Physics in the World of Ideas: Complexity as Energy*

Yuri I. Manin

Max–Planck–Institut für Mathematik, Bonn, Germany

Introduction: Two Faces of Computation

Since the origins of civilisation, computation methods adopted in various cultures could be roughly subdivided into two types.

One relied upon a system of notation, initially only for numbers, and rules of performing arithmetical operations on these notations. I will refer to this type as the *linguistic* one. Gradually it developed into *algebra* where generic or specific names could be ascribed to other mathematical objects, and then to operations on them as well.

The second type of computation methods stressed their *manipulation with objects*, such as arithmetical computations using abacus, Chinese counting sticks, arithmometer, slide rule, geometric measurements etc.

The discovery of the abstract theory of computability in the 1930's followed this pattern as well: Alonzo Church and Andrey Markov used linguistic models whereas Alan Turing introduced his celebrated "machine". In the powerful image of Turing machine a computation is "a physical process performing linguistic tasks". Finally, the opposition *software/hardware* in applied computer science is the contemporary avatar of this duality.

Introduction of entropy in the information theory by C. Shannon in 1948 was probably the key moment in the development of *"physics of the world of ideas".*

However, at least before 1980's, theory of computability did not really use conceptual richness of theoretical physics: principle of minimal action, conservation laws, basic symmetry groups of fundamental physical models. For example, space–time of a Turing machine is "Galilean" one: space dimension is stretched along the tape whereas time dimension is embodied in the sequence of states and positions of the head.

With advent of the idea of quantum computing, the necessity of bridging world of physics and Platonic world of (un)computable became urgent. There were many attempts to do this: see a lively survey [A1], and [Man4], [Man5] motivated by path integration in quantum field theory.

However, most of these attempts were dedicated to "mental engineering": attempts to devise quantum (or mixed quantum/classical, cf. [Man3]) gadgets capable to solve, say, NP–complete problems in polynomial time. S. Aaronson in

* Keynote talk at the 2014 Interdisciplinary Symposium on Complex Systems (ISCS'14), University of Florence, September 15–18, 2014.

A. Sanayei et al. (eds.), *ISCS 2014: Interdisciplinary Symposium on Complex Systems*,
Emergence, Complexity and Computation 14, DOI: 10.1007/978-3-319-10759-2_1

[A1] eloquently puts forward the idea that $P \neq NP$ principle might be elevated to the status of an important law of physics, cf. also [A2].

In my work of the last years, I was trying to develop the beautiful idea of Andrey Kolmogorov who connected the physical idea of "randomness" of, say, a finite sequence of bits ("string") w, with the mathematical idea of "incompessibility" of this string. Incompressibility here means impossibility to get w as the output of a Kolmogorov optimal Turing machine taking as input a considerably shorter than w sequence of bits, whereas "randomness" means absence of all nontrivial regularities or correlations. See [ZvLe] for an early development of this idea.

As I will explain in this talk, the Kolmogorov complexity of a combinatorial object (and more generally, of a semicomputable function) has strong similarities to the physical notion of *energy level* of an isolated system. I will show that partition functions and quantum evolution operators involving Kolmogorov complexity throw new light upon such problems as computation of asymptotical bounds of error–correcting codes ([VlaNoTsfa], [ManMar1], [ManMar3]), and origins of Zipf's law ([Zi1], [Zi2], [Mand], [Ma1], [Ma2]).

1 Complexity as Energy

1.1 Energy, Partition Function and Hamiltonian Evolution

Here I briefly remind two basic contexts in which the notion of energy is used in theoretical physics.

(i) In statistical thermodynamics, we imagine that a system can occupy an enumerable set of states, with energy E_n at the state number $n = 1, 2, 3, \ldots$. Its interaction with environment is described via thermal contact; the temperature of the system is denoted T. Put $\beta := -1/kT$ where k is the Boltzman constant. The partition function of the system

$$Z := \sum_n e^{-\beta E_n} \tag{1.1}$$

defines the probability $p_n := e^{-\beta E_n}/Z$ for the system to occupy the n–the state. Using this probability distribution, one calculates mean values of various observables as functions of temperature, volume etc. Singularities of such mean values determine phase transitions.

