



Contents lists available at ScienceDirect

## Journal of Quantitative Spectroscopy and Radiative Transfer

journal homepage: [www.elsevier.com/locate/jqsrt](http://www.elsevier.com/locate/jqsrt)

## Benford's law in atomic spectra and opacity databases

Jean-Christophe Pain<sup>a,b</sup>, Yuri Ralchenko<sup>c,\*</sup><sup>a</sup> CEA, DAM, DIF, F-91297 Arpajon, France<sup>b</sup> Université Paris-Saclay, CEA, Laboratoire Matière en Conditions Extrêmes, F-91680 Bruyères-le-Châtel, France<sup>c</sup> National Institute of Standards and Technology, Gaithersburg, MD 20899-8422, USA

## ARTICLE INFO

## Keywords:

Atomic physics  
Spectroscopy  
Databases  
Einstein coefficients  
Oscillator strengths  
Transition wavelengths

## ABSTRACT

The intriguing law of anomalous numbers, also named Benford's law, states that the significant digits of data follow a logarithmic distribution favoring the smallest values. In this work, we test the compliance with this law of the atomic databases developed at the National Institute of Standards and Technology (NIST) focusing on line energies, oscillator strengths, Einstein coefficients and radiative opacities. The considered databases are the Atomic Spectra Database (ASD) and the NIST-LANL (Los Alamos National Laboratory) Lanthanide/Actinide Opacity Database. The present study is not limited to the first digit and the case of multipole lines is also considered. The fact that the law is verified with a high accuracy means that the occurrence of digits reflects the constraints induced, in a given angular-momentum coupling, by the selection rules for atomic processes. As a consequence, Benford's law may be of great interest to detect inconsistencies in atomic databases.

## 1. Introduction

In the era of big data, quality control has become a key factor and it is important to establish scientific quality detection methods. Benford's law [1,2] (BL), sometimes called Newcomb–Benford's law, has become an effective tool for detection of data quality and identification of anomalies in various fields. Testing data quality and mining defects can be achieved by comparing the BL distribution with the actual distribution of the first digit in the data. It can also be combined with various theoretical techniques to solve other practical application problems (see extensive bibliography on BL in [3]). It has been widely used in the fields of natural and social sciences [4–6].

It was noticed (see Refs. [7–11]) that the distribution of the strengths of spectral lines in a given transition array follows very well Benford's logarithmic law of significant digits. Bhole et al. showed that BL is a useful paradigm for molecular spectroscopic analysis [12] and Bormashenko et al. discussed its applicability and breakdown in the infrared spectra of polymers [13].

The applicability of the law, however, was never checked over a large-size database of atomic spectroscopic data. In the present work, we study the validity of BL in the case of two widely-used databases developed at the National Institute of Standards and Technology (NIST). The first one, the Atomic Spectra Database (ASD), provides access and search capability for NIST critically evaluated data on atomic energy levels, wavelengths, and transition probabilities that are reasonably up-to-date [14]. The NIST Atomic Spectroscopy Data Center has carried out these critical compilations. The other collection of data analyzed in

this work is related to radiative opacities. An extensive set of opacities under the assumption of local thermodynamic equilibrium (LTE) has been calculated and made available to the community in the NIST-LANL (Los Alamos National Laboratory) Lanthanide/Actinide Opacity Database [15] for the lanthanide elements with atomic number  $57 \leq Z_N \leq 70$  [16] and actinides with  $89 \leq Z_N \leq 102$  [17]. Radiative opacities for a significant fraction of the elements in the periodic table are required to model the light curves and spectra produced by kilonovae.

The main features of Benford's law are recalled in Section 2 and its relevance in the context of complex atomic spectra is discussed in Section 3. Tests of the compliance to the law on ASD and the NIST-LANL Lanthanide/Actinide Opacity Database are presented in Sections 4 and 5, respectively. The case of electric and magnetic multipole transitions is addressed in Section 4 as well. The usefulness of the investigation of the first two significant digits as an additional source of information is presented in Section 6 which is followed by Conclusions.

## 2. Benford's law

## 2.1. Brief history and description

In 1881, Newcomb, an American astronomer and mathematician, discovered the law of the first digit [1]. Through statistical analysis of the data, he pointed out that many numbers characteristic of nature

\* Corresponding author.

E-mail addresses: [jean-christophe.pain@cea.fr](mailto:jean-christophe.pain@cea.fr) (J.-C. Pain), [yuri.ralchenko@nist.gov](mailto:yuri.ralchenko@nist.gov) (Y. Ralchenko).<https://doi.org/10.1016/j.jqsrt.2024.109010>

Received 15 February 2024; Received in revised form 5 April 2024; Accepted 16 April 2024

Available online 21 April 2024

0022-4073/© 2024 Elsevier Ltd. All rights reserved.

can be expressed in an exponential form, and that the logarithm of the mantissa is uniformly distributed. Later and independently, Benford observed the same phenomenon [2], and conducted deep research on 20229 numerical values in more than 20 data sets appearing randomly in physical and chemical constants, prime numbers and Fibonacci numbers, river lengths and lake areas. He arrived to the same conclusions as Newcomb, and proposed the formal “first digit law”, arousing the interest of many scientists. This law was named later after Benford [18].

Thus, Benford’s law means that the significant digits of many sets of naturally occurring data are not equi-probably distributed, but in a way that favors smaller significant digits through a uniform logarithmic distribution. For instance, the first significant digit, *i.e.*, the first digit which is non zero, will be 5 more frequently than 6, and the first three significant digits will be 548 more often than 576.

## 2.2. Applications

The verification of the applicability of BL has been checked in different fields [19]. For instance, relevant research has been carried out in economics, sociology, physics, computer science and biology. Among them, the most influential research is the systematic applicability to the detection of tax fraud [20]. It is found through statistics that computer file size [21], biological protein domain length, popular survival distribution [22], spectral line strength [7] in complex atomic spectroscopy, hadron lifetime or energy loss rate of pulsar self-rotation slowing [23], all verify the law very closely. Even the distributions of distances of galaxies and stars conform to BL [24]. However, many data sets also have not passed the applicability verification, such as the resident identity number or the height of the adult humans for instance. Yunxia studied the important economic data of the national development zone and found that some numbers were inconsistent with the Benford frequency [25]. Experiments prove that the data distribution in many fields obeys BL, but there are also data sets that violate it. Such data can be roughly classified into two categories, one in which the data do not satisfy the objective condition for applying the law and the other containing anomalous data, which break the law. In view of the former, the current theoretical basis cannot determine the reasonable data set *a priori*, and it still needs to be judged by rigorous statistical analysis, and this is precisely its limitation. For the latter, scientists skillfully use this feature to develop BL as the effective tool for assessing data quality and detecting potential problems or irregular data.

