Benford's law and complex atomic spectra

Jean-Christophe Pain*
Commissariat à l'Energie Atomique CEA/DIF, Boîte Postale 12, 91680 Bruyères-Le-Châtel Cedex, France
(Received 7 September 2007; revised manuscript received 24 October 2007; published 8 January 2008)

We found that in transition arrays of complex atomic spectra, the strengths of electric-dipolar lines obey Benford's law, which means that their significant digits follow a logarithmic distribution favoring the smallest values. This indicates that atomic processes result from the superposition of uncorrelated probability laws and that the occurrence of digits reflects the constraints induced by the selection rules. Furthermore, Benford' law can be a useful test of theoretical spectroscopic models. Its applicability to the statistics of electric-dipolar lines can be understood in the framework of random matrix theory and is consistent with the Porter-Thomas law.

DOI: 10.1103/PhysRevE.77.012102 PACS number(s): 02.50.Cw, 32.70.Cs, 31.90.+s

I. INTRODUCTION

Newcomb [1] noticed that the beginning pages of logarithm books were more used than the last pages. This observation led to the conjecture that the significant digits of many sets of naturally occurring data describing the physical world are not equiprobably distributed, but in a way that favors smaller significant digits. For instance, the first significant digit (i.e., the first digit which is nonzero) will be 5 more frequently than 6 and the first three significant digits will be 458 more often than 462. Moreover, this result is scale invariant. Benford provided a probability distribution function for significant digits. The probability that the N first significant digits d_i $(i=1,N)$ are equal, respectively, to k_i $(k_i=1,9)$ is given by [2]

$$P(d_i=k_i)=\log_{10}\left(1+\left[\sum_{i=1}^{N} d_i\times 10^{N-i}\right]^{-1}\right). \qquad (1)$$

For instance, since some accounting data satisfy this rule, it is used to detect fraud [3]. River lengths, mountain heights, populations of cities [4], radioactive decay half lives [5], physical constants, and some mathematical series satisfy Benford's law. However, if many data roughly satisfy the law, perfect agreement is not the general case and the law is not universal. In the present work, focus will be put on the first significant digit, whose probability of being equal to k is given by

$$P(d_1=k)=\log_{10}(1+k^{-1}). \qquad (2)$$

II. DIGITS OF LINE STRENGTHS

To our knowledge, although much attention has been paid to the modeling of electric-dipolar $(E1)$ transition lines in a complex atomic spectrum typical of a hot plasma, it has never been checked whether the statistics of the lines obey Benford's law. Most of the studies focused on the distribution of the values of the line strengths (see, for instance, Ref. [6] or [7]), but not on the significant digits. Figure 1 represents the transition array $3d^6\rightarrow 3d^5 4p$ of Ge IX calculated with Cowan's *ab initio* atomic-structure code [8], which provides the Slater direct and exchange integrals, and the one-particle dipole-moment integrals evaluated from Hartree-Fock wave functions. Figure 2 displays the number of lines of that transition array for each value of the first significant digit calculated from our approach and predicted from Benford's law. The total number of lines is 3245. It is clear that the two distributions of lines are very similar. The same property has been observed on other transition arrays for different elements. Table I displays the fraction of lines per significant digit for four transitions: $3d^6\rightarrow 3d^5 4p$ (3245 lines) for Ge IX (T_1), $3d^4\rightarrow 3d^3 4f$ (2825 lines) for Ta LII (T_2), $3d^2 4s 4p\rightarrow 3d 4s 4p^2$ (2722 lines) for Br XIV (T_3) and $3d^2 4s 4p^3\rightarrow 3d 4s 4p^4$ (8231 lines) for Br XII (T_4). It is worth mentioning that even for transition arrays with a much smaller number of lines, the law still applies very well.

