV)
THEORETICAL VALUE		44.09	20.82	0.42	5.1	0.159	36.67	17.58	0.41	5.5	0.165	38.91	18.36	0.45	6.0	0.151
TRUE VALUE	5200	43.70	20.77	0.415	5.05	0.158	36.43	17.60	0.415	5.3	0.165	39.10	18.19	0.453	5.6	0.150
A	5200	39.±4.		$0.42^{+0.06}_{-0.04}$		0.157±0.008	31.±3.		$0.45^{+0.05}_{-0.04}$		0.164±0.007	31.±3.		$0.48^{+0.06}_{-0.05}$		0.167±0.012
B	5200	48.9±6.6	21.4±2.0	0.408±0.006	5.7±0.5		39.1±4.8	17.8±1.4	0.414±0.009	6.49±0.53		42.±5.3	18.8±1.5	0.449±0.008	6.63±0.56	
C	1800	45.±3.		$0.28^{+0.08}_{-0.06}$	5.±1.2		38.±2.		$0.27^{+0.07}_{-0.06}$	4.2±1.0		38.±2.		$0.30^{+0.07}_{-0.06}$	5.2±1.2	
D	5000.	17.5		0.49		0.165	15.8		0.50		0.162	16.4		0.54		0.158
E	2990/2840	36.4±1.8		0.35±0.025	6.3		33.5±1.7		0.33±0.025	5.75		32.7±1.8		0.389±0.027	7.1	
	500-3000	35.8±2.5		0.38±0.04	6.3		31.±2.		0.385±0.04	5.75		33.±2.3		0.399±0.040	7.1	
F	1500.	33.0±3.2		0.30±0.09			30.8±2.9		0.28±0.07			38.±3.6		0.28±0.08		
G	5189.						34.5±5.0	18.5	0.43	4.9						
H	2600	32. 3.		0.34±0.06	6.		28.±3.		0.35±0.06			30.±3.		0.37±0.06	6.	
I						0.175±0.010					0.170±0.006					0.165±0.007

Table 2. Results for nuclei 5

		Energy range	Nuclei 5 - A $\overline{D}(\ell = o)$	$S^o(10^{-4})$	$\overline{\Gamma}_\gamma(\ell = o)$	Nuclei 5 - B $\overline{D}(\ell = o$	$S^o(10^{-4})$	$\overline{\Gamma}_\gamma(\ell = o)$	Nuclei 5 - C $\overline{D}(\ell = o)$	$S^o(10^{-4})$	$\overline{\Gamma}_\gamma(\ell = o$	Remarks
Theoritical value			1.850	2.2	0.072	1.425	2.5	0.080	1.811	1.8	0.075	
True value		320 - 310	1.849	2.22	0.073	1.428	2.47	0.081	1.824	1.785	0.076	
Results from	A	320	2.02±0.08	$2.23^{+0.30}_{-0.24}$	0.075±0.005	1.56±0.05	$2.52^{+0.34}_{-0.28}$	0.079±0.007	1.90±0.09	$1.81^{+0.25}_{-0.19}$	0.081±0.006	smaller E range for Γ_γ
	B	320	1.95±0.15	2.23±0.03		1.54±0.11	2.55±0.04		1.98±0.16	1.80±0.03		
	C	320	1.85±0.10	2.24±0.3		1.52±0.08	2.53±0.3		1.92±0.10	1.81±0.25		
	D 1	300	2.27	2.24	0.077	1.84	2.53	0.087	2.20	1.81	0.081	
	D 2(R)	300	2.15	2.24	0.077	1.79	2.53	0.087	2.13	1.81	0.081	recommended
	D 3		2.16	1.84	0.074	1.56	2.47	0.081	2.21	1.36	0.079	ΔE for
	D 4		2.04	1.85	0.074	1.51	2.59	0.081	2.04	1.37	0.079	50 resonances
	D 5		1.96	2.33	0.078	1.56	2.65	0.088	1.84	1.92	0.082	
	E	81/140/199	1.89±0.10	2.15±0.20		1.58±0.07	2.50±0.16		1.98±0.12	1.76±0.14		
	F	320	1.93±0.09	2.25±0.33		1.52±0.07	2.56±0.38		1.94±0.09	1.82±0.27		
	G	180	1.95±0.29	2.10								
	H	320	1.88±0.14	2.3 ±0.3		1.55±0.10	2.6±0.3		1.92±0.15	1.8±0.3		≈ same result with ΔE = 160 eV
	I				0.077±0.002			0.087±0.002			0.082±0.002	

TABLE 3

Results for Case A in the Second Exercise

(Objective: to find the average value of $2g\Gamma_n^o$, treating the width distribution as a truncated Porter-Thomas distribution)

	Case A1	Case A2
Theoretical Value	1.702	1.648
True Value	1.91	1.46
A	1.91 + 0.25	1.45 + 0.19
B	2.12 + 0.19	1.50 + 0.14
C	1.98 + 0.28	1.51 + 0.21
D		
E	1.89	1.495
F		
H	1.91	1.465

TABLE 4

Results for Case B in the Second Exercise

(To find average parameters from input data that contain no experimental effects)

	CASE B2				CASE B1			
	D_o (eV)	S_o ($\cdot 10^{-4}$)	D_1 (eV)	S_1 ($\cdot 10^{-4}$)	D_o (eV)	S_o ($\cdot 10^{-4}$)	D_1 (eV)	S_1 (10^{-4})
Theoretical Value	42.56	0.320	20.41	4.80	32.97	0.625	15.39	6.25
True Value	42.65	0.358	20.52	4.50	33.68	0.537	15.80	5.48
A	40 ± 10.	0.38 $^{+.05}_{-.03}$			40 ± 7.	0.50 $^{+.08}_{-.05}$		
B	45.0 ± 4.3 (45.7)	0.36 ± 0.05 (0.359)	19.6 ± 1.2 (19.0)	2.9 ± 1.0 (4.46)	32.3 ± 3.0 (34.5)	0.55 ± 0.08 (0.539)	14.0 ± 0.8 (14.6)	4.4 ± 1.4 (6.31)
C	44. ± 3.	0.43 ± 0.06		4.2 ± 0.6	33 ± 2.	0.60 ± 0.09		4.6 ± 0.7
D	31.8	0.37			29.6	0.53		
E								
F	38.1 ± 3.	0.42 ± 0.07			39.0 ± 3.0	0.53 ± 0.09		
H	42. ± 3.	0.37		8.1	40. ± 3.	0.53		11.

TABLE 5

Results for Case C in the Second Exercise

(Simulation of Experimental Data for I = 1/2, A = 100, as in Case 1 of the First Exercise)

	CASE C1				CASE C2			
	D_o (eV)	S_o ($\cdot 10^{-4}$)	D_1 (eV)	S_1 ($\cdot 10^{-4}$)	D_o (eV)	S_o ($\cdot 10^{-4}$)	D_1 (eV)	S_1 ($\cdot 10^{-4}$)
Theoretical Value	34.05	0.400	16.32	6.00	41.21	0.500	19.24	5.00
True Value	34.12	0.447	16.41	5.63	42.11	0.430	19.75	4.39
A	34. $\pm$ 5	0.47 $^{+.06}_{-.04}$			44. $\pm$ 6.	0.43 $^{+.06}_{-.05}$		
B	38.3 $\pm$ 3.7	0.450 $\pm$ 0.009	16.2 $\pm$ 1.0	5.98 $\pm$ 0.56	42.6 $\pm$ 3.9	0.442 $\pm$ 0.005	19.6 $\pm$ 1.2	4.28 $\pm$ 0.34
C	42. $\pm$ 3. (38. $\pm$ 3.)	0.47 $\pm$ 0.08		6.6 $\pm$ 1.1	52. $\pm$ 3. (46. $\pm$ 3.)	0.45 $\pm$ 0.07		4.7 $\pm$ 0.8
D	46.1 (34.7)	0.40 (0.43)			52.1 (42.4)	0.41 (0.43)		
E	32.7	0.456			44.7	0.43		
F	44.6 $\pm$ 3.5	0.47 $\pm$ 0.08			42.1 $\pm$ 3.5	0.46 $\pm$ 0.08		
H	46. $\pm$ 4.	0.42		12.	44. $\pm$ 3	0.42		6.7

