Assembly Theory of Binary Messages

Szymon Łukaszyk * and Wawrzyniec Bieniawski

Łukaszyk Patent Attorneys, ul. Głowackiego 8, 40-052 Katowice, Poland
* Correspondence: szymon@patent.pl

Abstract: Using assembly theory, we investigate the assembly pathways of binary strings of length N formed by joining bits present in the assembly pool and the strings that entered the pool as a result of previous joining operations. We show that the string assembly index is bounded from below by the shortest addition chain for N, and we conjecture about the form of the upper bound. We define the degree of causation for the minimum assembly index and show that for certain N it features regularities that can be used to determine a shortest addition chain. We show that a string with the smallest assembly index for N can be assembled by a binary program of length equal to this index if the length of this string is expressible as a product of Fibonacci numbers. We conjecture that there is no binary program that has a length shorter than the length of the string having the largest assembly index for N that could assemble this string. The results confirm that four Planck areas provide a minimum information capacity that corresponds to a minimum thermodynamic (Bekenstein-Hawking) entropy. Knowing that the problem of determining the assembly index is at least NP-complete, we conjecture that this problem is NP-complete, while the problem of creating the string so that it would have a predetermined largest assembly index is NP-hard. The proof of this conjecture would imply P ≠ NP, since every computable problem and every computable solution can be encoded as a finite binary string.

Keywords: assembly theory; emergent dimensionality; shortest addition chains; P versus NP problem; mathematical physics

1. Introduction

Assembly Theory (AT) [1–7] provides a distinctive complexity measure, superior to established complexity measures used in information theory, such as Shannon entropy or Kolmogorov complexity [1,5]. AT does not alter the fundamental laws of physics [6]. Instead, it redefines *objects* on which these laws operate. In AT, *objects* are not considered sets of point *particles* (as in most physics), but instead are defined by the histories of their formation (assembly pathways) as an intrinsic property, where, in general, there are multiple assembly pathways to create a given *object*. Therefore, we use the emphasis for *object* as this term, understood as a collection of *matter*, is a misnomer, as it neglects the (quantum) nonlocality [8]. The nonlocality is independent of the entanglement among *particles* [9], as well as the quantum contextuality [10], and increases as the number of *particles* [11] grows [12,13]. Furthermore, the ugly duckling theorem [14,15] asserts that every two *objects* we perceive are equally similar (or equally dissimilar).

AT explains and quantifies selection and evolution, capturing the amount of memory necessary to produce a given *object* [6] (this memory *is* the *object* [16]). This is because the more complex a given *object* is, the less likely an identical copy can be observed without the selection of some information-driven mechanism that generated that *object*. Formalizing assembly pathways as sequences of joining operations, AT begins with basic units (such as chemical bonds) and ends with a final *object*. This conceptual shift captures evidence of selection in *objects* [1,2,6].

The assembly index of an *object* corresponds to the smallest number of steps required to assemble this *object*, and - in general - increases with the *object's* size, but decreases with symmetry, so large *objects* with repeating substructures may have a smaller assembly index than smaller *objects* with greater heterogeneity [1]. The copy number specifies the observed number of copies of an *object*. Only these two quantities describe the evolutionary concept of selection by showing how many alternatives were excluded to assemble a given *object* [6,16].

AT has been experimentally confirmed in the case of molecules and has been probed directly experimentally with high accuracy with spectroscopy techniques, including mass spectroscopy, IR, and NMR spectroscopy [6]. It is a versatile concept with applications in various domains. Beyond its application in the field of biology and chemistry [7], its adaptability to different data structures, such as text, graphs, groups, music notations, image files, compression algorithms, human languages, memes, etc., showcases its potential in diverse fields [2].

In this study, we investigate the assembly pathways of binary strings by joining individual bits present in the assembly pool [6] and strings that entered the pool as a result of previous joining operations. It is bit that is the smallest amount and the quantum of information.

> In particular, we investigate the assembly of black-body objects (BBs: black holes (BHs), white dwarfs, and neutron stars) considered binary strings [17–19]. It is known [2,17–28] that information in the universe evolves toward increased structural complexity, decreasing information entropy.
>
> This study extends the findings of previous research [11,17–19] within the framework of AT and emergent dimensionality [8,11,17–19,24,26,27,29]. However, our results generally apply to information theory. Therefore, we put the BB-related content in frames like this one. The reader not interested in BBs may skip the text in these frames.