(ii) In quantum mechanics, (E_n) appear as eigenvalues of the Hamiltonian operator H acting upon a Hilbert space with orthonormal basis $(\,|n\rangle)$ of eigenvectors of H. The Hamiltonian defines the time evolution which is given by the operator e^{iHt}. Formally replacing in H time t by the imaginary inverse temperature i/kT, we get an interpretation of the partition function (1.1) as "trace of of the quantum evolution operator in imaginary time".

In the recent paper [ManMar2], we suggested that this formal substitution may have a physical meaning in the "thermodynamics of the Universe", where one of the versions of global cosmological time is given by the inverse temperature of the cosmic microwave background (CMB) radiation, and transition of time (or temperature) to the imaginary axis may describe the statistical physics of the Big Bang.

Below we argue that the values of Kolmogorov complexity in the computability theory can be fruitfully interpreted as such "energy levels" E_n.

1.2 Constructive Worlds

Markov algorithms process finite strings in a given finite alphabet. Turing machines do the same, although usually the respective alphabet consists of bits $\{0,1\}$. Church's lambda–calculus processes certain *syntactically correct* finite strings. Partial recursive functions are certain "(semi)computable" partial functions $f: D(f) \to \mathbf{Z}_+^n$, with definition domains $D(f) \subset \mathbf{Z}_+^m$. Positive integers $\mathbf{Z}_+$ here are not supposed to be encoded in any specific way.

Proofs that any two of these (and other) methods calculate "the same" class of partial functions can be reduced to the following general scheme.

Define an (infinite) *constructive world* as a countable set X (usually of some finite Bourbaki structures, such as the set of all error–correcting codes in a fixed alphabet) given together with a set of *structural numberings:* bijections $\nu : \mathbf{Z}_+ \to X$. This set must have the following formal property: for any two numberings ν_1, ν_2 there exists a *total recursive permutation* $\sigma : \mathbf{Z}_+ \to \mathbf{Z}_+$ such that $\nu_2 = \nu_1 \circ \sigma$, and conversely, if ν is structural numbering, all $\nu \circ \sigma$ are also such numberings.

Informally, structural numberings must be algorithmically computable, together with their inverses.

A finite constructive world is any finite set.

Categorical Church–Turing thesis, Part I. Let X, Y be two infinite constructive worlds, $\nu_X : \mathbf{Z}_+ \to X$ $\nu_Y : \mathbf{Z}_+ \to X$ their structural numberings, and F an (intuitive) algorithm that takes as input an object $x \in X$ and produces an object $F(x) \in Y$ whenever x lies in the domain of definition of F; otherwise it outputs "NO" or works indefinitely.

Conversely, any algorithmically (semi) computable partial map $F : X \to Y$ is induced by a partial recursive function f in this way.

Then $f := \nu_Y^{-1} \circ F \circ \nu_X : \mathbf{Z}_+ \to \mathbf{Z}_+$ is a partial recursive function.

Categorical Church–Turing thesis, Part II. Let $\mathcal{C}$ be a category, whose objects are some infinite constructive worlds, and some finite constructive worlds of all finite cardinalities. Define the set of morphisms $\mathcal{C}(X,Y)$ to be the set of all partial maps that can be algorithmically (semi)computed in an intuitive sense.

Then $\mathcal{C}$ is equivalent to the category having one infinite object $\mathbf{Z}$, one finite object $\{1,\dots,n\}$ of each cardinality, and partial recursive functions as morphisms. If X is finite, then $\mathcal{C}(X,Y)$ consists of all partial maps.

One must keep in mind, that if X is infinite, then $\mathcal{C}(X,Y)$ has *no* natural structure of a constructive world. In particular, various numberings of partial recursive functions discussed in the literature are *not* structural numberings in our sense, they enumerate "programs" rather than functions themselves, and the binary relations “two programs calculate one and the same partial function” is undecidable.

1.3 Kolmogorov Complexity of Constructive Objects

Consider now an infinite a constructive world. For any (semi)–computable function $u:\ \mathbf{Z}_+ \to X$, the (exponential) complexity of an object $x \in X$ relative to u is

$$K_u(x) := \min\{m \in \mathbf{Z}_+ \,|\, u(m) = x\}.$$

If such m does not exist, we put $K_u(x) = \infty$.

Claim: there exists such u (“an optimal Kolmogorov numbering”, or “decompressor”) that for each other $v:\ \mathbf{Z}_+ \to X$, some constant $c_{u,v} > 0$, and all $x \in X$,

$$K_u(x) \leq c_{u,v} K_v(x).$$

This $K_u(x)$ is called *Kolmogorov complexity* of x.