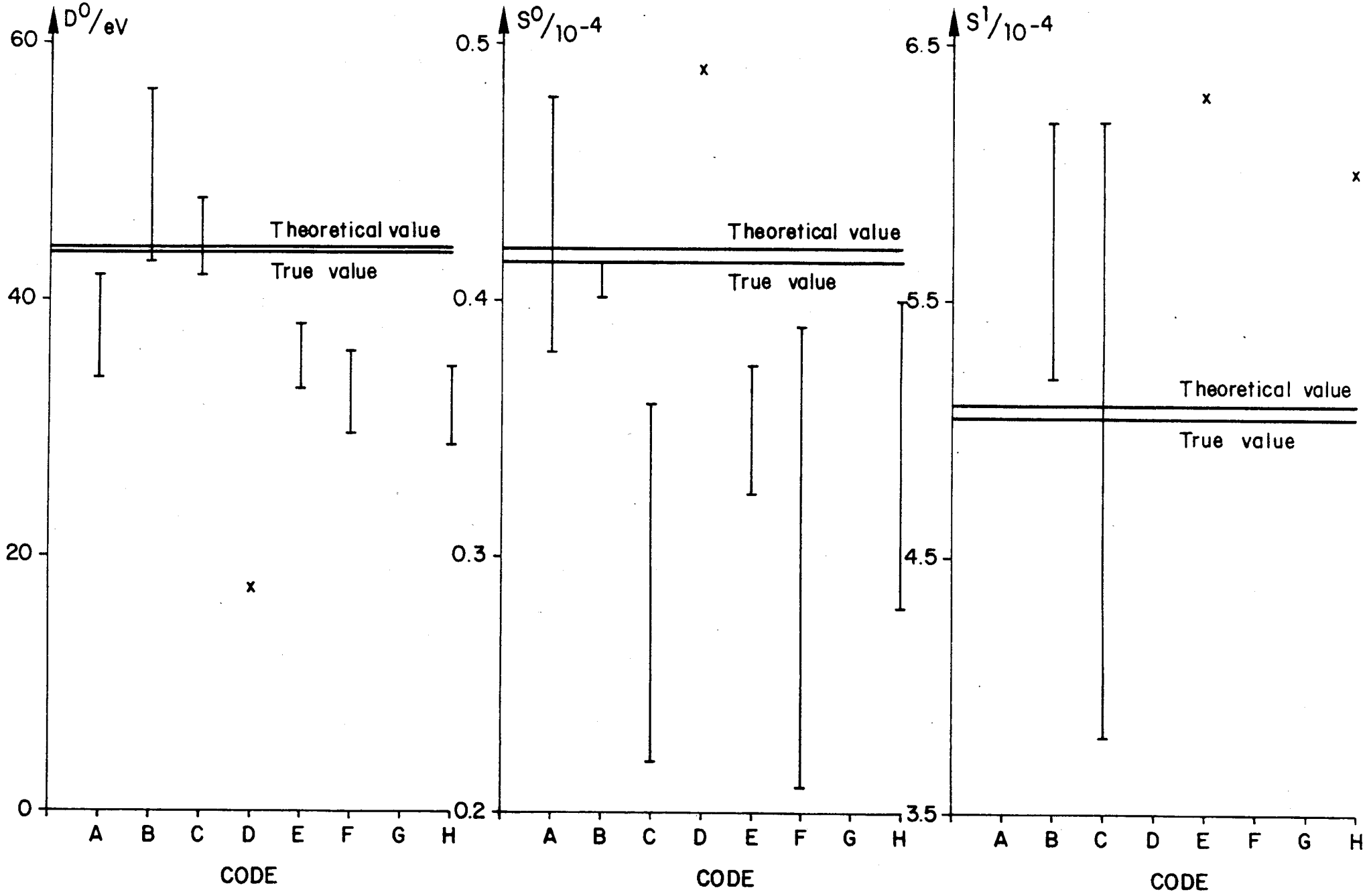

FIGURE 1 (see Table 1). Results for Case 1A

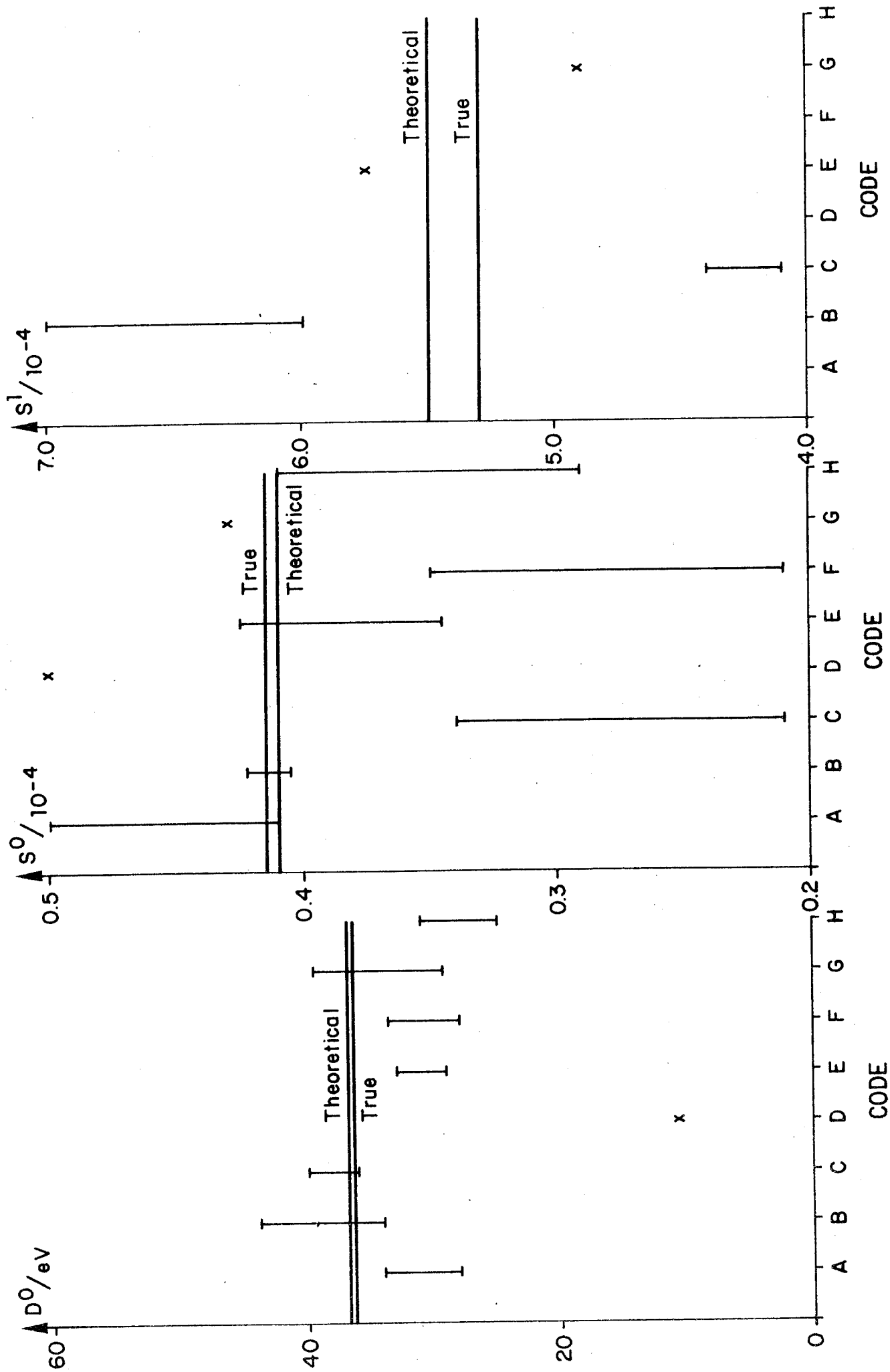

FIGURE 2 (see Table 1). Results for Case 1B

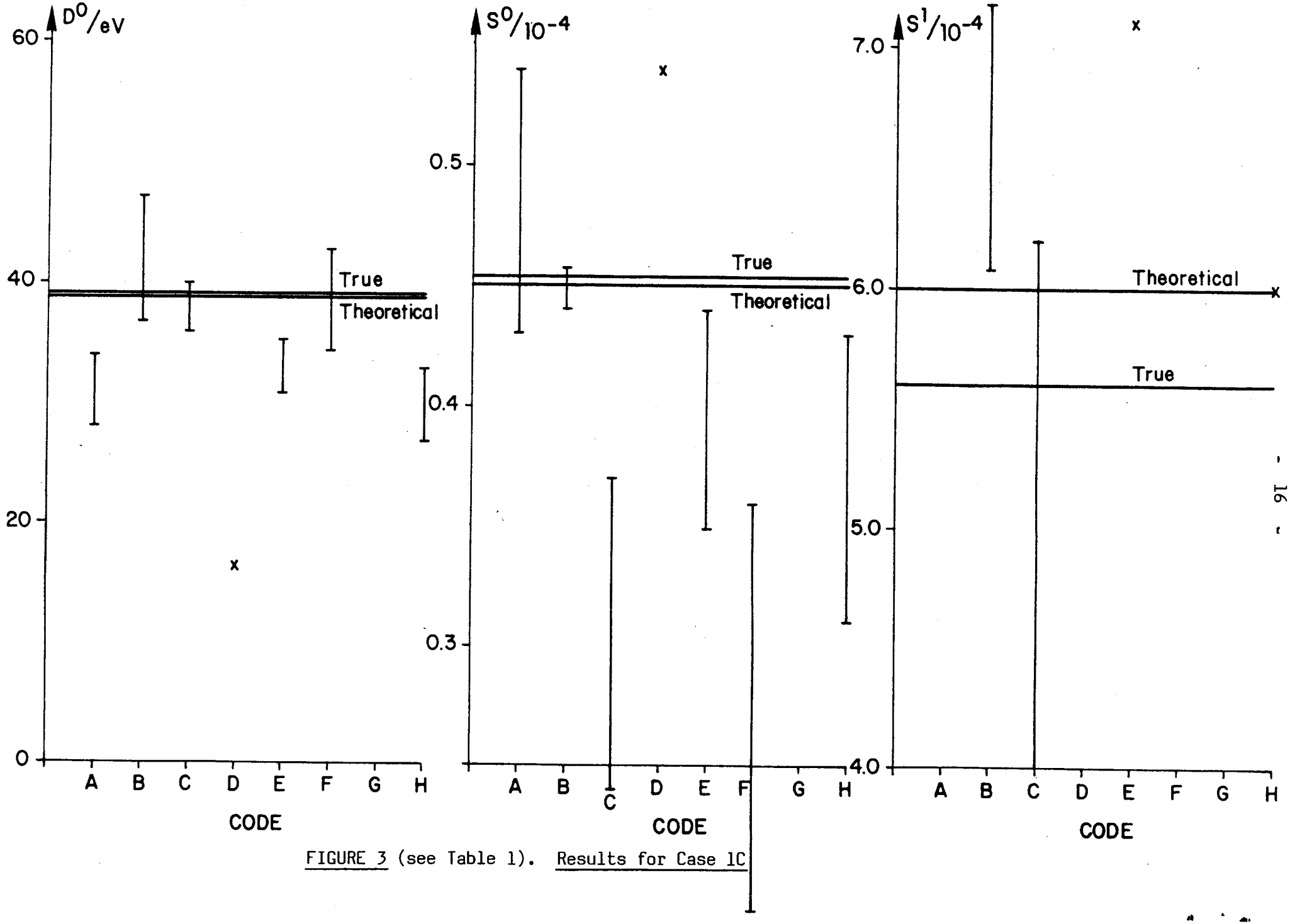

FIGURE 3 (see Table 1). Results for Case 1C

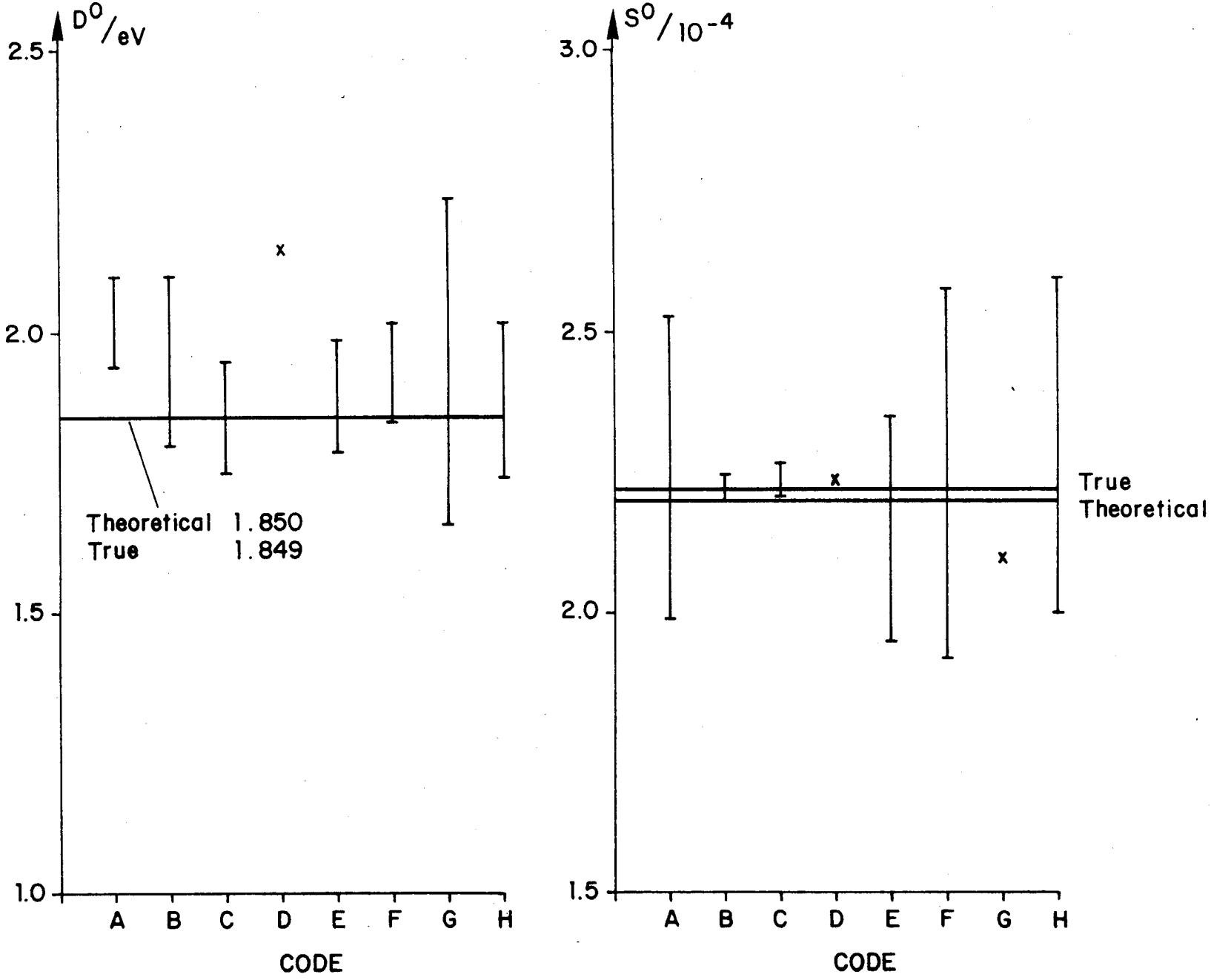

FIGURE 4 (see Table 2). Results for Case 5A

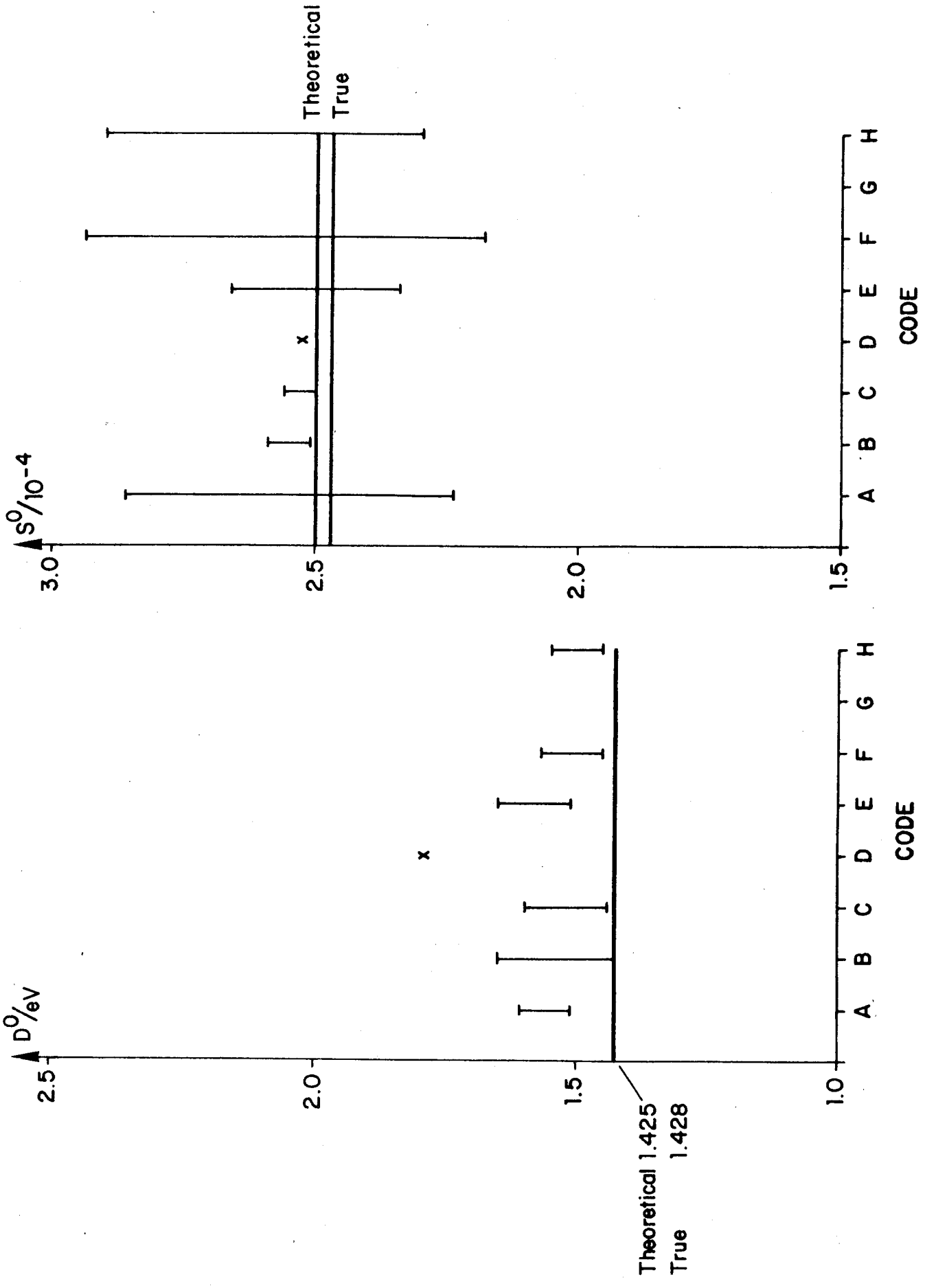

FIGURE 5 (see Table 2). Results for Case 5B

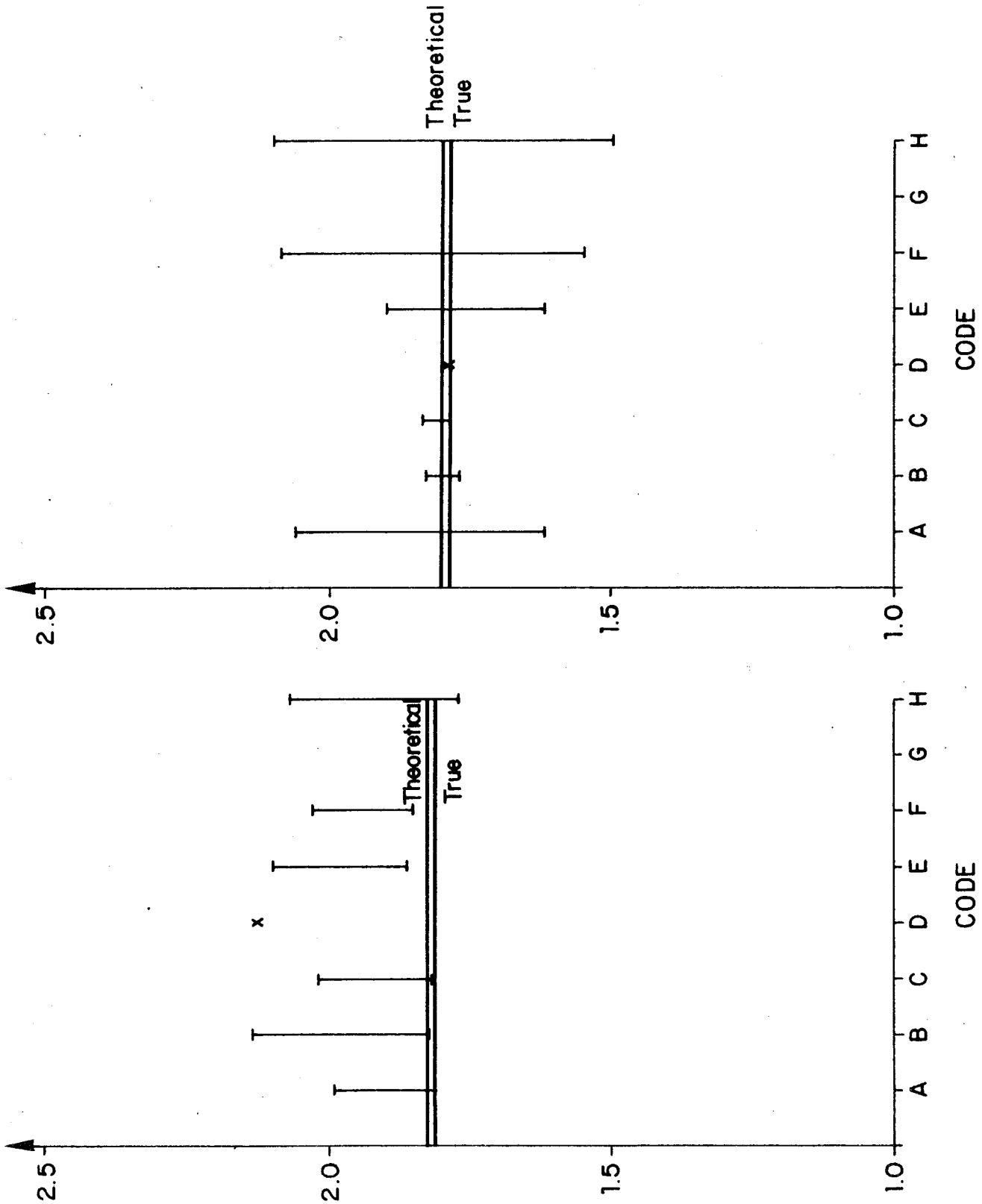

FIGURE 6 (see Table 2). Results for Case 5C

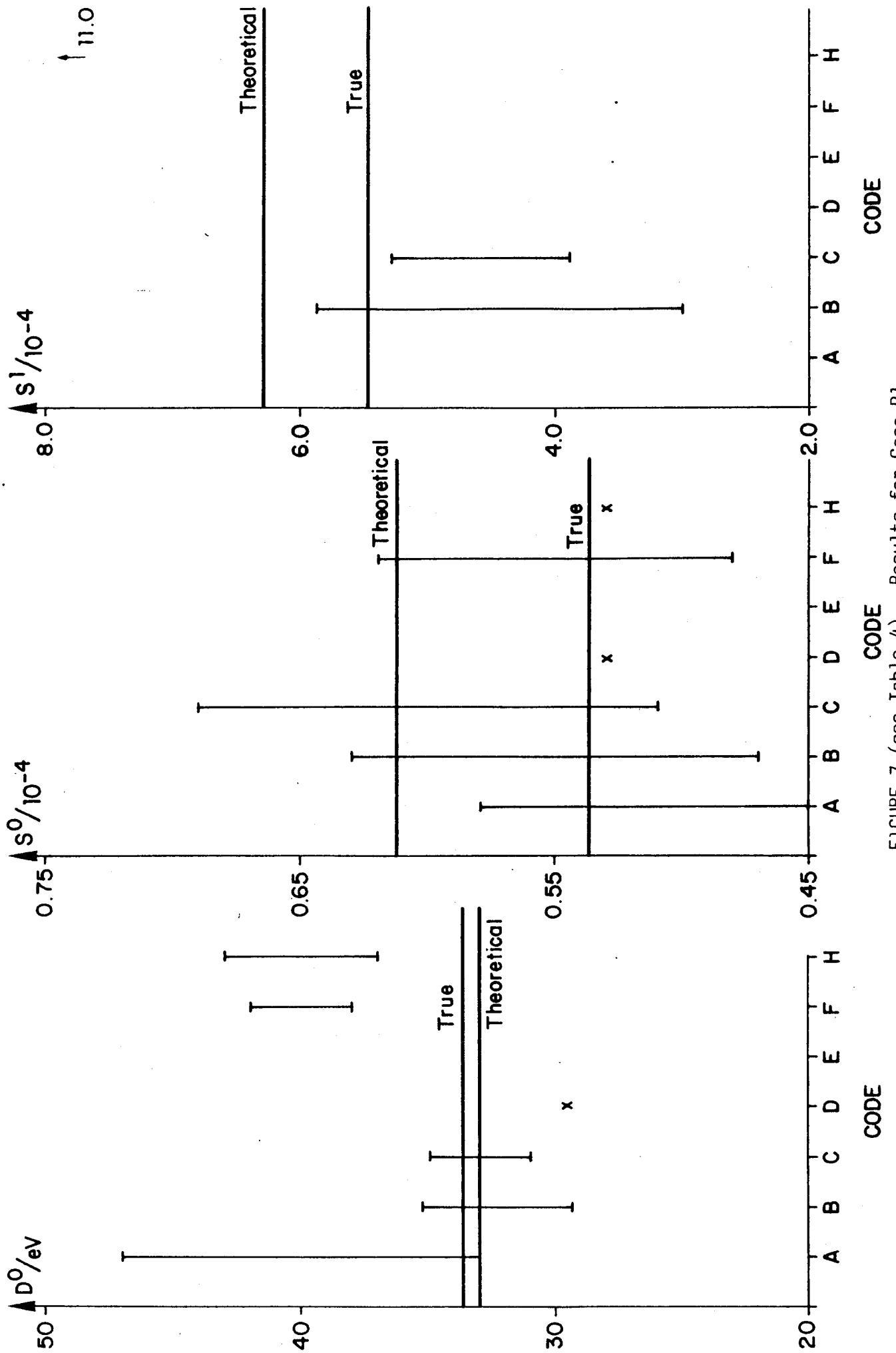

FIGURE 7 (see Table 4). Results for Case B1

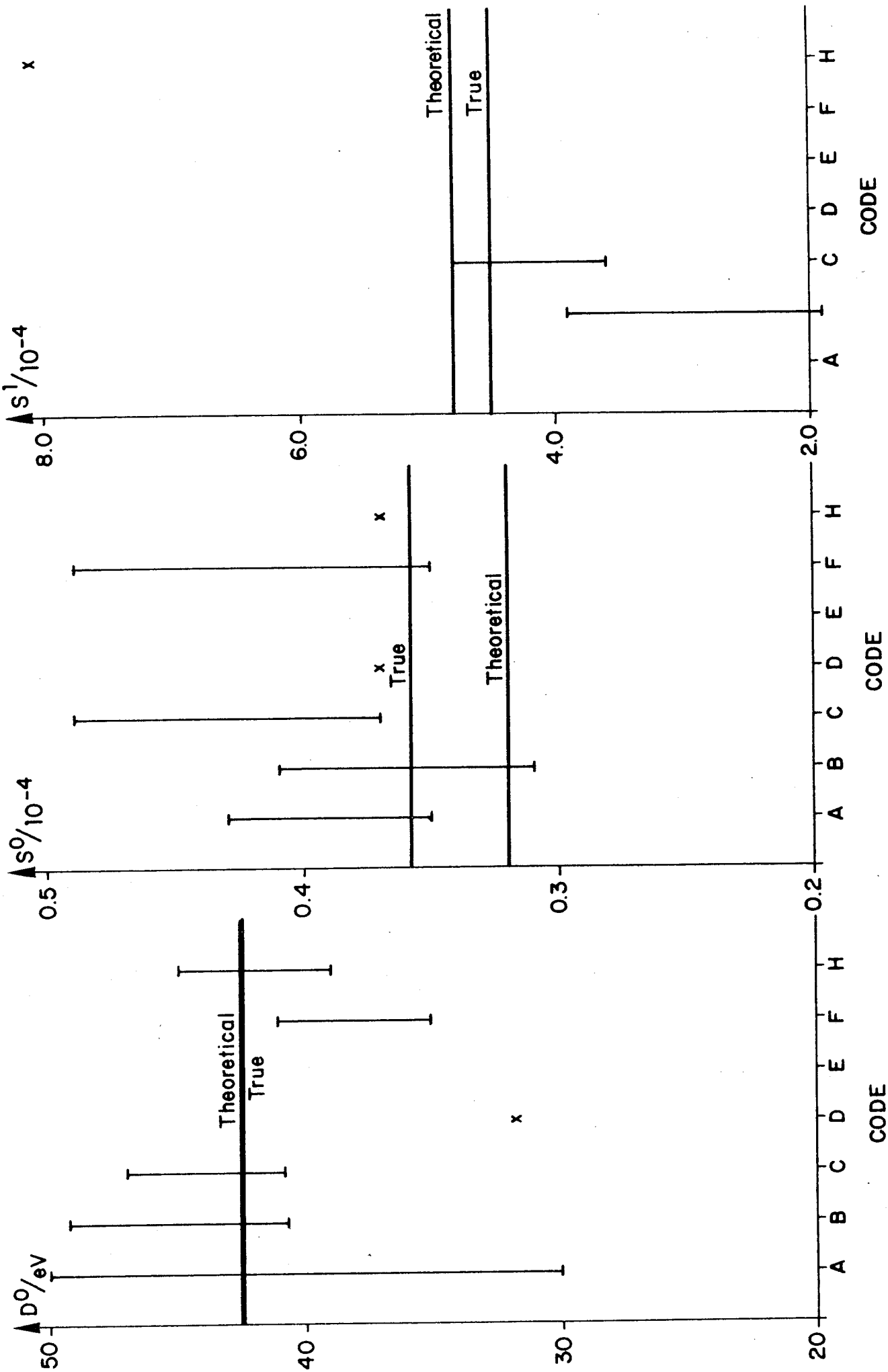

FIGURE 8 (see Table 4). Results for Case B2

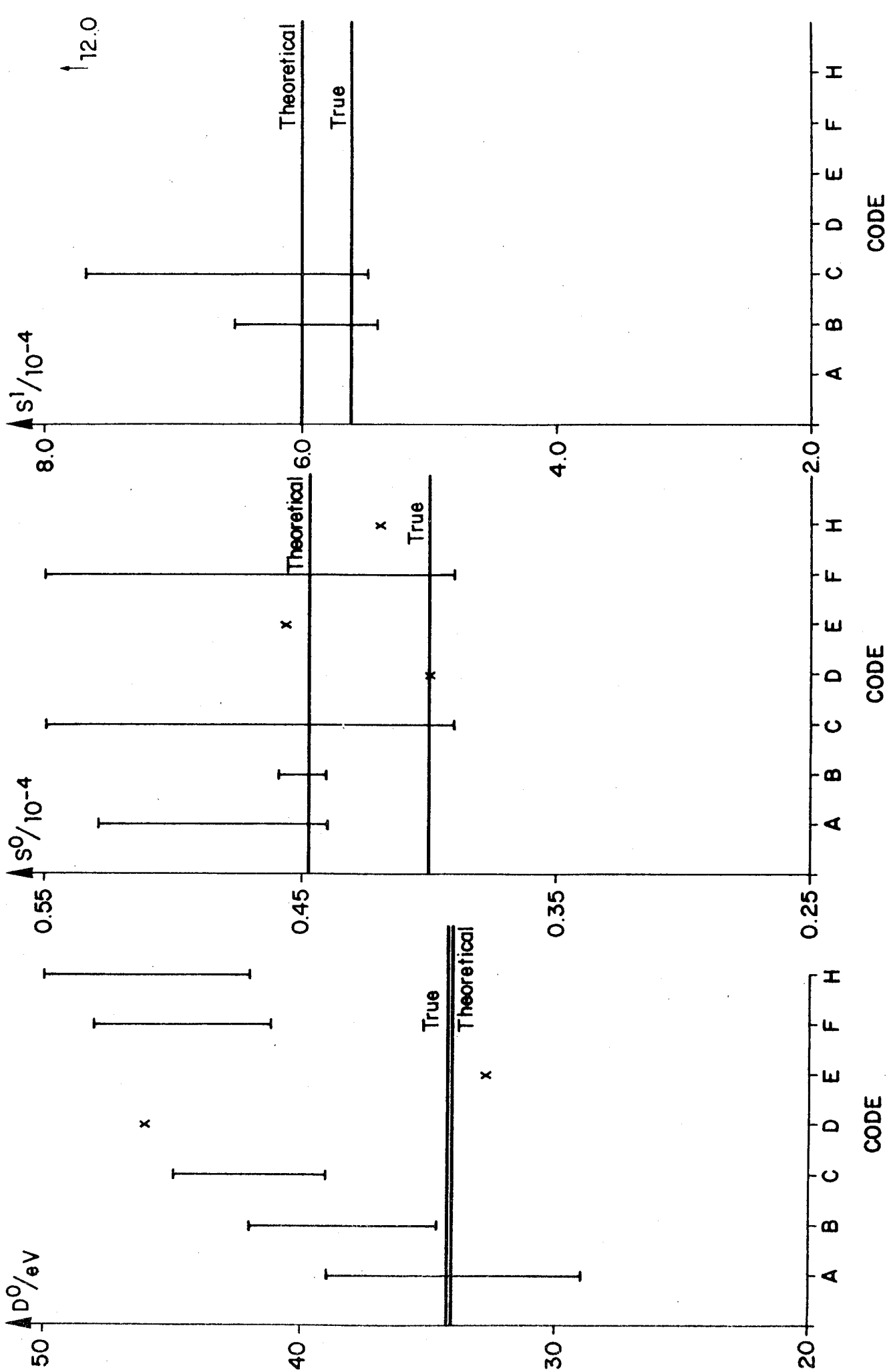

FIGURE 9 (see Table 5). Results for Case C1

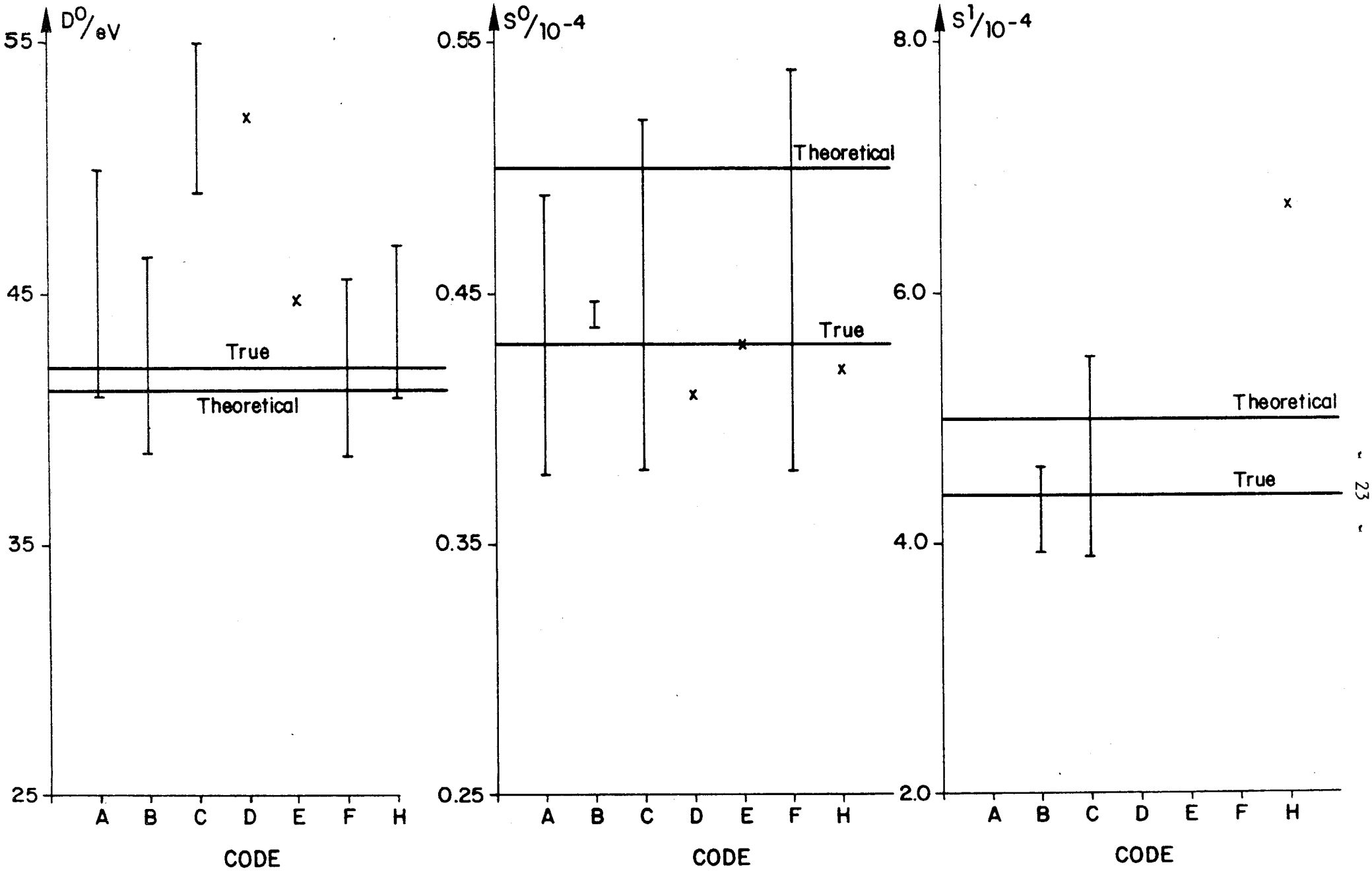

FIGURE 10 (see Table 5). Results for Case C2

THE THIRD STAGE OF THE AVERAGE RESONANCE PARAMETER BENCHMARK: A 'BLIND' TEST

A. Thompson, NEA Data Bank, Saclay, France

1. INTRODUCTION

This report describes the third stage of a benchmark test organised by P. Ribon to compare various methods used to determine the average parameters from sets of resonance parameters derived from experimental data.