## 2.3. Mathematical formulation

Benford proposed a probability distribution function for significant digits, which states that the probability that the first significant digit  $d_1$  is equal to  $k$  is given by [2]:

$$\mathcal{P}(d_1 = k) = \log_{10} \left( 1 + \frac{1}{k} \right). \quad (1)$$

Although some of the properties of BL were explained, it was not rigorously proved from a mathematical point of view and the literature on the subject remains substantial. In 1976, a paper whose claimed purpose was to review all the proposed explanations of the law, gives not less than thirty-seven Refs. [26]. Hill showed that the fact that the first digit obeys BL [27] is consistent with the central limit theorem. In addition, he generalized the theory, obtained the distribution law of higher-order digits, and derived the joint distribution function between the first digit and the higher-order digits. Recently, Wang and Ma proposed a concise proof of the famous BL when the distribution has a Riemann integrable probability density function. These authors also provided a criterion to judge whether a distribution obeys the law [28].

## 3. Validity for complex atomic spectra

Actually, Benford’s law can be proved to apply if the system is governed by random multiplicative processes [29], *i.e.*, processes which are additive in a logarithmic space. In Wigner’s Random Matrix Theory (RMT) [30,31], the Hamiltonian is defined in the Gaussian Orthogonal Ensemble (GOE) as a set of real symmetric matrices whose probability distribution is a product of the distributions for the individual matrix elements [32], considered as stochastic variables, and the variance of the distribution for the diagonal elements is twice the one for the off-diagonal elements. The matrix elements of the Hamiltonian are correlated stochastic variables and the product of such variables, arising through the diagonalization process, leads to Benford’s logarithmic distribution of digits. Since BL can be explained in terms of a dynamics governed by multiplicative stochastic processes (additive in logarithmic space), RMT is an interesting tool for the calculation of large electric-dipole (E1) transition arrays [33], and Benford’s law can help clarifying the existence of different classes of stochastic Gaussian variables.

## 4. Test on the Atomic Spectra Database

### 4.1. Physical quantities

Let us start with a reminder of definitions of the main physical quantities to be analyzed below. The oscillator strength  $f_{ij}$  of an absorption (electric-dipole) transition between lower level  $i$  and higher level  $j$  is defined as [34]:

$$f_{ij} = \frac{8\pi\epsilon_0 m_e}{3\hbar^2} \frac{|\Delta E_{ij}| S_{ij}}{e^2 g_i}, \quad (2)$$

where  $g_i$  is the degeneracy of level  $i$ ,  $S_{ij} = S_{ji}$  is the line strength,  $|\Delta E_{ij}| = |E_j - E_i|$  ( $E_i$  being the energy of level  $i$ ) is the transition energy, and  $m_e$  is the electron mass.  $\hbar$  represents the reduced Planck constant and  $e$  the electron charge while  $\epsilon_0$  is the permittivity of vacuum. The line strength [35] is defined as

$$S_{ij} = |R_{ij}|^2, \quad (3)$$

where  $R_{ij} = \langle \psi_j | \hat{D} | \psi_i \rangle$  with  $\psi_i$  and  $\psi_j$  being the initial- and final-state wave functions, respectively, and  $R_{ij}$  the transition matrix element of the dipole operator  $\hat{D}$  ( $R_{ij}$  involves an integration over spatial and spin coordinates of all  $N$  electrons of the atom or ion). The Einstein absorption coefficient  $B_{ij}$  reads

$$B_{ij} = \frac{4\pi^2}{3\hbar^2 g_i} S_{ij} \quad (4)$$

and the Einstein spontaneous emission coefficient

$$A_{ji} = \frac{4|\Delta E_{ij}|^3}{3\hbar^4 c^3 g_j} S_{ij}. \quad (5)$$

The connection between transition probabilities, oscillator strengths, and line strengths is given by ( $A$  in  $s^{-1}$ , wavelength  $\lambda$  in m,  $S$  in  $m^2 C^2$ ):

$$A_{ji} = \frac{2\pi e^2}{m c \epsilon_0 \lambda^2} \frac{g_i}{g_j} f_{ij} = \frac{16\pi^3}{3\hbar c \epsilon_0 \lambda^3 g_j} S_{ij}. \quad (6)$$

Numerically, in customary units -  $A$  in  $s^{-1}$ ,  $\lambda$  in nm,  $S$  in atomic units ( $m_e = \hbar = e = 1$ ) -

$$A_{ji} = \frac{6.6703 \times 10^{13}}{\lambda^2} \frac{g_i}{g_j} f_{ij} = \frac{2.0261 \times 10^{15}}{\lambda^3 g_j} S_{ij}. \quad (7)$$

It is worth recalling that for  $S_{ij}$  and  $\Delta E_{ij}$  in atomic units, one has

$$f_{ij} = \frac{2}{3} \frac{|\Delta E_{ij}|}{g_i} S_{ij}. \quad (8)$$

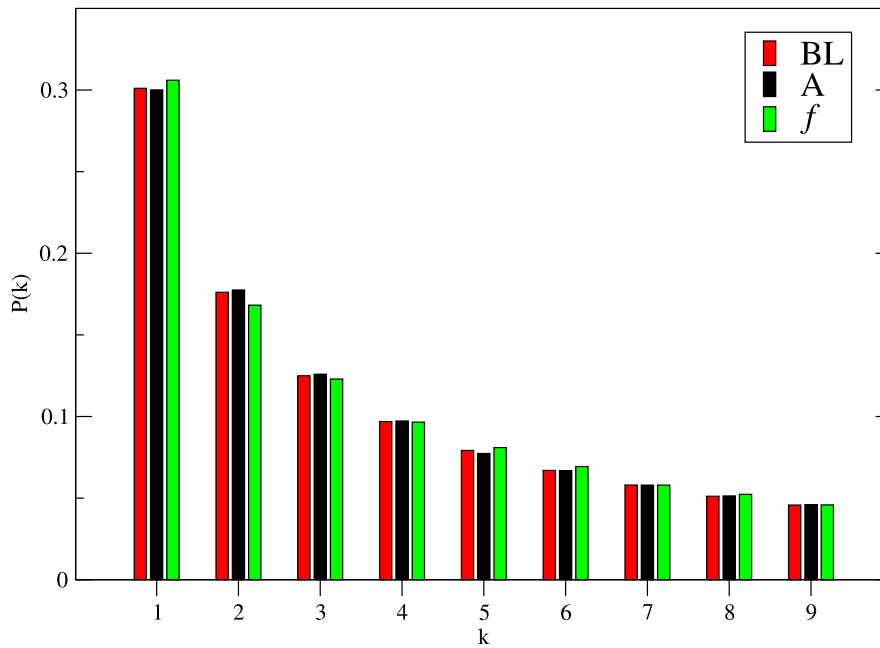


Fig. 1. Test of Benford's law (BL) for Einstein coefficients  $A$  and oscillator strengths  $f$  from the NIST ASD database.

#### 4.2. Description of the database

The NIST Atomic Spectra Database [14,36] contains about 300,000 spectral lines (about 40% of lines with transition probabilities) and 119,000 energy levels from more than 1,000 atoms and ions. Most of the data, about 60%, are for low- to mid- $Z_N$  elements with nuclear charges  $Z_N \leq 30$ . The spectroscopic data in ASD are mostly critically evaluated which, in particular, means that there is only one spectral line per transition. The database offers numerous search and selection choices for data extraction and presentation as well as rich graphics options, e.g., Grotrian diagrams and Saha-Local Thermodynamic Equilibrium (LTE) utility. Also, ASD provides ionization potentials for all ions of elements from H ( $Z_N = 1$ ) to Ds ( $Z_N = 110$ ).