III. TESTING SPECTROSCOPIC MODELS

In a hot plasma, the total spectrum consists of a huge number of transition arrays. Therefore, the total number of lines can be immense. Usually, when the lines coalesce due to the physical broadening processes (natural width, electron impact, Stark effect, and Doppler effect), the distribution of

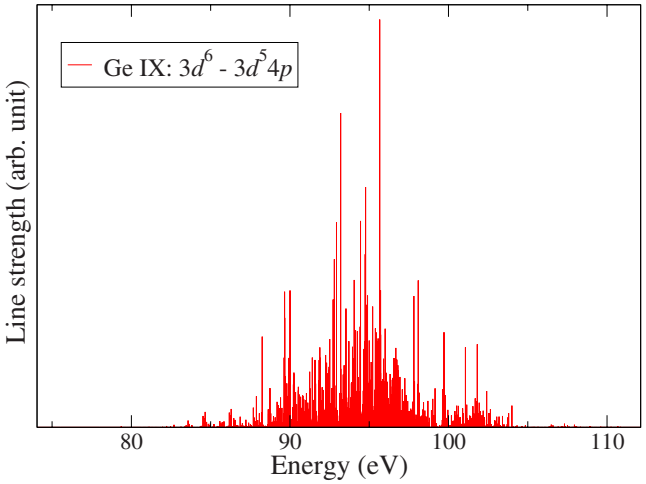

FIG. 1. (Color online) Array of lines for the transition $3d^6\rightarrow 3d^5 4p$ of Ge IX.

*jean-christophe.pain@cea.fr

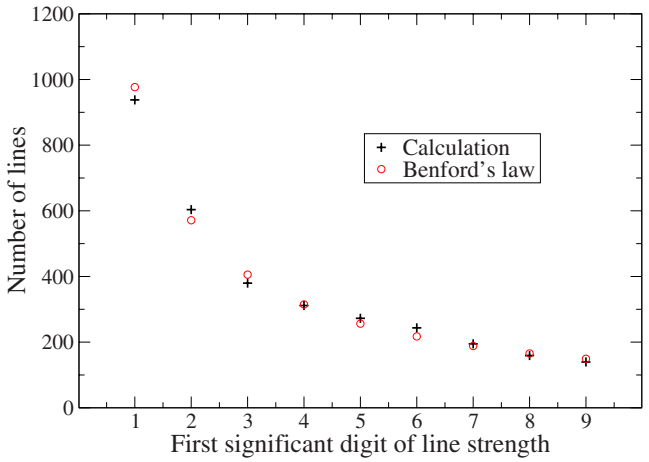

FIG. 2. (Color online) Fraction of lines versus first significant digit for the transition $3d^6 \rightarrow 3d^5 4p$ of Ge IX. The total number of lines is 3245.

lines in intermediate coupling is modeled by a continuous (Gaussian) distribution, called the unresolved transition array whose average energy and variance are evaluated following the work of Bauche *et al.* [9]. However, in that case some information is lost, and random calculations can help to approximately reconstruct the resolved distribution of lines ensuring known physical rules [10].

The resolved transition array (RTA) approach developed by Bauche *et al.* [11] and recently improved by Gilleron *et al.* [12] enables one to simulate $E1$ lines of a transition array avoiding direct diagonalization of the atomic Hamiltonian. The strength of a line and the energies of the lower and upper levels are picked up at random satisfying some specific constraints: preservation of the number of lines, of the total strength and weighted variance of the transition array, as well as of the unweighted variances of levels of lower and upper configurations. In that approach, the distribution function also includes a correlation between the energies and amplitudes of the lines (propensity law). Therefore, Benford's significant-digit law can be a performing test of the validity of the correlation rules included in such approaches.

IV. INTERPRETATION

The fact that $E1$ lines satisfy Benford's law indicates that the distribution of digits reflects the symmetry due to the selection rules. If transitions were governed by uncorrelated random processes, each digit would be equiprobable.

Benford's law is still not fully understood mathematically. However, some key elements have been known since the middle of the 1990s. For instance, Benford's law holds for any rescaling of the data, since it does not depend on any particular choice of units [13]. It is the only scale-invariant law referring to digits since the logarithmic distribution is the unique continuous base-invariant distribution [14]. More precisely, if probability distributions are selected at random and random samples are taken from each of these distributions so that the overall process is base invariant, then the significant-digit frequencies of the sample will converge to the logarithmic distribution. In other words, the more diversified the probability distributions are, the better the data sets fit Benford's law. The understanding of why systems with many interacting particles spontaneously organize into scale-invariant states is a difficult task of statistical physics.