Nine physicists participated in the first stage of the exercise, analyzing sets of generated pseudo-experimental resonance parameters. The results, circulated in March 1981 (NEANDC(E)231-AL), demonstrated that the average level spacings calculated by various methods contained systematic errors.

Further sets of pseudo-experimental parameters were generated, with the aim of determining the origin of these systematic errors. Seven of the nine original physicists participated in this second stage. The results were discussed at a working group meeting at the NEA Data Bank in October 1981.

It was decided to undertake a third stage. In order to determine whether the subjective judgements inherent in many of the methods required the level of knowledge about a program possessed by the author, some of the second stage tests were rerun by a physicist unfamiliar with the problem.

The third stage was performed by A. Thompson at the NEA Data Bank during March and April 1982 and involved the following four programs: H. Gruppelaar's CAVECN, (a modification of M. Stephanon's CAVE), E. Fort's ESTIMA, M. Moore's BAYESZ, and F. Fröhner's STARA.

2. THE PROBLEM

All four programs were written to calculate average parameters from a set of resonance energies Eo and neutron widths $g\Gamma_n$ together with radiation widths Γ_γ or total widths Γ_{tot}. These widths characterise the decay of a compound nucleus of spin J via various "channels". In the case of neutron widths, a "channel" consists of neutron plus nucleus and is specified by channel spin s (the vector sum of neutron and nuclear spin) and orbital angular momentum. l.

A set of widths associated with the decay of a compound nucleus via a single channel is expected to obey the Porter-Thomas distribution:

$$P(x)dx = e^{-1/2x} / \sqrt{2\pi x}\ dx \qquad (1)$$

where $$x = \frac{g\Gamma_n^o}{\langle g\Gamma_n^o \rangle} \quad \text{for s-wave neutron widths} \qquad (2)$$

$$g\Gamma_n^o = g\Gamma_n \sqrt{\frac{1eV}{Eo}} = \text{reduced s-wave neutron width} \qquad (3)$$

$$g = \text{statistical weight factor} \qquad (4)$$

$$= \frac{2J+1}{2(2I+1)} \quad \text{for a neutron induced reaction}$$

J = compound nucleus spin

I = target nucleus spin

$$\langle g\Gamma_n^o \rangle = \text{mean reduced s-wave neutron width} \qquad (5)$$

and $$S^{l=0} = \frac{\langle g\Gamma_n^o \rangle}{D^{l=0}} = \text{s-wave strength function,} \qquad (6)$$

where $D^{l=0}$ is the mean s-wave level spacing

An experimental set of widths will not be a pure s-wave. In the present benchmark calculations target nucleus spin I = 1/2, and the neutron widths will consists of two s-wave (l=0) and four p-wave (l=1) series characterized by the following spin quantum numbers:

	Neutron Spin	I	s	l	J
i)	1/2	1/2	1	0	1
ii)	1/2	1/2	0	0	0
iii)	1/2	1/2	1	1	2
iv)	1/2	1/2	1	1	1
v)	1/2	1/2	1	1	0
vi)	1/2	1/2	0	1	1

The inclusion of the spin factor g in (2) allows the two series of s-wave widths i) and ii) to be treated as a single Porter-Thomas distribution. The same is true of the two sets of p-wave widths iii) and v), except that the variable x appearing in the distribution must be redefined.

$$x = g\Gamma_n^1 / \langle g\Gamma_n^1 \rangle \qquad (7)$$

where $$\Gamma_n^1 = \Gamma_n \sqrt{\frac{1eV}{Eo}} \left(1 + \frac{1}{k^2R^2}\right) \qquad (8)$$

Γ_n^1 = reduced p-wave neutron width

k = neutron wavenumber

R = nuclear radius

$$S^{l=1} = \frac{\langle g\Gamma_n^1 \rangle}{3D^{l=1}} = \text{p-wave strength function} \qquad (9)$$

$D^{l=1}$ = mean p-wave level spacing

There are two p-wave decay channels for a J=1 resonance ((iv) and (vi) above) and the corresponding widths obey not a Porter-Thomas, but an exponential distribution

$$P(x)\,dx = e^{-x}\,dx \qquad (10)$$

where x is given by (7)

To calculate the average parameters from a set of widths distributed according to a pure s-wave Porter-Thomas distribution would not be difficult, however the real problem is complicated in three important ways.

i) The presence of p-wave resonances.