A Kolmogorov order of a constructive world X is a bijection $\mathbf{K} = \mathbf{K}_u:\ X \to \mathbf{Z}$ arranging elements of X in the increasing order of their complexities K_u.

Notice that any optimal numbering is only partial function, and its definition domain is not decidable. Moreover, the Kolmogorov complexity K_u itself is *not computable*: it is the lower bound of a sequence of computable functions.

The same can be said about the Kolmogorov order. Moreover, on $\mathbf{Z}_+$ it cardinally differs from the natural order in the following sense: it puts in the initial segments very large numbers that can be at the same time Kolmogorov simple. For example, let $a_n := n^{n^{\cdot^{\cdot^{\cdot^{n}}}}}$ (n times). Then $K_u(a_n) \leq cn$ for some $c > 0$.

Finally, the indeterminacy of the complexity related to different choices of optimal functions u, v is multiplicatively $\exp(O(1))$. The same is true for the Kolmogorov order.

For a concise treatment of Kolmogorov complexity, see [Man1], pp. 226–231, and for a thorough one, see [LiVi]. Notice that in the literature one often uses the *logarithmic* Kolmogorov complexity which is defined as the length of the binary presentation of $K_u(x)$. It is interpreted as the length of the *maximally compressed description* of x. For our purposes, exponential version is more convenient, in particular, because as soon as an optimal programming method u is chosen, K_u allows us to define an unambiguous Kolmogorov order on $\mathbf{Z}_+$ or on any infinite constructive world given together with its structural numbering.

1.4 Fractality and Symmetries of the Kolmogorov Complexity

In [LiVi], pp. 103, 105, 178, one can find a schematic graph of logarithmic complexity of integers. The visible“continuity” of this graph reflects the fact that complexity of $n+1$ in any reasonable encoding is almost the same as complexity of n. The graph “most of the time” follows closely the graph of $\log_2 n$: as soon as an optimal family is chosen, the respective complexity differs from $\log_2 K_u(n)$ by a bounded function. However, infinitely often the graph drops down, lower than any given computable function of n. We will sometimes write K in place of K_u when change of optimal u is not essential.

Looking only at such graphs one does not see or suspect any kind of self–similarity. But it is there: if one restricts this graph onto any infinite decidable subset of $\mathbf{Z}_+$ in increasing order, one will get the same complexity relief as for the whole $\mathbf{Z}_+$: in fact, for any recursive bijection f of $\mathbf{Z}_+$ with a subset of $\mathbf{Z}_+$ we have $K(f(x)) = exp(O(1)) \cdot K(x)$.

The natural symmetry group of the logarithmic Kolmogorov complexity is the group S_∞^{comp} of total recursive permutations of $\mathbf{Z}_+$: for each such permutation σ, there exists a constant $c(\sigma)$ such that for all n we have

$$|\log_2 K(\sigma(n)) - \log_2 K(n)| < c(\sigma).$$

This symmetry reflects the essential independence of complexity on a chosen initial structural numbering. The group S_∞^{comp} is pretty mysterious one.

However, an inner conjugation inside the total permutation group of $\mathbf{Z}_+$ allows one to embed it into a less mysterious group of permutations S_∞^{lin} of $\mathbf{Z}_+$, consisting of all permutations which together with their inverses have no more than linear growth.

To achieve this, it suffices to replace the natural order of $\mathbf{Z}_+$ by its Kolmogorov order. This remark is essentially used in the proof of the Theorem 3.5 of the next section.

2 Probability Distributions on Constructive Worlds and Zipf's Law

2.1 Zipf's Law

Zipf ([Zi1], [Zi2]) studied statistics of usage of words in natural languages. His main empirical observation was this: if all words $\{w_k\}$ of a language in a representative corpus of texts are ranked according to decreasing frequency of their appearance, and the rank k and the number of the occurrences of w_k are plotted in the logarithmic scale, then the data approximately fit on a line. This means that the frequency of usage of the word of rank k is approximatetely C/k^α where C and α are some constants. Moreover, for natural languages α is close to one: see e. g. Fig. 1 in [Ma1] based upon a corpus containing $4 \cdot 10^7$ Russian words.