#### 4.3. Results for the NIST ASD database

Fig. 1 represent comparisons between the predictions of Benford's law and Einstein coefficients  $A_{ji}$  as well as oscillator strengths  $f_{ij}$ . The total number of data is about 128,000 and 119,000, respectively. The difference in these numbers is due to the fact that it is customary in atomic databases to list oscillator strengths only for electric-dipole (E1) transitions. It is clear that the first-digit distributions of both quantities in Fig. 1 fit the law very closely. Again, as shown in Eq. (6),  $A$ - and  $f$ -values are not simply connected by a constant factor but rather a combination of energy differences and statistical weights and therefore such an agreement is an independent evidence of Benford's law (BL) validity for this type of atomic data in ASD.

However, the first-digit distributions for the *observed* wavelengths and energy levels in ASD reveal significant departures from BL (Fig. 2 (a)). This conclusion is likely to be explained by the fact that historically spectroscopic measurements were and still are mostly performed in the ultraviolet (UV) to visible spectral ranges. This is directly reflected in the contents of the database: the distribution of the number of lines vs. *actual wavelengths* (Fig. 2(b)) strongly peaks between 200 nm and 800 nm. In fact, out of about 235,000 observed wavelengths in ASD the number of UV-visible lines constitutes more than 51% of the total. As has been discussed in the literature (see, e.g., [37]), the distributions need to span more than an order of magnitude to obey BL and therefore the found deviations are not unexpected. As for the distribution of the transition probabilities in ASD (Fig. 3), most of them

also group near the values which are typical for UV-visible transitions, that is,  $A \approx (10^6-10^9) \text{ s}^{-1}$ . However, the span significantly exceeds one order of magnitude which brings  $A$  distribution much closer to the BL predictions.

Fig. 2(a) also presents the distribution of the available 5853 ionization potentials (in  $\text{cm}^{-1}$ ) in ASD that extend over more than 4 orders of magnitude. Clearly, this distribution much better agrees with Benford's law than the observed wavelengths or transition energies.

#### 4.4. Forbidden transitions

It is interesting to check whether the agreement with Benford's law remains for forbidden transitions. For E2 (electric-quadrupole) lines, one has ( $A^{(E2)}$  in  $\text{s}^{-1}$ ,  $\lambda$  in m,  $S^{(E2)}$  in  $\text{m}^4\text{C}^2$ ):

$$A_{ji}^{(E2)} = \frac{16\pi^5}{15h\epsilon_0\lambda^5 g_j} S_{ij}^{(E2)}. \quad (9)$$

Numerically, still in customary units -  $A^{(E2)}$  in  $\text{s}^{-1}$ ,  $\lambda$  in nm,  $S_{ij}^{(E2)}$  in atomic units ( $m_e = \hbar = e = 1$ ,  $a_0^4 e^2 = 2.013 \times 10^{-79} \text{ m}^4\text{C}^2$ ) - one has

$$A_{ji}^{(E2)} = \frac{1.1199 \times 10^{13}}{\lambda^5 g_j} S_{ij}^{(E2)}. \quad (10)$$

For M1 (magnetic dipole) lines, one has

$$A_{ji}^{(M1)} = \frac{16\pi^3 \mu_0}{3h\lambda^3 g_j} S_{ij}^{(M1)}, \quad (11)$$

where  $\mu_0$  is the permeability of vacuum. Numerically, in customary units -  $A$  in  $\text{s}^{-1}$ ,  $\lambda$  in nm,  $S^{(E2)}$  in atomic units ( $m_e = \hbar = e = 1$ ,  $e^2 \hbar^2 / (16\pi^2 m_e^2) = \mu_B^2 = 8.601 \times 10^{-47} \text{ J}^2\text{T}^{-2}$ ,  $\mu_B$  being the Bohr magneton) - one can write

$$A_{ji}^{(M1)} = \frac{2.6974 \times 10^{10}}{\lambda^3 g_j} S_{ij}^{(M1)}. \quad (12)$$

Finally, the M2 (magnetic quadrupole) transition probabilities are (in customary units):

$$A_{ji}^{(M2)} = \frac{1.4910 \times 10^8}{\lambda^5 g_j} S_{ij}^{(M2)}. \quad (13)$$

ASD contains transition probabilities for about 9000 M1, E2, and M2 forbidden lines. The forbidden  $A$ -values vary between approximately

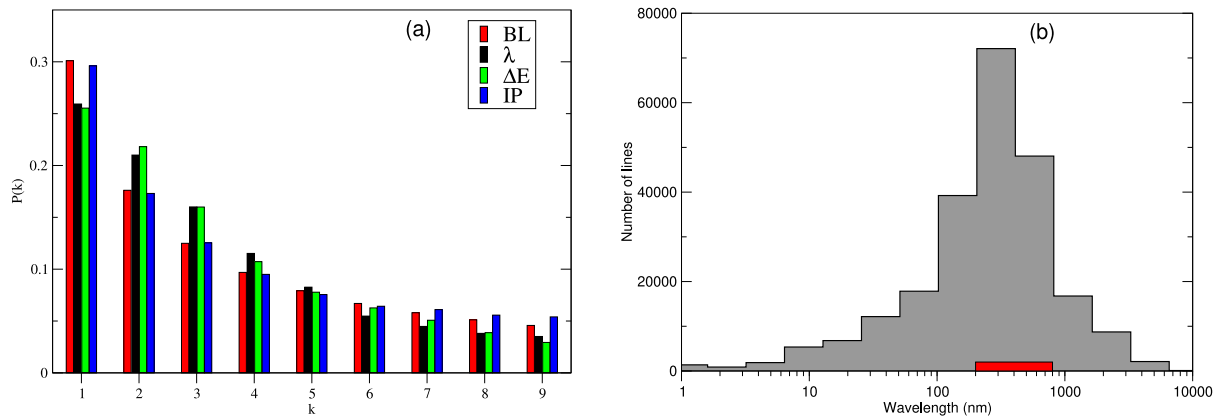


Fig. 2. (a) Test of Benford’s law (BL) for observed wavelengths (black), observed level energies (green), and ionization potentials (blue) from the NIST ASD database. (b) Distribution of the observed wavelengths in ASD. Red bar represents the ultraviolet–visible range of wavelengths from 200 nm to 800 nm.

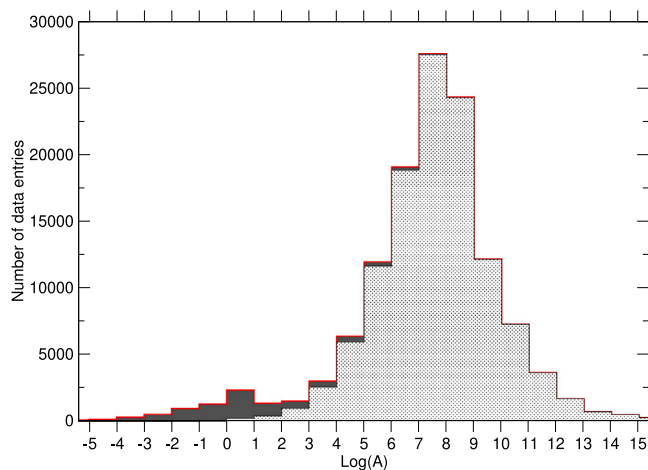


Fig. 3. Distribution of Einstein coefficients  $A$  vs.  $\log(A, \text{s}^{-1})$  from the NIST ASD database. The light-shaded area shows the distribution of allowed electric-dipole transitions.