An experimental set of resonance parameters will consist of a superposition of three different distibutions, which cannot be completely separated since it is usually not possible to assign a definite l value to each individual resonance. However use can be made of an expression, derived from Bayes' Theorem, for the probability that a resonance of width $g\Gamma_n$ is p-wave.

ii) There will be an energy-dependant minimum detectable width. The missing levels must be considered.

iii) There will be a minimum resolution, two resonances closely spaced in energy may appear as a single resonance with over-large width.

Phenomena ii) and iii) are interdependent. For example, the minimum detectable width at any given energy is increased by the presence of a resonance.

3. THE ALGORITHMS

Program STARA performs a maximum likelihood fit of the neutron width to a composite distribution consisting of a weighted sum of two Porter-Thomas distributions, (for s-wave and p-wave) and a p-wave exponential. An energy-dependent, diffuse, minimum detection threshold is also included in the likelihood function, which is maximised by a Newton iteration technique. Two parameters are adjusted during the process:

i) The mean s-wave neutron width $\langle g\Gamma_n^0 \rangle$

ii) The fraction of observed levels.

When convergence is achieved, the s-wave Porter-Thomas distribution is replaced by one distorted to account for resonance overlap, and a new converged solution is sought.

The p-wave strength function $S^{l=1}$ is given as an input, and its value remains constant throughout the calculation. The user has the option to analyze only resonances within a given energy range, or those with reduced widths greater than a given threshold THRESH. The effect of increasing this threshold is to reduce the proportion of p-wave resonances in the analyzed data.

Program BAYESZ uses a rather different method, the missing level estimator method based on the following properties of the Porter-Thomas distribution (1)

$$\int_{1/4}^{\infty} P(x)dx = 0.617 \qquad (12)$$

$$\int_{1/4}^{\infty} \sqrt{g\Gamma_n^o}\, P(x)dx = 0.70\sqrt{\langle g\Gamma_n^o\rangle} \qquad (13)$$

$$\int_{1/4}^{\infty} g\Gamma_n^o\, P(x)dx = 0.969\,\langle g\Gamma_n^o\rangle \qquad (14)$$

These integrals can be approximated by weighted sums over the input resonance data, if these summations are computed in the order of decreasing reduced neutron width and stopped when

$$\frac{\left(\sum_i W_i\right)\left(\sum_i W_i\, g_{ni}^o\right)}{\left(\sum_i W_i\, g\Gamma_{ni}^o\right)^2} = \frac{0.627 \times 0.969}{(0.70)^2} = 1.2062 \qquad (15)$$

it follows that the mean s-wave neutron width g_n^o can be estimated from

$$\frac{\sum_i W_i\, g\Gamma_{ni}^o}{\sum_i W_i} = \frac{0.969}{0.617}\langle g\Gamma_n^o\rangle = 1.715\langle g\Gamma_n^o\rangle \qquad (16)$$

The weights W_i include contributions from:

i) the probability that an individual resonance is p-wave, calculated from Bayes' Theorem.

ii) the probability that a particular resonance is an unresolved doublet.

The cut-off at $x = 1/4$, or $g\Gamma_n^o = 1/4\ g\Gamma_n^o$ avoids the need to take special account of a minimum detection threshold.

The iterative procedure used by BAYESZ involves calculating the weights W_i from initial guesses for the average parameters $D^{I=0}$, $S^{I=0}$, $S^{I=1}$ computing the summations appearing in (15) for a range of cut-offs, and selecting improved $D^{I=0}$, $S^{I=0}$, $S^{I=1}$ values. The user must also supply a reduced width threshold above which all resonances are assumed to be s-wave. This parameter is automatically adjusted to achieve the expected s to p wave level spacing.

The philosophy behind the ESTIMA program is slightly different. It contains three different algorithms, all of which require a pure s-wave sample of resonances, and none of which account for resonance overlap. Before beginning the calculations, ESTIMA rejects those resonances with too high a probability of being p-wave. The probability is calculated from Bayes' Theorem using input guesses of $\langle g\Gamma_n^0 \rangle$ and $\langle g\Gamma_n^1 \rangle$, and the resonance is rejected if this value is greater than another input parameters CONSTE. Resonances parameters which are known to be p-wave should be excluded from the input data.

The three algorithms are:

i) A maximum likelihood fit to a truncated Porter-Thomas distribution, using a Newton-Raphson technique and adjusting $\langle g\Gamma_n^0 \rangle$. The fit is performed for a maximum of twenty truncation thresholds, and can also be restricted to a given energy range.

   The quantity $\langle g\Gamma_n^1 \rangle$ must be given as an input variable, and remains fixed throughout the calculation.

ii) A simplified missing level estimator method, similar to that of BAYESZ, but with all the weights W_i set to unity. This simplification removes the need for iterations.

iii) The GAMN method, based on the following properties of the Porter-Thomas distribution (1):

$$\int_1^\infty P(x)dx = 0.32 \; ; \quad \int_{0.5}^1 P(x)dx = 0.17 \qquad (17)$$

$$\int_{0.2}^{0.5} P(x)dx = 0.18 \; ; \quad \int_{0.05}^{0.2} P(x)dx = 0.17 \qquad (18)$$

Using an energy-dependent threshold to remove all p-wave (and inevitably also some s-wave) resonances, $\langle g\Gamma_n^0 \rangle$ can be manually adjusted until the number of widths lying in the above range is proportional to the integrals.

CAVECN is another program which uses a mximum likelihood technique to fit a set of resonance widths to a truncated Porter-Thomas distribution. Since, in the event, it was not extensively tested, the algorithm need not be described here.

The following table summarizes the required input parameters of the three codes STARA, BAYESZ and ESTIMA.

	STARA	BAYESZ	ESTIMA
Required Resonance parameters and their errors	E_o Γ_{tot} E_o $\delta\Gamma_{tot}$ $2g\Gamma_n$ $\delta(2g\Gamma_n)$	E_o Γ_{tot} $\delta\Gamma_{tot}$ $2g\Gamma_n$ $\delta(2g\Gamma_n)$	E_o Γ_{tot} $2g\Gamma_n$, Γ_γ
Parameters describing the nucleus	A,I,R,	A,I,σ,R	A,I, σ , R
Parameters to be manually adjusted	$S^{l=1}$, THRESH	none	$\langle g\Gamma_n^1 \rangle$ CONSTE thresholds for max. likelihood method
Parameters automatically adjusted by the program	$\langle g\Gamma_n^o \rangle$	$D^{l=0}$, $S^{l=0}$, $S^{l=1}$ threshold above which resonances are assumed s-wave	$\langle g\Gamma_n^o \rangle$

All the above parameters have been already described except A = mass number and the spin cut-off parameter σ, used to calculate the s to p level spacing.

4. TESTING THE PROGRAMS - A USERS VIEWPOINT

The programs were tested in the order CAVECN, ESTIMA, BAYESZ, STARA for the resonance parameters sets, B1, C1, C2 used in the second stage of this benchmark exercise. CPU time and storage requirements are summarized in the accompanying table:

TIME AND CORE STORAGE REQUIREMENTS

	CAVECN	ESTIMA	BAYESZ	STARA
Computer	CRAY-1	IBM 3033-11	CRAY-1	IBM 3033-11
Storage requirements	37450_8 words	270K bytes including graphics routines 190K bytes otherwise	32400_8 words	300K bytes excluding graphics routines
CPU Time 275 resonance parameters analysed	30 secs. for 1.5 iterations	1.2 secs. for one execution of the three algorithms.	2.25 secs for 11 iterations.	0.55 secs. for 2 analyses.
Number of runs required per resonance parameter set		4 - 8	2	About 25

i) CAVECN

The tester was supplied with the FORTRAN source, test problem data (but no example output), and documentation consisting of a list of input variables and formats. Modifications to the input format statements had to be made before the test data could be read.

Several problems arose due to the fact that the tester was, at that time, unfamiliar with both the CRAY-1 operating system and the file editting/job submission system being used. These problems were soon overcome, but it was quickly discovered that CAVECN required in excess of 30 seconds of CRAY CP time to complete two iterations of the test problem. Since the author requested that the code should not be made available for distrbution it was decided not to continue testing.

ii) ESTIMA

ESTIMA was supplied, not only with FORTRAN source, test problem data together with the corresponding lineprinter and graphical output, draft documentation describing the algorithms and input parameters, but also with a full day visit from the author, E. Fort. This proved to be helpful, clarifying various aspects of the physics and statistics of resonance parameters.

Because the code was run on the same machine that it was developed on, no significant problems arose implementing it, and, as a bonus, the graphical output could be obtained without difficulty. A fair degree of judgement is required to choose the most physical result. The graphics was found to be helpful, although lineprinter plots were usually adequate.

The most sophisticated of the three ESTIMA algorithms is the maximum likelihood technique, and all results quoted in the next chapter are calculated by this method. A good result requires a truncation threshold low enough to include as many resonances as possible, but high enough to exclude all p-waves. An optimum is chosen by searching for a region where the calculated average parameters are insensitive to changes in threshold.

In practice, the missing level estimator method was used to calculate a range of values for $\langle g\Gamma^0_n \rangle$ and then consistency between all three methods was sought, varying the s_n and p-wave mean neutron widths and keeping CONSTE fixed. This was achieved in about 4 to 8 runs, taking about 1/2 hour between each to analyze results and decide the input parameters for the next run.

iii) BAYESZ

The source of BAYESZ was developed on a CDC machine, and minor modifications were required to implement it on the CRAY.

Since all parameters are automatically adjusted by the program, there is very little that the user need do, except submit the job. Two runs were performed for each case, one using order of magnitude guesses for the quantities $D^{l=0}$, $S^{l=0}$, $S^{l=1}$, and a second confirmatory run using the results of the first as input.