This general power law, and its particular $\alpha = 1$ realization turned out to be very universal, see [MurSo], [Pi]. It was observed in the texts of an extinct and not deciphered Meroitic language ([Sm]), and in the frequency of certain patterns in databases of financial reports of businesses: see [HuYeYaHua]. In the latter article, it was suggested that systemtic deviations from Zipf's law might be useful for fraud detection.

Zipf himself ([Zi2], [Zi1]) conjectured that his distribution "minimizes effort". Mandelbrot in [Mand] made this mathematically precise in the following way. If we postulate and denote by C_k a certain "cost" (of producing, using etc.) of the word of rank k, then the frequency distribution $p_k \sim 2^{-h^{-1}C_k}$ minimizes the ratio $h = C/H$, where $C := \sum_k p_k C_k$ is the average cost per word, and $H := -\sum_k p_k \log_2 p_k$ is the average entropy: see a discussion in [Ma2]. This produces a power law, if $C_k \sim \log k$. An additional problem, explanation of $\alpha = -1$, must be addressed separately. For one possibility, see [MurSo], sec. III.

In [Man7], I suggested a similar background but replaced logarithm in $C_k \sim \log k$ by the logarithmic Kolmogorov complexity. This means, in particular, that Zipf's rank corresponds to the ranking in order of growing complexity, and identifies complexity with Zipf's "effort". Underlying this metaphor is the image of brain/society dealing with compressed descriptions of units of information. Intuitively, whenever an individual mind, or a society, finds a compressed description of something, this something becomes usable, and is used more often than other "something" whose description length is longer. In experimental psychology, this corresponds to the *"availability bias"* studied by A. Tversky and D. Kahneman: as they write in [TvKa], this is *"a judgemental heuristic in which a person evaluates the frequency of classes or the probability of events by availability, i. e., by the ease with which relevant instances come to mind."*

For an expanded version of this metaphor applied to the history of science, see [Man6], and for related suggestions see [DeMe], [Del], [Ve].

In the context of [Man7], the $\alpha = -1$ enigma is solved by appealing to another mathematical discovery. Namely, L. Levin has established that the power law with $\alpha = -1$ appears as the *maximal* (up to a multiplicative constant) probability distribution that is *computable from below*: see [Lev1], [Lev2].

In all theoretical discussions, it is more or less implicitly assumed that empirically observed distributions concern fragments of a potential countable infinity of objects. I also postulate this, and work in a "constructive world".

2.2 Zipf's Law from Complexity

Here is the summary of my arguments. Zipf's law emerges as the combined effect of two factors:

(A) Zipf's rank ordering of a constructive world coincides with the ordering with respect to the growing (exponential) Kolmogorov complexity $K(w)$.

(B) The probability distribution producing Zipf's law (with exponent -1) is (an approximation to) the L. Levin maximal computable from below distribution: see [ZvLe], [Le] and [LiVi].

The $\alpha = -1$ power law follows from the fact that Levin's distribution assigns to an object w probability $\sim KP(w)^{-1}$ where KP is the exponentiated prefix Kolmogorov complexity, and we have, up to *exp* $(O(1))$–factors,

$$K(w) \preceq KP(w) \preceq K(w) \cdot \log^{1+\varepsilon} K(w)$$

with arbitrary $\varepsilon > 0$.

Discrepancy between the growth orders of K and KP is the reason why a probability distribution on infinity of objects cannot be constructed from K: the series $\sum_m K(m)^{-1}$ diverges. However, on finite sets of data this small discrepancy is additionally masked by the dependence of both K and KP on the choice of an optimal encoding. Therefore, when speaking about Zipf's Law, we will mostly disregard this difference. One could however argue that it is at least partly responsible for *"numerous minima and maxima in the error of [Zipf's] fit"*: see [Pi], Sec. 2.

3 Error–Correcting Codes and Their Asymptotic Bounds

3.1 Codes and Code Points

Fix an alphabet A which a finite set of cardinality $q \geq 2$. A *code* $C \subset A^n$ is a subset of words of length n. *Hamming distance* between two words of the same length is defined as

$$d((a_i),(b_i)) := \mathrm{card}\{i \in (1,\dots,n)\,|\,a_i \neq b_i\}.$$

Code parameters are the cardinality of the alphabet q and the numbers $n(C)$, $k(C), d(C)$ defined by:

$$n(C) := n, \quad k(C) := k := [\log_q \mathrm{card}(C)],$$

where $[x]$ is the maximal integer $\leq x$;

$$d(C) := d = \min\{d(a,b)\,|\,a,b \in C, a \neq b\}.$$

Briefly, C is an $[n,k,d]_q$–code. Its *code point* is the point

$$x(C) := \left(\frac{k(C)}{n(C)}, \frac{d(C)}{n(C)},\right) \in [0,1]^2$$

Coordinates of $x(C) = (R(C), \delta(C))$ are called *transmission rate* and *relative distance* respectively.