$4 \cdot 10^{-24} \text{ s}^{-1}$  to  $3 \cdot 10^{12} \text{ s}^{-1}$  thus covering many orders of magnitude. Fig. 4 shows the comparisons of the M1, E2, and M2 distributions with Benford’s law. Although the statistics is not exceedingly high, nonetheless, both M1 and E2 Einstein coefficients agree with the law very well, on par with the much more extensive E1 data from ASD. The more substantial deviations for the M2 lines may be attributed to their relatively smaller statistics of just under 600 lines.

## 5. Test on the NIST-LANL Lanthanide/Actinide Opacity Database [15]

### 5.1. Description of the database

Kilonovae (or macronovae) are astronomical transient events that produce electromagnetic emission from matter ejected by the merger of two neutron stars or a neutron star and a black hole. This ejected matter is thought to be roughly 1/1000 to a few times 1/100 the mass of the Sun and move at about 1/10 the speed of light. The first kilonova associated with gravitational waves (detected by LIGO/Virgo) was observed in 2017 [38], in signals that spanned the electromagnetic spectrum. This was a significant discovery in astronomy, galvanizing research into the broad range of physics underlying neutron star mergers.

Radiative opacities for a significant fraction of the elements in the periodic table are required to model the light curves and spectra produced by kilonovae. In 2020, Fontes et al. [16] published an

extensive set of opacities calculated under the assumption of local thermodynamic equilibrium for the lanthanide elements with atomic number  $57 \leq Z_N \leq 70$ . Another set of opacities for actinides with  $89 \leq Z_N \leq 102$  was published by the same group in a recent paper [17].

The number of (bound-bound) absorption features, or lines, associated with these open  $f$ -shell elements can render the spectral simulations computationally intractable. Therefore, it is common to consider methods that group many lines within a particular photon-energy bin. The present opacities were generated with a line-binned treatment producing tabular results that are a function of temperature and density, but independent of the particular form of the hydrodynamic expansion chosen to model a kilonova. This independence eliminates the computational expense of repeatedly calculating opacities at specific conditions within the spectral simulations of kilonovae [16].

The total opacities presented here are split into various contributions according to the following formulae:

$$\kappa_{\text{tot}}(h\nu) = \kappa_{\text{scat}}(h\nu) + \kappa_{\text{abs}}(h\nu), \quad (14)$$

where  $h\nu$  is the photon energy,  $\kappa_{\text{scat}}(h\nu)$  is the scattering contribution due to Compton scattering and  $\kappa_{\text{abs}}(h\nu)$  is the absorption contribution given by

$$\kappa_{\text{abs}}(h\nu) = \kappa_{\text{bb}}(h\nu) + \kappa_{\text{bf}}(h\nu) + \kappa_{\text{ff}}(h\nu), \quad (15)$$

where  $\kappa_{\text{bb}}(h\nu)$  is the bound-bound (bb) contribution due to photo-excitation,  $\kappa_{\text{bf}}(h\nu)$  is the bound-free (bf) contribution due to photo-ionization, and  $\kappa_{\text{ff}}(h\nu)$  is the free-free (ff) contribution due to inverse Bremsstrahlung.

As mentioned above, the bound-bound opacities discussed here are computed with the line-binned treatment, *i.e.*, they are obtained via a discrete sum of all lines that occur between two points in the prescribed photon-energy grid, according to Eq. (2) in Ref. [16]. A complete description of the relevant temperature, density and photon-energy grids can be found in the aforementioned article [16]. The temperature grid consists of 27 values (in eV): 0.01, 0.07, 0.1, 0.14, 0.17, 0.2, 0.22, 0.24, 0.27, 0.3, 0.34, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.2, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0. Due to the fact that only the first four ion stages were considered for each element, the opacities are most accurate for temperatures below about 2 eV, even though data are provided up to a maximum temperature of 5 eV. The density grid contains seventeen values ranging from  $10^{-20}$  to  $10^{-4} \text{ g/cm}^3$ , with one value per decade. The photon energy grid contains 14,900 points for the (dimensionless) temperature-scaled quantity  $u = h\nu/k_B T$ , where  $k_B T$  is the thermal kinetic energy ( $k_B$  is the Boltzmann constant) expressed in the same units as the photon energy. A description of this  $u$  grid is available in Table 1 of Ref. [39].

The total number of analyzed data points per element is  $17 \times 27 \times 14900 = 6\,839\,000$  — this is actually the maximal possible number

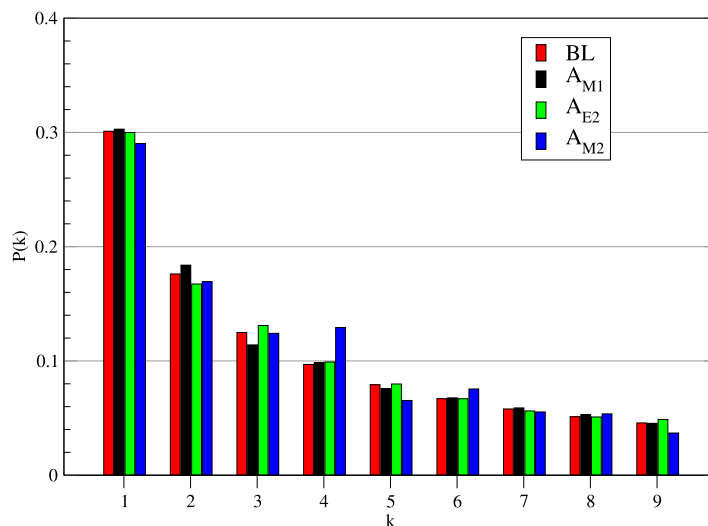


Fig. 4. Comparison of Einstein coefficients of M1 (3411 transitions), E2 (4572 transitions), and M2 (596 transitions) lines in ASD with Benford’s law (BL).

of data points. However, we did not include zero values in the present analysis and thus the actual number per element is typically smaller, about 4.5 millions.

### 5.2. Results for the NIST-LANL database

The comparison between Benford’s law and the available bound-bound and bound-free opacities for lanthanides from  $Z_N = 59$  to  $Z_N = 70$  and for actinides from  $Z_N = 89$  to  $Z_N = 102$  are presented in Figs. 5 and 6. Globally, the agreement for bound-bound opacities is very good, which is not exactly the case for the bound-free opacity. For instance, for Er  $Z_N = 68$  (Fig. 5, right), the bf-opacity distribution is almost 20% higher than BL predictions for the lowest digits 1 and 2, while for Th  $Z_N = 90$ , the values are too low in the beginning (digits 1 and 2) but exceed the BL values by about 25% for digit 4 (Fig. 6, left). Such noticeable deviations for bf-opacities are observed for all elements in the database.