TABLE I. Fraction of lines whose first digit is k for transitions T_1, T_2, T_3, and T_4 compared to the values predicted by Benford's law.

k	T_1	T_2	T_3	T_4	Benford
1	0.289	0.307	0.307	0.300	0.301
2	0.186	0.182	0.179	0.186	0.176
3	0.117	0.121	0.126	0.117	0.125
4	0.096	0.096	0.093	0.097	0.097
5	0.084	0.079	0.079	0.076	0.079
6	0.075	0.064	0.071	0.071	0.067
7	0.060	0.060	0.055	0.055	0.058
8	0.049	0.051	0.049	0.050	0.051
9	0.043	0.039	0.043	0.048	0.046

Moreover, Benford's law applies if the system is governed by multiplicative processes [15]. Indeed, a random multiplicative process corresponds to an additive process (dynamical description of a Brownian process) in a logarithmic space. In Wigner's random matrix theory [16,17], the Hamiltonian is defined in the Gaussian orthogonal ensemble (GOE) by an ensemble of real symmetric matrices whose probability distribution is a product of the distributions for the individual matrix elements H_{kl}, which are considered to be stochastic variables. The variance of the distribution for the diagonal elements is twice the one for the off-diagonal elements. The line strength S is proportional to $|\langle i|\vec{D}|j\rangle|^2$, where $\vec{D}$ is the dipolar operator and $|i\rangle$ and $|j\rangle$ are eigenvectors of the Hamiltonian. Therefore, the line strengths involve quantities such as products of H_{kl} and one has

$$\frac{S}{S'} = \zeta, \tag{3}$$

where S and S' are the strengths of two lines belonging to the transition array and ζ is a stochastic variable. Equation (3) can be written as

$$\ln S = \ln \zeta + \ln S'. \tag{4}$$

The central-limit theorem states that the probability distribution that the value of the nth strength is S will be Gaussian with a variance $\propto n^{1/2}$. In the infinite limit, the distribution will approach the uniform one, characterized by constant K. Therefore, one has

$$\int P(\ln S) d(\ln S) = K \int \frac{dS}{S}. \tag{5}$$

The probability P that the first significant digit d_1 of S is k in base 10 is given by

$$P(d_1 = k) = \int_k^{k+1} \frac{dS}{S} \Big/ \int_1^{10} \frac{dS}{S} = \log_{10}\left(1 + \frac{1}{k}\right), \quad (6)$$

which is exactly Eq. (2). This heuristic argument relies on the idea that fluctuations are governed by multiplicative processes involving a stochastic variable. The matrix elements of the Hamiltonian are correlated stochastic variables and the product of such variables leads to Benford's logarithmic distribution of digits.

Another argument can be invoked to understand the mathematical foundation of Benford's law. Porter and Thomas have shown that the amplitudes of the lines between all the levels of two random matrices obey a Gaussian distribution [18,19]. The strength being the square of the amplitude, its distribution is [6,20]

$$D(S) = \frac{L}{\sqrt{2\pi\langle S\rangle S}} \exp\left[-\frac{S}{2\langle S\rangle}\right], \quad (7)$$

where L and $\langle S\rangle$ are, respectively, the number of lines and the average value of the line strength in a (J,J') set. When numbers are taken from an exponential distribution, they automatically obey Benford's law. Therefore, when the exponential term dominates in $D(S)$, which is often the case except close to the origin (i.e., for very weak lines), Benford's law applies.

The random matrix theory contains approximate symmetries, which are not sufficient to describe the vicinity of Russell-Saunders and jj couplings [21,22]. In the model of Wilson *et al.* [23], diagonal terms are calculated in a pure coupling using Cowan's code mentioned above and off-diagonal elements are populated statistically beyond the GOE according to a bi-Gaussian distribution function where elements are correlated. They observed a disproportionably large number of off-diagonal elements of small amplitude. However, we found that even when the random matrix theory is expected to be inappropriate, i.e., close to a pure coupling, the line strengths still fit Benford's law. This is due to the fact that even if the number of weak emerging lines is important [when the term $\frac{1}{\sqrt{S}}$ dominates in the Porter-Thomas law (7)], the number of decades is sufficiently large so that the weight of those weak lines is negligible in the statistical occurrence of digits.