It is very difficult to adjust the results of BAYESZ, for example case C1 was repeated for a range of threshold parameters spread over 4 decades. Although the number of iterations varied, the final results agreed to within 3 figures of precision.

iv) STARA

STARA was supplied as FORTRAN source, together with documentation, test problem input file and sample outputs. The program includes graphical output options, but these could not be easily implemented at Saclay. The test problem was found to include an execution step error, due to a redundant FORTRAN statement left over from an earlier version. The offending statement was deleted.

The program requires that $S^{l=1}$ and the threshold parameter be adjusted manually. The problem is very similar to the ESTIMA maximum likelihood technique - a value of $S^{l=1}$ is sought where the calculated s-wave parameters are more or less independent of threshold. STARA is, however, not as convenient to use in this way as ESTIMA, the former requires one run for each pair of values for $S^{l=1}$ and threshold, the latter performs the fitting for up to 20 threshold values at once. This parameter search technique was not used by the author in the first and second stages of the benchmark test.

It was also found that in certain cases, at low $S^{l=1}$ values and low, but non zero thresholds, e.g. 0.1 or 0.2 meV, the program failed due to an attempt to square root a negative number.

5. RESULTS

The first two stages of this benchmark comparison tested the accuracy of various numerical algorithms designed to compute average resonance parameters from experimental data. This third stage tested the programs as a whole; from the point of view of a user who is not an author, the accompanying documentation is as much part of the computer program as the coding of the algorithm.

There are two things to report in this chapter: the usefulness of the supplied documentation and a comparison of the numerical results with the values computed by the code authors, and those used to generate the pseudo-experimental resonance parameters.

Documentation

Comments on the adequacy of documentation are inevitably somewhat subjective. It is, therefore, useful to state what constitutes a well-documented program. It includes:

i) The coding itself

ii) Technical details needed to run the code, e.g. operating system utilities and subroutine libraries used, storage space requirements etc. These details are best summarized by a set of job control commands.

iii) A brief description of the problem to be solved and the algorithm used, including any physics or mathematics approximations and the implied limitation on the accuracy of the numerical results.

iv) A full description of the required input parameters.

v) A sample input dataset and corresponding output.

vi) A description of how the output should be analysed.

vii) Any other comments on the behaviour of the program for different types of problem.

In the final analysis, documentation cannot include everything. The only way to learn how to use a program is to run it. Good documentation should enable the user to get the best out of the code after a reasonably short learning time.

Judged on these criteria, the documentation supplied with both BAYESZ and ESTIMA was good. The former, since it requires no analysis of the output, is very easy to use. The draft description supplied with the latter contained many helpful comments; these were supplemented by a day visit from the author.

The STARA documentation contained all the above items except vi) and vii), that is it gave no information on how to choose the input parameters to obtain self-consistent results. By way of comparison the present tester learnt how to use ESTIMA after 4 or 5 separate runs; it required in excess of 25 to create a technique for using STARA.

Since the testing was stopped at an early stage, it is difficult to make comments about the CAVECN documentation.

Numerical Results

The values obtained for the s-wave level spacing $D^{l=0}$, s- and p- wave strength functions $S^{l=0}$ and $S^{l=1}$ and radiation width $\langle\Gamma_\gamma\rangle$ are summarized in Tables 1 and 3. They are compared with:

i) the "theoretical values" introduced into the computation of the synthesized parameter sets.

ii) the "true values", i.e. the average value of each set of randomly generated parameters.

iii) the values obtained by the authors using their own codes.

The difference between i) and ii) is a measure of the statistical fluctuations in the sample.

The following comments can be made:

i) General

All three cases contained about 275 resonance parameters, of which about 120 were s-wave. It would be difficult to construct a better case which realistically simulates experimental data.

Table 2 summarizes the standard deviations of the author and Thompson results, computed as the root mean square deviation from the "true value" for the three cases B1, C1 and C2. These values are compared with the fractional difference between "true" and "theoretical value".

The dispersion, due to the analysis, is less than the statistical dispersion for $S^{l=0}$: around 2% in the case of ESTIMA and BAYESZ compared to a 12% difference between "true" and "theoretical" values. The analytic dispersion for the s-wave level spacing $D^{l=0}$ is, at best, about 6%, somewhat greater values are obtained for $S^{l=1}$.

ii) s-wave strength function $S^{l=0}$

There is good agreement in all cases between the true result, author's result and the Thompson result.

iii) s-wave level spacing $D^{l=0}$

The agreement between results is not so good. The fractional error of the Thompson and author results, compared to the true result, have been calculated, and a covariance matrix formed for each program.

$$\text{covariance matrix} \quad \sigma = \begin{bmatrix} \langle x_T x_T \rangle & \langle x_T x_A \rangle \\ \langle x_T x_A \rangle & \langle x_A x_A \rangle \end{bmatrix} \tag{19}$$

$$\text{where } X_T = (D^{l=0}_{\text{Thompson}} - D^{l=0}_{\text{True}}) / D^{l=0}_{\text{True}} \tag{20}$$

$$X_A = (D^{l=0}_{\text{Author}} - D^{l=0}_{\text{True}}) / D^{l=0}_{\text{True}} \tag{21}$$

and the averaging, denoted by $\langle\ \rangle$ is performed over the three cases B1, C1, C2 for BAYESZ and STARA, and over the two cases C1, C2 for ESTIMA.

The diagonal terms are measures of the error in the results from a particular code and particular user and are the squares of the standard deviations given in Table 2. The errors obtained by Thompson and the authors are of comparable size for all three codes, however the errors of STARA results are significantly larger than those of the other two codes.

The off-diagonal terms are measures of the correlation between the two sets of calculated results. They are best expressed by the correlation coefficient:

$$r = \frac{\langle x_T x_A \rangle}{\sqrt{\langle x_T x_T \rangle \langle x_A x_A \rangle}} \tag{22}$$

Since the errors X are expected to be primarily due to systematic errors in the codes which should be independent of user, the Thompson and author results are expected to be highly correlated, i.e. close to unity. This result is found for BAYESZ and ESTIMA. The correlation for STARA is, however, negative. This is partly due to the fact that the code was used in rather different ways by the tester and the author.

It is, however, difficult to make firm conclusions on the basis of covariances derived from such a small statistical sample.

iv) p-wave strength function $S^{l=1}$

Of the three programs for which results were obtained, only BAYESZ calculates a value for the p-wave strength function $S^{l=1}$ by an automatic iteration technique. The results obtained for cases C1 and C2 with this code are markedly better than those for case B1, probably because the data for this case contained no information on the l value of the resonances.

ESTIMA and STARA both estimate the effect of the p-wave resonances in the data on the basis of an input value either for the p-wave mean reduced width, or the p-wave strength function. The $S^{l=1}$ values quoted in Table 1 are optimum input parameters rather than calculated results, and for this reason, their errors have not be assigned.

6. CONCLUSIONS

The main conclusions of this third stage of the benchmark test can be summarized:

i) The two programs BAYESZ and ESTIMA can be used equally well by an unfamiliar user as by their authors, to obtain average resonance parameters from a simulation of experimental results with a standard deviation of about 7.5% on $D^{l=0}$ and about 2% on $S^{l=0}$

ii) The corresponding errors using STARA are 11-17% for $D^{l=0}$ and about 5% for $S^{l=0}$. A significant amount of judgement is required to obtain recommended values from the output and it is not clear which is the best way to apply it.

7. POSTSCRIPT - STATUS OF PROGRAMS

The codes ESTIMA and BAYESZ are available from the NEA Data Bank. The authors of both CAVECN and STARA have requested that their programs should not be distributed.

TABLE 1

RESULTS FOR $D^{l=0}$, $S^{l=0}$, $S^{l=1}$ FOR THE BENCHMARK RESONANCE PARAMETER SETS B1,C1,C2 USING THE CODES ESTIMA, BAYESZ AND STARA

	CASE B1			CASE C1			CASE C2		
	$D^{l=0}$/eV	$S^{l=0}/10^{-4}$	$S^{l=1}/10^{-4}$	$D^{l=0}$/eV	$S^{l=0}/10^{-4}$	$S^{l=1}/10^{-4}$	$D^{l=0}$/eV	$S^{l=0}/10^{-4}$	$S^{l=1}/10^{-4}$
TRUE VALUES	33.7	0.54	5.5	34.1	0.45	5.6	42.1	0.43	4.4
THEORETICAL VALUES	33.0	0.63	6.3	34.1	0.40	6.0	41.2	0.50	5.0
ESTIMA									
E. Fort				32.7	0.46		44.7	0.43	
A. Thompson	29.4 +/-4.3	0.56 +/-0.08	4.9	32.5 +/-4.6	0.46 +/-0.07	3.5	44.1 +/-6.7	0.43 +/-0.07	3.1
BAYESZ									
M. Moore (BAYESX)	32.3 +/-3.0	0.55 +/-0.08	4.4 +/-1.4	38.3 +/-3.7	0.45 +/-0.01	6.0 +/-0.6	42.6 +/-3.9	0.44 +/-0.01	4.3 +/-0.3
A. Thompson	31.8 +/-2.9	0.55 +/-0.08	3.4 +/-0.5	37.6 +/-3.7	0.45 +/-0.01	5.34 +/-0.2	43.1 +/-4.1	0.436 +/-0.005	4.05 +/-0.13
STARA									
F. Fröhner	40.0 +/-7.00	0.50 +/-0.07		34.00 +/-5.0	0.47 +/-0.05		44.0 +/-6.0	0.43 +/-0.06	
A. Thompson	29.8 +/-3.8	0.54 +/-0.06	3.5	43.3 +/-5.5	0.414 +/-0.05	7.0	39.4 +/-4.5	0.43 +/-0.05	2.5

TABLE 2

STANDARD DEVIATIONS OF THE AUTHOR AND THOMPSON RESULT FROM THE "TRUE VALUES"

		$D^{I=0}$		$S^{I=0}$		$S^{I=1}$	
Code		author	Thompson	author	Thompson	author	Thompson
ESTIMA	a)	0.052	0.047	0.016	0.016		0.28
	b)		0.083		0.025		
BAYESZ	b)	0.075	0.068	0.017	0.013	0.12	0.23
STARA	b)	0.111	0.173	0.053	0.046		0.29
Fractional average difference between "true" and "theoretical" values.		0.015		0.12		0.10	

a) Calculated from cases C1 and C2 only.
b) Calculated from all three cases.