The fact that the bound-free opacities depart from BL may be explained by the statistical distribution of the data. For bound-bound transitions, we found opacity values in all the decades, but for bound-free there are abrupt variations (due to the photo-ionization edges) with the opacity changing between very different values. One can see in Fig. 7 that the photo-excitation (bb) opacity (here  $Z_N = 94$ ) looks much more random while the photo-ionization (bf) covers almost fifty orders of magnitude and exhibits “quasi-plateaus” separated by edges. The structure of the photo-ionization emphasizes therefore some digits at the expense of others. In a sense, it is the same kind of phenomenon as for the observed energies/wavelengths (see Fig. 2), where we had a dominating number of lines in some specific ranges.

However, when we gather all the opacities of the elements of the database, one can see that the agreement of the bound-free opacity with the BL significantly improves (see Fig. 8). This is due to the fact that since all elements have their photo-ionization edges at different energies, the spread of opacity values is more homogeneous. If we considered a wider range of atomic numbers (not only lanthanides and actinides but also light and mid- $Z_N$  elements), the agreement with BL would probably be even better although this statement obviously requires tests with newly calculated opacities for other elements.

### 6. Test of the second-digit distribution

It is possible to extend the law to digits beyond the first digit [27]. Considering the first  $n$  digits ( $n > 1$ ) brings more stringent information about the data. For instance, in order to distinguish two distributions in

Table 1  
Probabilities (in %) that the first, second and third digits are equal to a given figure.

Figure	First digit	Second digit	Third digit
0	/	12.0	10.2
1	30.1	11.4	10.1
2	17.6	10.9	10.0
3	12.5	10.4	10.1
4	9.7	10.0	10.0
5	7.9	9.7	10.0
6	6.7	9.3	9.9
7	5.8	9.0	9.9
8	5.1	8.8	9.9
9	4.6	8.5	9.8

the perspective of their conformation to the law, if both of them follow the law with the same accuracy, looking at the first  $n$  digits may enable one to distinguish them. In particular, for any given number of digits, the probability of encountering a number starting with the string of digits  $n$  of that length – discarding leading zeros – is given by

$$\log_{10}(n + 1) - \log_{10}(n) = \log_{10} \left( 1 + \frac{1}{n} \right). \tag{16}$$

For example, the probability that a number starts with the digits 6, 5, 3 is  $\log_{10}(1 + 1/653) \approx 0.00066$ .

This result can be used to find the probability that a particular digit occurs at a given position within a number. For instance, the probability that a 2 is encountered as the second digit is [27]:

$$\log_{10} \left( 1 + \frac{1}{12} \right) + \log_{10} \left( 1 + \frac{1}{22} \right) + \dots + \log_{10} \left( 1 + \frac{1}{92} \right) \approx 0.109, \tag{17}$$

and the probability that  $d$  ( $d = 0, 1, \dots, 9$ ) is encountered as the  $n$ th ( $n > 1$ ) digit is

$$\sum_{k=10^{n-2}}^{10^n-1} \log_{10} \left( 1 + \frac{1}{10k + d} \right). \tag{18}$$

The distribution of the  $n$ th digit, as  $n$  increases, rapidly approaches a uniform distribution with 10% for each of the ten digits, as shown in Table 1.

More generally, for all positive integers  $p$ , the probability that a number begins by digits  $k_1, k_2, \dots, k_p$  is given by

$$\mathcal{P} (d_1 = k_1, \dots, d_p = k_p) = \log_{10} \left[ 1 + \left( \sum_{i=1}^p k_i \cdot 10^{p-i} \right)^{-1} \right]. \tag{19}$$

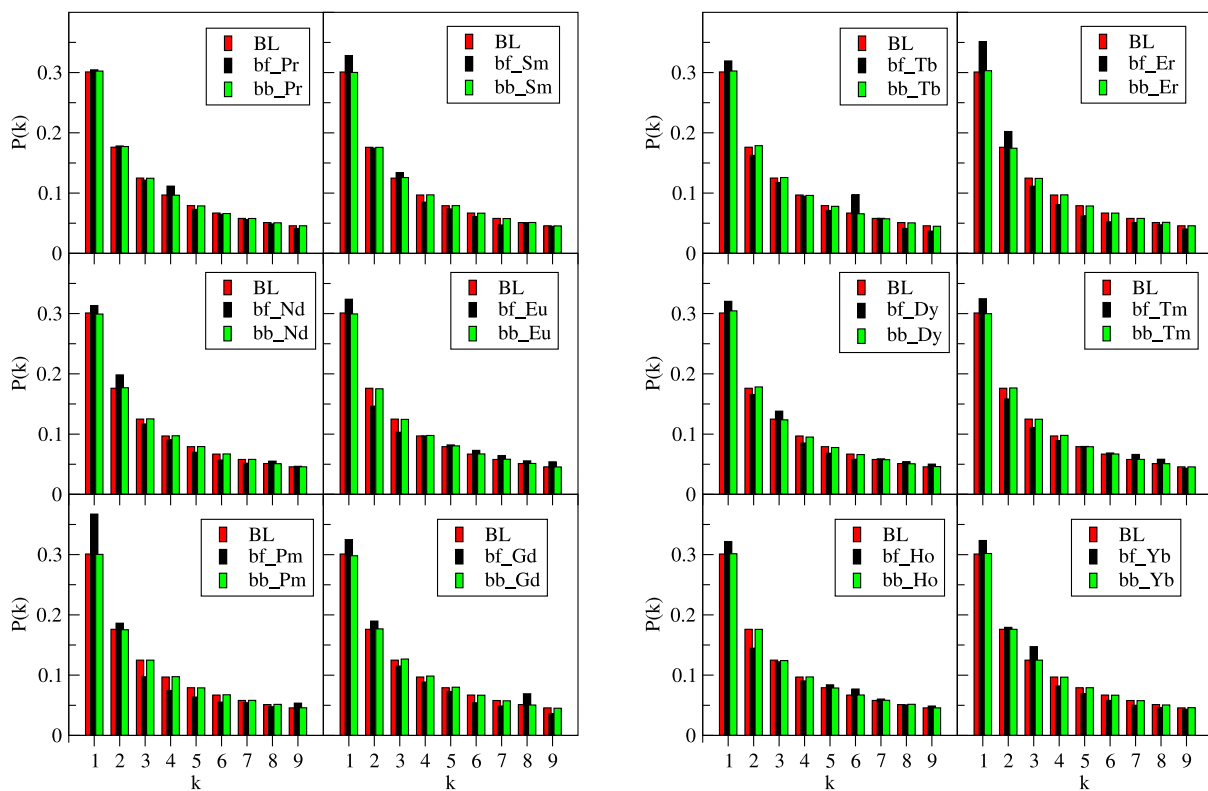


Fig. 5. Test of Benford's law (BL) for absorption coefficients for lanthanides ( $59 \leq Z_N \leq 70$ ) from the NIST-LANL Lanthanide/Actinide Opacity Database.