V. CONCLUSIONS

The results of this study are following. First, the distribution of lines in a given transition array follows Benford's logarithmic law of significant digits very well, which was, to our knowledge, never observed and rather unexpected. Second, this indicates that the correlations due to the selection rules manifest themselves in the distribution of digits, even in intermediate coupling, and that such a symmetry results from a superposition of different probability laws. Third, computational methods avoiding diagonalization such as the resolved transition arrays model, which ensures known physical rules (propensity law, correlation laws), can be improved and/or tested by imposing the constraint that the statistics of the lines must verify Benford's law with a high accuracy. Finally, since Benford's law can be explained in terms of a dynamics governed by multiplicative stochastic processes, the random matrix theory is probably an interesting pathway for the calculation of large atomic-dipole transition arrays and Benford's law can help clarifying the existence of different classes of stochastic Gaussian variables. Moreover, Benford's law is a signature of the Porter-Thomas distribution of lines.

ACKNOWLEDGMENT

The author would like to thank F. Gilleron for numerous helpful discussions.

[1] S. Newcomb, Am. J. Math. **4**, 39 (1881).
[2] F. Benford, Proc. Am. Philos. Soc. **78**, 551 (1938).
[3] M. Nigrini, J. Am. Tax. Assoc. **18**, 72 (1996).
[4] J. Torres, S. Fernandez, A. Gamero, and A. Sola, Eur. J. Phys. **28**, L17 (2007).
[5] B. Buck, A. C. Merchant, and S. M. Perez, Eur. J. Phys. **14**, 59 (1993).
[6] J. Bauche and C. Bauche-Arnoult, Comput. Phys. Rep. **12**, 3 (1990).
[7] R. C. W. Learner, J. Phys. B **15**, L891 (1982).
[8] R. D. Cowan, *The Theory of Atomic Structure and Spectra* (University of California Press, Berkeley, Los Angeles, 1981).
[9] C. Bauche-Arnoult, J. Bauche, and M. Klapisch, Phys. Rev. A **20**, 2424 (1979).
[10] U. Fano, Phys. Rev. A **32**, 617 (1985).
[11] J. Bauche, C. Bauche-Arnoult, and M. Klapisch, J. Phys. B **24**, 1 (1991).
[12] F. Gilleron, J. Bauche, and C. Bauche-Arnoult, J. Phys. B **40**, 3057 (2007).
[13] R. A. Raimi, Am. Math. Mon. **83**, 521 (1976).
[14] T. Hill, Stat. Sci. **10**, 354 (1995).
[15] L. Pietronero, E. Tossati, V. Tossati, and A. Vespignani, Physica A **293**, 297 (2001).
[16] M. Mehta, *Random Matrices and the Statistical Theory of Energy Levels* (Academic, New York, 1967).
[17] H. S. Camarda and P. D. Georgopulos, Phys. Rev. Lett. **50**, 492 (1983).
[18] C. E. Porter and R. G. Thomas, Phys. Rev. **104**, 483 (1956).
[19] C. E. Porter, *Statistical Theories of Spectra: Fluctuations* (Academic Press, New York, NY, 1965).
[20] J. Bauche, C. Bauche-Arnoult, and J.-F. Wyart, J. Quant. Spectrosc. Radiat. Transf. **51**, 41 (1994).
[21] J. Bauche, C. Bauche-Arnoult, J.-F. Wyart, P. Duffy, and M. Klapisch, Phys. Rev. A **44**, 5707 (1991).
[22] J. Bauche and C. Bauche-Arnoult, Phys. Rev. Lett. **69**, 1038 (1992).
[23] B. G. Wilson, F. Rogers, and C. Iglesias, Phys. Rev. A **37**, 2695 (1988).