The idealized scheme of using error–correcting codes for information transmission can be described as follows. Some source data are encoded by a sequence of code words. After transmission through a noisy channel at the receiving end we will get a sequence of possibly corrupted words. If we know probability of corruption of a single letter, we can calculate, how many corrupted letters in a word we may allow for safe transmission; pairs of code words must be then separated by a larger Hamming distance. This necessity puts an upper bound on the achievable transmission rate.

A good code must maximize minimal relative distance when the transmission rate is chosen.

Our discussion up to now was restricted to *unstructured* codes: arbitrary subsets of words. Arguably, one more property of good codes is the existence of efficient algorithms of encoding and decoding. This can be achieved by introduction of *structured* codes. A typical choice is represented by *linear codes*: for them, A is a finite field of q elements, and C is a linear subspace of $\mathbf{F}_q^n$.

Now, let us call *the multiplicity* of a code point the number of codes that project onto it.

All codes (with fixed q), and all rational points in $[0,1]^2$ constitute examples of constructive worlds. The map $C \mapsto x(C)$ is a computable function. The same refers to structured codes considered in the literature, in particular, to linear codes.

Theorem 1. (Yu. M., 1981 + 2011). *There exists a continuous function $\alpha_q(\delta)$, $\delta \in [0,1]$, with the following properties:*

(i) The set of code points of infinite multiplicity is exactly the set of rational points $(R,\delta) \in [0,1]^2$ satisfying $R \le \alpha_q(\delta)$. The curve $R = \alpha_q(\delta)$ is called the asymptotic bound.

(ii) Code points x of finite multiplicity all lie strictly above the asymptotic bound and are called isolated ones: for each such point there is an open neighborhood containing x as the only code point.

(iii) The same statements are true for linear codes, with a possibly different asymptotic bound $R = \alpha_q^{lin}(\delta)$.

In a paper published in 1981, I proved that all limit points of the set of code points (q being fixed) lie under the graph of a monotone function α_q that was later called the asymptotic bound. Its characterisation via multiplicity used here was proved in [Man2].

3.2 Can One Compute an Asymptotic Bound?

During the thirty years since the discovery of the asymptotic bounds, many upper and lower estimates were established for them, especially for the linear case: see the monograph [VlaNoTsfa]. Upper bounds helped pinpoint a number of isolated codes.

However, the following most natural problems remain unsolved: – To find an explicit formula for α_q or α_q^{lin}.

– To find any single value of $\alpha_q(\delta)$ or $\alpha_q^{lin}(\delta)$ for $0 < \delta < 1 - q^{-1}$ (at the end segment $[1-q^{-1}, 1]$ these function vanish).

– To find any method of approximate computation of $\alpha_q(\delta)$ or $\alpha_q^{lin}(\delta)$.

– Clearly, $\alpha_q^{lin} \le \alpha_q$. Is this inequaliy *strict* somewhere?

3.3 A Brief Survey of Some Known Results

(i) One can count the number of codes of bounded block length n and plot their code points. The standard probabilistic methods then give the following *Gilbert–Varshamov bounds.*

Most unstructured q–ary codes lie lower or only slightly above the Hamming curve

$$R = 1 - H_q(\delta/2),$$

$$H_q(\delta) = \delta \log_q(q-1) - \delta \log_q \delta - (1-\delta)\log_q(1-\delta).$$

Most linear q–ary codes lie near or only slightly above the Gilbert–Varshamov bound

$$R = 1 - H_q(\delta).$$

In particular,

$$\alpha_q(R) \geq \alpha_q^{lin}(R) \geq 1 - H_q(\delta)$$

(ii) A useful combinatorial upper estimate is the Singleton bound:

$$R(C) + \delta(C) \leq 1 + \frac{1}{n(C)}.$$

Hence

$$\alpha_q(\delta) \leq 1 - \delta.$$

It follows that code points lying above this bound are isolated. The following Reed–Solomon (linear) codes $C \subset \mathbf{F}_q^n$ belong to this group.