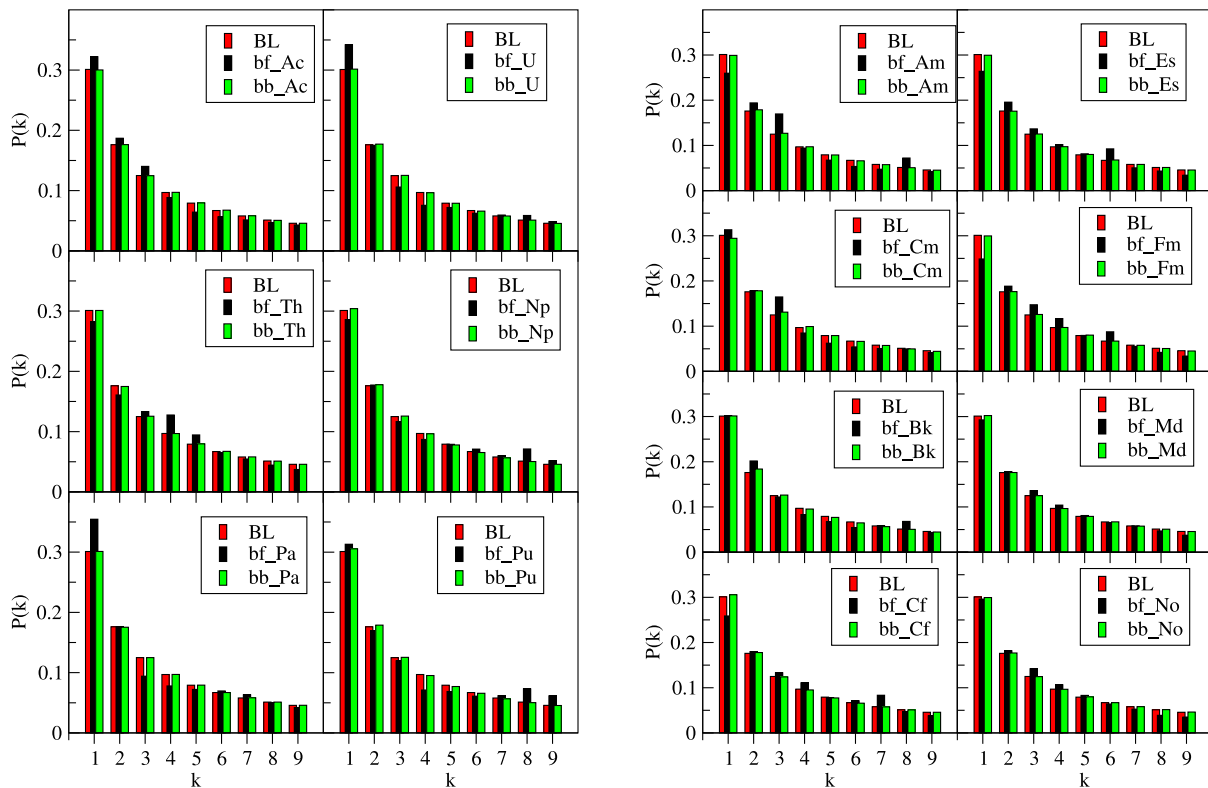


Fig. 6. Test of Benford's law (BL) for absorption coefficients for actinides ( $89 \leq Z_N \leq 102$ ) from the NIST-LANL Lanthanide/Actinide Opacity Database.

Thus, looking at just the first two significant digits, it is trivial to show that the probability of a number to begin with “10” is 4.14%, with “25” is 1.70% and so on.

Fig. 9 shows the ratio of the first two significant digit bb- and bf-distributions of all elements ( $Z_N = 59$  to  $Z_N = 70$  and  $Z_N = 89$  to  $Z_N = 102$ ) in the NIST-LANL opacity database to the corresponding BL

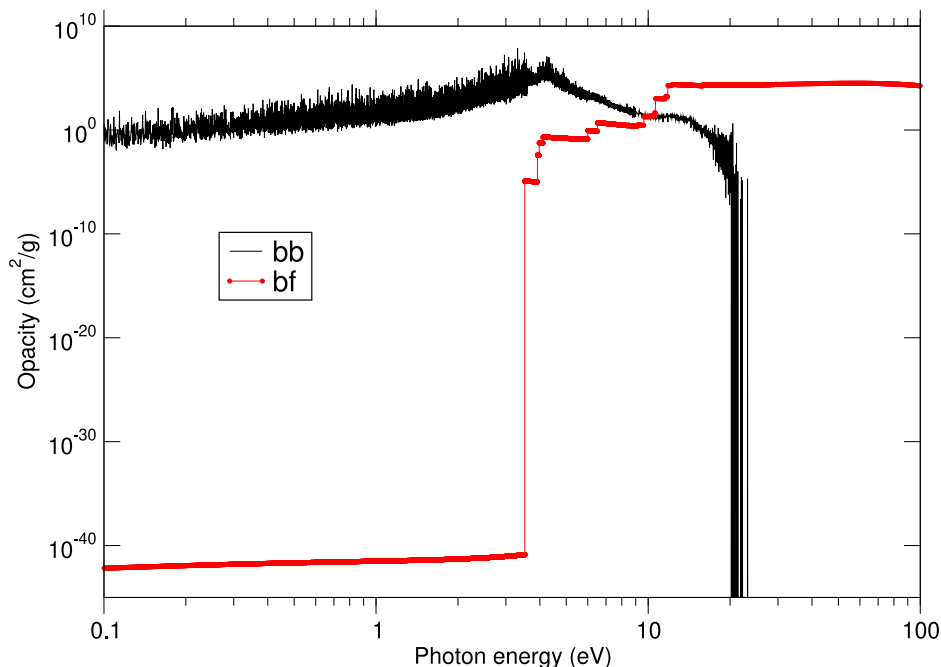


Fig. 7. Bound-bound and bound-free opacities of plutonium (12867 data points per opacity) from the NIST-LANL Lanthanide/Actinide Opacity Database at mass density of  $10^{-13} \text{ cm}^{-3}$  and temperature of 0.3 eV.

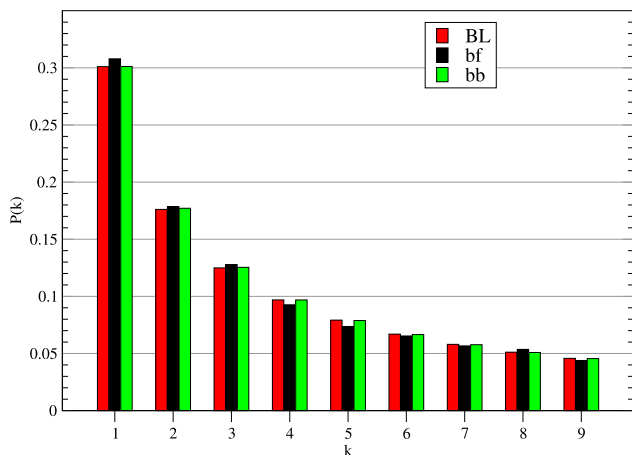


Fig. 8. Comparison between the first digits of bound-bound and bound-free opacities from the NIST-LANL Lanthanide/Actinide Opacity Database for all elements.

distribution (the latter is shown in the inset). Similar to the first-digit comparisons, the bb-opacities show better results that are just within 1% from the expected BL distribution. The bf-distribution agrees worse, the ratio bf/BL reaching 5% to 11% for some digit combinations.