TABLE 3

MEAN RADIATION WIDTH RESULTS (PROGRAM BAYESZ ONLY)

	CASE C1	CASE C2
True Value	0.150	0.160
Theoretical Value	0.150	0.160
M. Moore	0.159 +/-0.013	0.113 +/-0.018
A. Thompson	0.162 +/-0.013	0.188 +/-0.015

TABLE 4

COVARIANCE MATRIX FOR $D^{l=0}$ RESULTS

$$\sigma = \begin{bmatrix} \langle x_T x_T \rangle & \langle x_T x_A \rangle \\ \langle x_A x_A \rangle & \langle x_A x_A \rangle \end{bmatrix}$$

where:

X_T, X_A are the fractional error in the Thompson and author results for $D^{l=0}$ (see equation 19).

ESTIMA $\quad \sigma = \begin{pmatrix} 0.0023 & 0.0025 \\ 0.0025 & 0.0028 \end{pmatrix}$

Correlation coefficient $\quad r = \dfrac{\langle x_T x_A \rangle}{\sqrt{\langle x_T x_T \rangle \langle x_A x_A \rangle}}$

$= 0.99$

BAYESZ $\quad \sigma = \begin{pmatrix} 0.0033 & 0.0041 \\ 0.0041 & 0.0057 \end{pmatrix}$

$r = 0.95$

STARA $\quad \sigma = \begin{pmatrix} 0.030 & -0.0079 \\ -0.0079 & 0.012 \end{pmatrix}$

$r = -0.41$

COMMENTS ON THE RESULTS OF THE THIRD STAGE

Comments on the draft version of Mr. Thompson's report were received from Dr. Moore, Dr. Fort and Dr. Fröhner in letters dated 30th June, 1st July and 8th July respectively. Matters of textual detail were corrected in the final version; other comments which still apply are quoted below.

M. Moore, Los Alamos National Laboratory, USA

A. Your version of the code BAYESZ and mine gave slightly different answers when Adrian ran it. The reason seems to be that the results I reported at Paris, and which you used in Table 1, were not done with BAYESZ, but with the older version BAYESX, which ran a factor of 10 or more slower, and thus permitted only a few iterations. Running the problem with BAYESZ, I get precisely your answers for case B1, but not for C1 or C2. Here I get 37.99, 0.451, and 5.36 for D^o, S^0, and S^1 for C1, and 42.54, 0.442, and 4.19 for C2.

B. A comment on the text of the Third Stage report. At the foot of page 27 you have the sentence "the cutoff at x = 1/4... avoids the need to take special account of a minimum detection threshold". This was true of BAYESX and earlier versions, and was a weakness in the code, because if the number of resonances is too few to permit the algorithm to reach a value of 1.2, no results would be reported. You also note that in BAYESZ the results are calculated for various values of the algorithm (given in the data array xlmr in statement 43, and terminated at 1.24). But it may not be clear to the reader that this procedure overcomes the earlier weakness: the code uses however many resonances it is given, and can provide an estimate of the level density and strength function of p-wave levels if as few as 6 percent of them are observed. This feature can also lead to a problem when the sample contains only s-wave resonances, as in Cases 5A, B, C: if the reduced-width threshold is non-zero, the calculation may not converge. A caveat to the user is therefore perhaps in order: if the p-wave results are completely out of range, set the last word in the second card image preceding the data array, and/or the default value of gthr in statement 54, to zero.

F. Fröhner, KFK, Karlsruhe, Fed. Rep. of Germany

The following comments were received in a letter dated 8th July from Dr. Fröhner in reply to a letter from Mr. Thompson concerning the difficulties experienced running STARA at the NEA Data Bank.

A. General

My explanations on how STARA should be used were obviously too short. You should not have varied the p-wave strength functions in the search for the maximum likelihood estimates. Only the truncation threshold THRESH should be varied if one wants to check whether one is in the flat portion of the spacing-versus-threshold curve, exactly as in the case of ESTIMA. Variation of the p-wave strength functions leads to wild excursions and wrong results because the whole algorithm should be fairly insensitive to S_1. I think this becomes clearer if I now turn to the comments in your letter.

B. Concerning the difficulty experienced optimising the p-wave strength function input parameter

STARA was never meant to provide estimates of p-wave strength functions. These should be given as fixed input numbers, based on some reasonable guess, e.g. an optical model estimate for the mass range in question. They are only used to calculate a correction term, which is approximate for instance with respect to resolution effects. On the other hand the truncation threshold could be varied, but only to make sure that p-wave levels are effectively excluded or adequately taken into account.

C. Concerning the negative argument of a square root problem mentioned on page 32

This is one of the cases which I wanted to check by own calculations. At the moment I can say this: lowering the threshold should be effective only above the true threshold (which is estimated automatically by the code). If one prescribes too low values these should be overridden by this automatism. As a consequence, THRESH should not have any influence below a certain limiting value. Your observation that results differ for THRESH = 0 and 0.01 meV seems to point to an error in the program.

D. Concerning the insensitivity of the results to the value of $S^{l=0}$ at THRESH=0

A big dependence is not expected if the fraction of p-wave levels is not large.

E. Concerning the accuracy of calculating all radiation widths from the formula $\Gamma = \Gamma_{tot} \cdot 2g_n$

Completely correct - the calculation is dubious. For exactly this reason, I used only the lowest resonances for the radiation width estimates which I submitted as solutions to the first benchmark test. This is what evaluators usually do, the first few resonances being normally those with the best-known radiation and total widths.

E. Fort, CEN, Cadarache, France

I generally agree with the draft copy of A. Thompson's paper: at least what concerns the ESTIMA code algorithms and "philosophy" is correctly reported.

This test work appears to be of quality.

The remark I have concerns S_1 determination. I don't understand how it has been possible to obtain the values given by Mr. Thompson, 3.5 and 3.1 for cases C_1 and C_2.

In a general way S_1 is obtained (in ESTIMA) by the following formula:

$$S_1 = \frac{1}{3} \frac{\langle g\Gamma_n^1 \rangle}{D\,\alpha}$$

Since at this step of calculation α and D are known, S_1 derives directly from the knowledge of $\langle g\Gamma_n^1 \rangle$. I suspect there could be a mistake by Mr. Thompson in the determination of this quantity.

I would like to recall in the cases (C_1 and C_2) where the set of resonances contains some "a priori" p-wave resonances (declared as such by the experimentalist) the procedure for $\langle g\Gamma_n^1 \rangle$ determination is very simple:

Different guesses for $\langle g\Gamma_n^1 \rangle$ are used (starting from low values) up to the moment where all the a priori p-waves are below the curve $10 \times \langle g\Gamma_n^1 \rangle \left(1 + \frac{1}{K^2R^2}\right)$ in the plane $(E, g\Gamma_n^0)$. To be noted that this curve is dependent on R, nuclear radius. If the curve is such that the "highest" p resonance lies just below it, the corresponding parameter S^1 is an estimate of $\langle g\Gamma_n^1 \rangle$ by default. That is the main reason why S^1 is generally underestimated.

With the values I gave for D, I have obtained for S^1 (values not published in the frame of the intercomparison exercise):

Case C_1; D = 32.8 ev; $\langle g\Gamma_n^1 \rangle$ = 23.8 mev $\longrightarrow$ $S_1 = 5.44.10^{-4}$

Case C_2; D = 44.13ev; $\langle g\Gamma_n^1 \rangle$ = 23.7 mev $\longrightarrow$ $S_1 = 4.04.10^{-4}$

Mr. Thompson should have found similar results.

CONCLUSIONS

P. Ribon, CEN, Saclay, France

We can distinguish four stages in this exercise, with very different purposes and organisation:

Stage 1 - Assessing the problem

The purpose was to display the discrepancies, if they existed, and to convince physicists of their existence. The exercise was open to all physicists who wished to take part.

The results of this stage are clearly explained by M. Moore (part I of this report). The most difficult cases showed:

- results surprisingly bad: disagreements were outside error bars;
- physicists apparently had too much confidence in their own codes;
- the origin of the discrepancies was unclear.

Stage 2 - Investigation of the origins of disagreements and choice of a few basic codes

This stage was restricted to physicists having participated in the first stage. A second benchmark was conceived in order to try to provide information about the origins of discrepancies, and gave to participants the possibility of testing the improvements they applied to their own codes.

A two-day workshop terminating this stage showed that:

- origins of divergences were still not clear, even between codes using the same method (ESTIMA and CAVECN);
- very few codes (those of M. Moore and H. Weigmann) provide good estimates of p-wave strength functions;
- one code (BAYESX, by M. Moore) was rather automatic and requires little personal judgement. A faster version, BAYESZ, was later made available.

It was decided, at the workshop, to devote attention to three or four codes, to run them and to check their processing.

Stage 3 - Testing the codes by a "neutral" physicist and providing them to the scientific community

This stage confirms the limit in accuracy we can currently expect for a difficult nucleus. It showed that, when using codes according to the methods prescribed by their authors, the errors are to a large extent independent of the physicist using them (cf. part II of this report by A. Thompson).

It also showed the difficulty in obtaining from the authors final versions of their programs, properly documented for distribution.

Stage 4 - Follow-up

As a result of stage 3, the NEA Data Bank can send without restriction, upon request, two very different codes: BAYESZ and ESTIMA.

Stage 4 will consist in the use of these codes by physicists, and it is requested that users' comments be sent to NEA Data Bank, for transmission to the authors and possibly to other users.

Overall Conclusions

There was some delay in obtaining the program authors' agreement to take part in this exercise, due to some scepticism about its usefulness, and the fear that it could develop into a self-perpetuating project.

Considerable effort was invested in this benchmark, mainly by the organisers, and we believe that in the event it has proved fruitful:

- The eight to ten physicists working independently in different laboratories on this problem have come into direct contact with the others and with their programs. Their confidence in their own programs has perhaps been shaken: certainly all the authors taking part made some modifications as a result of the exercise.

- A more realistic assessment of the accuracy of these codes can now be made.

- In order to participate in Stage 3, authors made an effort to increase the portability and extend the documentation of their codes.

- The two codes finally made available for open distribution are likely to be widely used, and may form a common basis for comparing results and investigating difficulties in the use of other codes.