Choose parameters $1 \leq k \leq n \leq q, d = n + 1 - k$. Choose pairwise distinct $x_1, \dots, x_n \in \mathbf{F}_q$, Embed the space of polynomials $f(x) \in \mathbf{F}_q[x]$ of degree $\leq k - 1$ into $\mathbf{F}_q^n$ by

$$f \mapsto (f(x_1), \dots, f(x_n)) \in \mathbf{F}_q^n.$$

After works of Goppa, this construction was generalized. Points $x_1, \dots, x_n \in \mathbf{F}_q$ were replaced by rational points of any smooth algebraic curve over $\mathbf{F}_q$, and polynomials by sections of an invertible sheaf. This allowed one to construct non–isolated linear codes lying partly strictly above the Gilbert–Varshamov bound.

This implies that we cannot "see" the asymptotic bound, plotting the set of (linear) code points of bounded size: we will see a cloud of points, whose upper bound concentrates near the Hamming or Varshamov–Gilbert bounds.

The proof of the following theorem given in [ManMar1] again uses reordering codes by their growing Kolmogorov complexity and basic properties of the Levin distribution (in this context the difference between K and KP becomes negligible). Imaginatively, one can say that the asymptotic bound becomes plottable with help of oracle–assisted computation: oracle should produce for us codes in their Kolmogorov order.

In order to state our theorem, notice that the function $\alpha_q(\delta)$ is continuous and strictly decreasing for $\delta \in [1, 1 - q^{-1})$. Hence the limit points domain $R \leq \alpha_q(\delta)$ can be equally well described by the inequality $\delta \leq \beta_q(R)$ where β_q is the function inverse to α_q.

Fix an $R \in \mathbf{Q} \cap (0, 1)$. For $\Delta \in \mathbf{Q} \cap (0, 1)$, put

$$Z(R, \Delta; \beta) := \sum_{C:\, R(C)=R,\, \Delta \leq \delta(C) \leq 1} K_u(C)^{-\beta + \delta(C) - 1},$$

where K_u is an (exponential) Kolmogorov complexity on the constructive world of all codes in a given alphabet of cardinality q.

Theorem 2. *(i) If $\Delta > \beta_q(R)$, then $Z(R, \Delta; \beta)$ is a real analytic function of β.*

(ii) If $\Delta < \beta_q(R)$, then $Z(R, \Delta; \beta)$ is a real analytic function of β for $\beta > \beta_q(R)$ such that its limit for $\beta - \beta_q(R) \to +0$ does not exist.

The following thermodynamical analogies justify our interpretation of the asymptotic bound as a phase transition curve in the "statistical physics of Platonian world." a) The argument β of the partition function corresponds to the inverse temperature.

b) The transmission rate R corresponds to the density ρ.

c) Our asymptotic bound transported into $(T = \beta^{-1}, R)$–plane as $T = \beta_q(R)^{-1}$ becomes the phase transition boundary in the *(temperature, density)*–plane.

For other mathematical contexts in which our partition function and other characteristics of codes appear, in particular, for relevant quantum mechanical analogies, see [ManMar3].

References

[A1] Aaronson, S.: NP–complete problems and physical reality, 23. arXiv:quant–ph/0502072

[A2] Aaronson, S.: Why philosophers should care about computational complexity, 58. arXiv:1108.1791

[DeMe] Dehaene, S., Mehler, J.: Cross–linguistic regularities in the frequency of number words. Cognition 43, 1–29 (1992)

[Del] Delahaye, J.–P.: Les entiers ne naissent pas égaux. Pour la Science (421), 80–85 (2012)

[HuYeYaHua] Huang, S.-M., Yen, D.C., Yang, L.-W., Hua, J.-S.: An investigation of Zipf's Law for fraud detection. Decision Support Systems 46, 70–83 (2008)

[Lev1] Levin, L.A.: Various measures of complexity for finite objects (axiomatic description). Soviet Math. Dokl. 17(2), 522–526 (1976)

[Lev2] Levin, L.A.: Randomness conservation inequalities: information and independence in mathematical theories. Information and Control 61, 15–37 (1984)

[LiVi] Li, M., Vitányi, P.: An introduction to Kolmogorov complexity and its applications. Springer (1993)