### 7. Conclusions

Benford’s law is now known to apply to many sets of data concerning the physical world. Although its applicability has already been checked in atomic and molecular spectroscopy, it was never tested over entire open-access databases. We have investigated its validity in the case of two NIST databases, i.e., the Atomic Spectra Database and the NIST-LANL Lanthanide/Actinide Opacity Database, focusing on different physical quantities such as oscillator strengths, Einstein coefficients (transition probabilities), transition energies, observed transition wavelengths, photo-excitation opacity, and photo-ionization opacity. It appears that the law is very well reproduced for oscillator strengths and

Einstein coefficients in ASD. Strong deviations with respect to the law are observed for the transition energies and observed wavelengths in the case of ASD, but this can be explained by the fact that the database contains many lines in the UV-visible range that is most accessible in experimental spectroscopy. We also considered a few non-electric-dipole lines, and the agreement was found to be excellent, except for magnetic-quadrupole lines with relatively low statistics (their number and the spread of their strengths are not as important as for E1 lines for instance). In addition, absorption coefficients coming from the NIST-LANL Lanthanide/Actinide Opacity Database reveal also a very good agreement with the law, which tends to indicate that BL should be a powerful tool for checking the reliability of radiative-opacity tables [40–45] for local-thermodynamic-equilibrium plasmas, required for the simulation of stellar structure and evolution. It was found that the bound-free opacities from the NIST-LANL database do not comply as closely as the bound-bound ones with the predictions of BL. This was checked over twenty-six elements (lanthanides and actinides). This point was confirmed by the study of the two-digit distribution, for which the discrepancy with the corresponding Benford-like probability distribution is heightened.

Our main goal was to test the validity of Benford’s law in atomic spectra and opacity databases, i.e., to conduct a kind of numerical experiment, in order to pinpoint the problem and to define a perimeter (a set of conditions or characteristics of the database) within which the law applies. When it does not, we tried to explain why. For example, the fact that a large number of line wavelengths fall within a reduced range (due to experimental constraints or particular applications) has consequences (dependent on statistical representativity). Actually, if one finds that a collection of data does not follow Benford’s law when it should, this means that some data may be erroneous. However, this covers different kinds of anomalies: the data may be simply due to typographical errors or formatting issues, or may have a deeper meaning, and suggest that something went wrong in the calculation, and therefore motivate further investigations. Indeed, such a situation may itself cover different cases: a defect of the model or a computing mistake may either affect all the generated data (but in that case Benford’s law could be verified, either with wrong data), or only some lines in particular conditions.

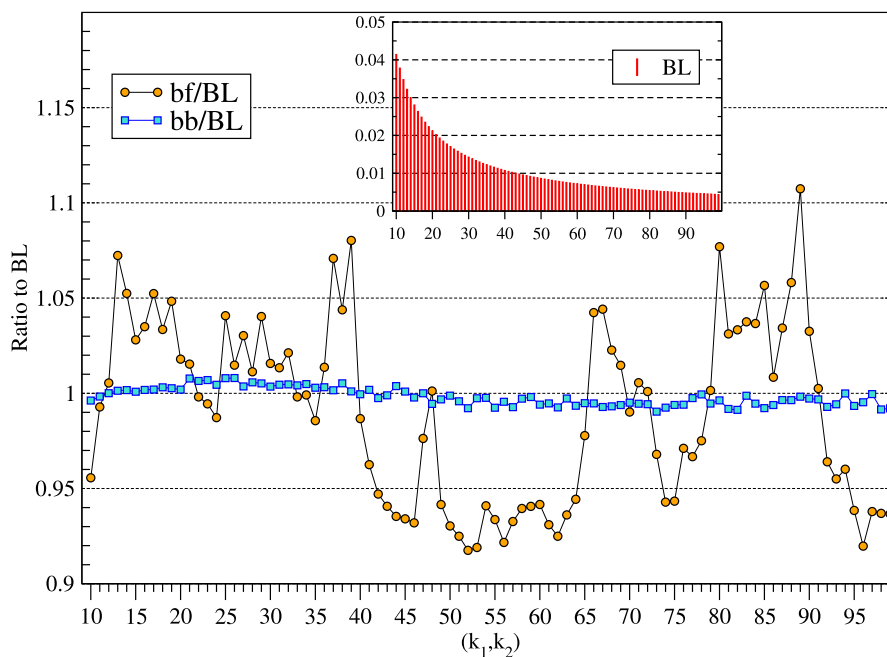


Fig. 9. Test of Benford's law (BL) for the first two digits in the case of photo-excitation (bb) and photo-ionization (bf) opacities from the NIST-LANL Lanthanide/Actinide Opacity Database. All chemical elements of the database are taken into account. The inset shows the 2-digit BL distribution.

The number of atomic databases is very limited, as building such data sets is a colossal and time-consuming task, requiring considerable expertise in atomic physics and computer science. The NIST databases are the most widely used in the community, this is the reason why we made the choice to focus on them. In addition, the NIST ASD database contains mostly pure atomic-physics quantities, and the NIST-LANL Lanthanide/Actinide Opacity Database gathers radiative opacities and their components, which include more ingredients than fundamental atomic-physics quantities. The two sets of databases are therefore complementary. In the case of the NIST-LANL Lanthanide/Actinide Opacity Database, this has enabled us to highlight problems with some photo-ionization cross-sections. We have also studied other databases, namely the Japan-Lithuania Opacity Database for Kilonova [46] and the Vienna Atomic Lines Database (VALD) [47]. The former gathers numerical data for opacities of heavy elements for ejecta of neutron star mergers. The latter is aimed at providing the relevant information for atomic transitions significantly contributing to absorption in stellar spectra. We analyzed the two databases and came to similar conclusions, but decided not to include them in the present article in order to improve the readability.

It would be interesting to investigate whether the law of anomalous numbers is also verified in the case of non-local-thermodynamic-equilibrium plasmas [48], for level populations and absorption and emission coefficients, for instance.

#### CRedit authorship contribution statement

**Jean-Christophe Pain:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Yuri Ralchenko:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Data curation, Conceptualization.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

All used data are available in the public databases referred to in the paper.

#### Declaration of Generative AI and AI-assisted technologies in the writing process

The authors did not use AI.

#### Acknowledgments

We are grateful to C.J. Fontes for valuable discussions.

#### References

- [1] Newcomb S. Note on the frequency of use of the different digits in natural numbers. *Am J Math* 1881;4:39–40.
- [2] Benford F. The law of anomalous numbers. *Proc Am Philos Soc* 1938;78:551–72.
- [3] Benford Online Bibliography, Berger A, Hill TP, Rogers E. 2009. <http://www.benfordonline.net>. [Last accessed 8 September 2023].
- [4] Barabesi L, Cerioli A, Perrotta D. Forum on Benford's law and statistical methods for the detection of frauds. *Stat Methods Appl* 2021;30:767–78.
- [5] Berger A, Hill TP. An introduction to Benford's law. Princeton, NJ: Princeton University Press; 2015.
- [6] Berger A, Hill TP. What is... Benford's law? *Not Am Math Soc* 2017;64:132–4.
- [7] Pain J-C. Benford's law and complex atomic spectra. *Phys Rev E* 2008;77:012102.
- [8] Pain J-C. Regularities and symmetries in atomic structure and spectra. *High Energy Density Phys* 2013;9:392–401.
- [9] Pain J-C. Structure atomique, équation d'état et propriétés radiatives des plasmas chauds [Habilitation Thesis], Paris-Saclay University; 2021, [in French] <https://tel.archives-ouvertes.fr/tel-03325468>.
- [10] Pain J-C, Croset P. Ideas and tools for error detection in opacity databases. *Atoms* 2023;11:27.
- [11] Pain J-C, Croset P. Checking the reliability of opacity databases. *Eur Phys J D* 2023;77:60.
- [12] Bhole G, Shukla A, Mahesh TS. Benford analysis: A useful paradigm for spectroscopic analysis. *Chem Phys Lett* 2015;639:36–40.
- [13] Bormashenko Ed, Shulzinger E, Whyman G, Bormashenko Ye. Benford's law, its applicability and breakdown in the IR spectra of polymers. *Phys A: Stat Mech Appl* 2016;444:524–9.
- [14] Kramida A, Ralchenko Yu, Reader J, NIST ASD Team. NIST Atomic spectra database (version 5.10). Gaithersburg, MD: National Institute of Standards and Technology; 2022, <http://dx.doi.org/10.18434/T4W30F>, [Online]. Available: <https://physics.nist.gov/asd>. [Mon Nov 21 2022].



- [15] Olsen K, Fontes CJ, Fryer CL, Hungerford AL, Wollaeger RT, Korobkin O, et al. NIST-LANL lanthanide opacity database (ver. 1.2). Gaithersburg, MD 20899: National Institute of Standards and Technology, <http://dx.doi.org/10.18434/mds2-2375>, [Online]. Available: <https://nslte.nist.gov/OPAC>. [Mon Apr 03 2023].
- [16] Fontes CJ, Fryer CL, Hungerford AL, Wollaeger RT, Korobkin O. A line-binned treatment of opacities for the spectra and light curves from neutron star mergers. *Mon Not R Astron Soc* 2020;493:4143–71.
- [17] Fontes CJ, Fryer CL, Wollaeger RT, Mumpower MR, Sprouse TM. Actinide opacities for modelling the spectra and light curves of kilonovae. *Mon Not R Astron Soc* 2023;519:2862–78.
- [18] Launay M. Le théorème du parapluie ou l'art d'observer le monde dans le bon sens. Flammarion; 2019, [in French].
- [19] Li F, Han S, Zhang H, Ding J, Zhang J, Wu J. Application of Benford's law in data analysis. *J Phys: Conf Ser* 2019;1168:032133.
- [20] Zhu WM, Wang H, Chen W. Research on fraud detection method based on Benford's law. *J Appl Stat Manage* 2007;1:41–6.
- [21] Torres J, Fernández S, Gamero A, Sola A. How do numbers begin? (The first digit law). *Eur J Phys* 2007;28:17–25.
- [22] Leemis LM, Schmeiser BW, Evans DL. Survival distributions satisfying Benford's law. *Am Stat* 2000;54:236–41.
- [23] Shao L, Ma BQ. Empirical mantissa distributions of pulsars. *Astrophys J* 2010;33:255–62.
- [24] Alexopoulos T, Leontsinis S. Benford's law in astronomy. *Astron Astrophys* 2014;35:639–48.
- [25] Liu YX, Wu XM, Zeng WY. Research on the comprehensive use of benford's law and panel model for detecting the quality of statistical data. *Stat Res* 2012;11:74–8.
- [26] Raimi RA. The first digit problem. *Am Math Mon* 1976;83:521–38.
- [27] Hill TP. The significant-digit phenomenon. *Am Math Mon* 1995;102:322–7.
- [28] Wang L, Ma B-Q. A concise proof of Benford's law. *Fundam Res* 2023. <http://dx.doi.org/10.1016/j.fmr.2023.01.002>, in press.
- [29] Pietronero L, Tossati E, Tossati V, Vespignani A. Explaining the uneven distribution of numbers in nature: The laws of Benford and Zipf. *Phys A* 2001;293:297–304.
- [30] Mehta ML. Random matrices and the statistical theory of energy levels. New York and London: Academic Press; 1967.
- [31] Izrailev FM. Simple models of quantum chaos: Spectrum and eigenfunctions. *Phys Rep* 1990;196:299–392.
- [32] Krivine H, Treiner J. Exercices et problèmes de physique statistique. De Boeck Supérieur; 2016, [in French].
- [33] Wilson BG, Rogers F, Iglesias C. Random-matrix method for the simulation of large atomic E1 transition arrays. *Phys Rev A* 1988;37:2695–7.
- [34] Cowan RD. The theory of atomic structure and spectra. Berkeley: University of California Press; 1981.
- [35] Drake GWF, editor. Atomic, molecular and optical physics handbook, chap. 21. NY: Woodbury; 1996.
- [36] Ralchenko Yu, Kramida A. Development of NIST atomic databases and online tools. *Atoms* 2020;8:56.
- [37] Hill TP. 2020. <https://arxiv.org/abs/2011.13015>.
- [38] Abbott BP, et al. GW170817: Observation of gravitational waves from a binary neutron star inspiral. *Phys Rev Lett* 2017;119:161101.
- [39] Frey LH, Even W, Whalen DJ, Fryer CL, Hungerford AL, Fontes CJ, et al. The Los Alamos supernova light-curve project: Computational methods. *Astrophys J Supp Ser* 2013;204:16.
- [40] The Opacity Project Team. The opacity project 1. Bristol, UK: Institute of Physics Publications; 1995.
- [41] Seaton MJ, Yu Y, Mihalas D, Pradhan AK. Opacities for stellar envelopes. *Mon Not R Astron Soc* 1994;266:805.
- [42] Badnell NR, Bautista MA, Butler K, Delahaye F, Mendoza C, Palmeri P, et al. Updated opacities from the opacity project. *Mon Not R Astron Soc* 2005;360:458–64.
- [43] Rogers FJ, Iglesias CA. Radiative atomic Rosseland mean opacity tables. *Astrophys J Supp Ser* 1992;79:507.
- [44] Iglesias CA, Rogers FJ. Updated OPAL opacities. *Astrophys J* 1996;464:943.
- [45] Colgan J, Kilcrease DP, Magee NH, Sherrill ME, Abdallah Jr J, Hakel P, et al. A new generation of Los Alamos opacity tables. *Astrophys J* 2016;817:116.
- [46] Kato D, Murakami I, Tanaka M, Banerjee S, Gaigalas G, Kitoviene L, et al. Japan-Lithuania opacity database for kilonova. 2021, <http://dpc.nifs.ac.jp/DB/Opacity-Database/>, (version 1.1); Tanaka M, Kato D, Gaigalas G, Kawaguchi K. Systematic opacity calculations for kilonovae. *Mon Not R Astron Soc* 2020;496:1369.
- [47] Ryabchikova T, Piskunov N, Kurucz RL, Stempels HC, Heiter U, Pakhomov Yu, et al. *Phys Scr* 2015;90:054005.
- [48] Ralchenko Yu, editor. Modern methods in collisional-radiative modeling of plasmas. Springer; 2016.