[Mand] Mandelbrot, B.: An information theory of the statistical structure of languages. In: Jackson, W. (ed.) Communication Theory, pp. 486–502. Butterworth, Woburn (1953)

[Ma1] Manin, D.Y.: Zipf's Law and Avoidance of Excessive Synonymy. Cognitive Science 32(7), 1075–1078 (2008), arXiv:0710.0105

[Ma2] Manin, D.Y.: Mandelbrot's model for Zipf's Law. Can Mandelbrot's model explain Zipf's Law for language? Journ of Quantitative Linguistics 16(3), 274–285 (2009)

[Man1] Manin, Y.I.: A Course in Mathematical Logic for Mathematicians, 2nd edn. Graduate Texts in Mathematics. Springer (2010)

[Man2] Manin, Y.I.: A computability challenge: Asymptotic bounds and isolated error-correcting codes. In: Dinneen, M.J., Khoussainov, B., Nies, A. (eds.) WTCS 2012 (Calude Festschrift). LNCS, vol. 7160, pp. 174–182. Springer, Heidelberg (2012)

[Man3] Manin, Y.I.: Classical computing, quantum computing, and Shor's factoring algorithm, Séminaire Bourbaki, no. 862, vol. 266, pp. 375–404. Astérisque (June 1999/2000), arXiv: quant-ph/9903008

[Man4] Manin, Y.I.: Renormalization and computation I. Motivation and background. In: Loday, J., Vallette, B. (eds.) Proceedings OPERADS 2009. Séminaires et Congrès, vol. 26, pp. 181–223. Soc. Math. de France (2012), math.QA/0904.492

[Man5] Manin, Y.I.: Renormalization and computation II: Time cut–off and the Halting Problem. Math. Struct. in Comp. Science 22, 729–751 (2012), Special issue, Cambridge UP. math.QA/0908.3430

[Man6] Manin, Y.I.: Kolmogorov complexity as a hidden factor of scientific discourse: From Newton's law to data mining. In: Talk at the Plenary Session of the Pontifical Academy of Sciences on Complexity and Analogy in Science: Theoretical, Methodological and Epistemological Aspects, Vatican, November 5–7 (2012) arXiv:1301.0081

[Man7] Manin, Y.I.: Zipf's law and L. Levin's probability distributions. Functional Analysis and its Applications, Vol. 48(2) (2014), arXiv:1301.0427

[ManMar1] Manin, Y., Marcoll, M.: Kolmogorov complexity and the asymptotic bound for error-correcting codes. Journ. of Diff. Geometry (to appear in 2014), arXiv:1203.0653

[ManMar2] Manin, Y., Marcolli, M.: Big Bang, Blow Up, and Modular Curves: Algebraic Geometry in Cosmology. arXiv:1402.2158

[ManMar3] Manin, Y., Marcolli, M.: Error–correcting codes and phase transitions. Math. in Comp. Science 5, 133–170 (2011), arXiv:mat.QA/0910.5135

[MurSo] Murtra, B.C., Solé, R.: On the Universality of Zipf's Law. Phys. Rev. E 82 (2010)

[Pi] Piantadosi, S.T.: Zipf's word frequency law in natural language: A critical review and future directions, `http://colala.bcs.rochester.edu/papers/piantadosi2014zipfs.pdf`

[Sm] Smith, R.: Investigation of the Zipf–plot of the extinct Meroitic language. Glottometrics 15, 53–61 (2007)

[TvKa] Tversky, A., Kahneman, D.: Availability; A Heuristic for Judging Frequency and Probability. Cognitive Psychology 5, 207–232 (1973)

[Ve] Todd, L.: Veldhuizen. Software Libraries and Their Reuse: Entropy, Kolmogorov Complexity, and Zipf's Law. arXiv:cs/0508023

[VlaNoTsfa] Vladut, S.G., Nogin, D.Y., Tsfasman, M.A.: Algebraic geometric codes: Basic notions. Mathematical Surveys and Monographs. American Mathematical Society, Providence (2007)

[Zi1] Zipf, G.K.: The psycho–biology of language. Routledge, London (1936)

[Zi2] Zipf, G.K.: Human behavior and the principle of least effort. Addison–Wesley (1949)

[ZvLe] Zvonkin, A.K., Levin, L.A.: The complexity of finite objects and the basing of the concepts of information and randomness on the theory of algorithms. Uspehi Mat. Nauk 25(6(156)), 8–127 (1970), (Russian)