The paper is structured as follows. Section 2 introduces basic concepts and definitions used in the paper. Section 3 shows that the assembly index of binary strings is bounded from below and provides the form of this bound. Section 4 defines the degree of causation for strings with the smallest assembly index. Section 5 shows that the assembly index of binary strings is bounded from above and conjectures about the exact form of this bound. Section 6 introduces the concept of a binary assembling program and shows that the trivial assembling program assembles binary strings that have the smallest assembly index. Section 7 discusses and Section 8 concludes the findings of this study.

2. Preliminaries

For K subunits of an *object O* the assembly index a_O of this *object* is bounded [1] from below by

$$\min(a_O) = \log_2(K), \tag{1}$$

and from above by

$$\max(a_O) = K - 1, \tag{2}$$

The lower bound (1) represents the fact that the simplest way to increase the size of a subunit in a pathway is to take the largest subunit assembled so far and join it to itself [1] and, in the case of the upper bound (2), subunits must be distinct so that they cannot be reused from the pool, decreasing the index.

Here, we consider binary strings $C_k^{(N)}$ containing bits $\{1,0\}$, with N_0 zeros and N_1 ones, having length $N = N_0 + N_1$. N_1 is called the binary Hamming weight or bit summation of a binary string. Binary strings are our basis AT *objects* [2] and we consider the process of their formation within the AT framework. Where the bit value can be either 1 or 0, we write $* = \{1,0\}$ with $*$ being the same within the string $C_k^{(N)}$. If we allow for the 2nd possibility that can be the same as or different from $*$, we write $\star = \{1,0\}$. Thus, $C_k^{(2)} = [**]$, for example, is a placeholder for all four 2-bit strings.

We consider strings $C_k^{(N)}$ to be *messages* transmitted through a communication channel between a source and a receiver, similarly to the Claude Shannon approach [30] used in the derivation of binary information entropy

$$H\left(C_k^{(N)}\right) = -p_0 \log_2(p_0) - p_1 \log_2(p_1), \tag{3}$$

where

$$p_0 = \frac{N_0}{N} \quad \text{and} \quad p_1 = \frac{N_1}{N}, \tag{4}$$

are the ratios of occurrences of zeros and ones within the string $C_k^{(N)}$ and the unit of entropy (3) is bit.

Definition 1. *A string assembly index $a^{(N)}$ is the smallest number of steps s required to assemble a binary string $C_k^{(N)}$ of length N by joining two distinct bits contained in the initial assembly pool $P = \{1, 0\}$ and strings assembled in previous steps that were added to the assembly pool. Therefore, the assembly index $a^{(N)}(C_k)$ is a function of the string $C_k^{(N)}$.*

For example, the 8-bit string

$$C_k^{(8)} = [01010101] \tag{5}$$

can be assembled in at most seven steps:

1. join 0 with 1 to form $C_k^{(2)} = [01]$, adding $[01]$ to $P = \{1, 0, 01\}$,
2. join $C_k^{(2)} = [01]$ with 0 to form $C_k^{(3)} = [010]$, adding $[010]$ to $P = \{1, 0, 01, 010\}$,
3. ...
7. join $C_k^{(7)} = [0101010]$ with 1 to form $C_k^{(8)} = [01010101]$

(i.e. not using the assembly pool P), six, five, or four steps:

1. join 0 with 1 to form $C_k^{(2)} = [01]$, adding $[01]$ to P,
2. join $C_k^{(2)} = [01]$ with $[01]$ taken from P to form $C_k^{(4)} = [0101]$, adding $[0101]$ to P,
3. join $C_k^{(4)} = [0101]$ with $[01]$ taken from P to form $C_k^{(6)} = [010101]$, adding $[010101]$ to P,
4. join $C_k^{(6)} = [010101]$ with $[01]$ taken from P to form $C_k^{(8)} = [01010101]$,

or at least three steps:

1. join 0 with 1 to form $C_k^{(2)} = [01]$, adding $[01]$ to P,
2. join $C_k^{(2)} = [01]$ with $[01]$ taken from P to form $C_k^{(4)} = [0101]$, adding $[0101]$ to P,
3. join $C_k^{(4)} = [0101]$ with $[0101]$ taken from P to form $C_k^{(8)} = [01010101]$,

while the 8-bit string

$$C_l^{(8)} = [00010111] \tag{6}$$

can be assembled in at least six steps:

1. join 0 with 1 to form $C_l^{(2)} = [01]$, adding $[01]$ to P,
2. join $C_l^{(2)} = [01]$ with $[01]$ taken from P to form $C_l^{(4)} = [0101]$, adding $[0101]$ to P,
3. join 0 with 0 adding $[00]$ to P,
4. join $C_l^{(4)} = [0101]$ with $[00]$ taken from P to form $C_l^{(6)} = [000101]$, adding $[000101]$ to P,
5. join $C_l^{(6)} = [000101]$ with 1 to form $C_l^{(7)} = [0001011]$, adding $[0001011]$ to P,
6. join $C_l^{(7)} = [0001011]$ with 1 to form $C_l^{(8)} = [00010111]$,

as only the doublet $[01]$ can be reused from the pool. Therefore, strings (5) and (6), despite having the same length $N = 8$, Hamming weight $N_1 = 4$, and Shannon entropy (3), have respective assembly indices $a^{(8)}(C_k) = 3$ and $a^{(8)}(C_l) = 6$ that represent the lengths of their shortest assembly pathways, which in turn ensures that their assembly pools P are distinct sets for a given assembly pathway.

Tables 1 and A6–A13 (Appendix D) show the distributions of the assembly indices among 2^N strings for $4 \leq N \leq 12$ taking into account the number of ones N_1. The sums of each column form Pascal's triangle read by rows (OEIS sequence A007318).

Table 1. Distribution of the assembly indices for $N = 4$.

$a^{(4)}(C)$	$\lvert a^{(4)}(C)\rvert$	N_1				
		0	1	2	3	4
2	4	1		2		1
3	12		4	4	4	
	16	1	4	$\lvert B^{(4)}\rvert = 6$	4	1

The following definition is commonly known, but we provide it here for clarity.

Definition 2. *A string $B_k^{(N)}$ is a balanced string if its Hamming weight $N_1 = \lfloor N/2 \rfloor$ or $N_1 = \lceil N/2 \rceil$.*

Without loss of generality, we shall assume that if N is odd, $N_1 < N_0$ (e.g., for $N = 5$, $N_1 = 2$, and $N_0 = 3$). However, our results are equivalently applicable if we assume the opposite (i.e. a larger number of ones for an odd N). The number $\lvert B^{(N)}\rvert$ of balanced strings among all 2^N strings is[1]

$$\lvert B^{(N)}\rvert = \binom{N}{\lfloor N/2 \rfloor} = \binom{N}{\lceil N/2 \rceil} \approx \sqrt{\frac{2}{\pi N}} 2^N. \tag{7}$$

This is the OEIS A001405 sequence, the maximal number of subsets of an N-set such that no one contains another, as asserted by Sperner's theorem, and approximated using Stirling's approximation for large N. Balanced and even length strings $B_k^{(N)}$ have natural binary entropies (3) $H(B_k^{(N)}) = \{0, 1\}$. Conversely, non-balanced and/or odd-length strings $C_k^{(N)}$ have binary entropies $0 < H(C_k^{(N)}) < 1$.

BBs emit Hawking black-body radiation having a continuous spectrum that depends only on one factor, the BB temperature $\lvert T_{\text{BB}}\rvert = T_{\text{P}}/(2\pi d_{\text{BB}})$ corresponding to the BB diameter $D_{\text{BB}} := d_{\text{BB}}\ell_{\text{P}}, d_{\text{BB}} \in \mathbb{R}$, where ℓ_{P} and T_{P} denotes respectively the Planck length and temperature [17]. Triangulated BB surfaces contain a balanced number of Planck area triangles, each having binary potential $\delta\varphi = -c^2 \cdot \{0, 1\}$, where c denotes speed of light in vacuum, as has been shown [17,19], based on the Bekenstein-Hawking entropy [31–33] $S_{\text{BB}} = k_{\text{B}} N_{\text{BB}}/4$. Here k_{B} is the Boltzmann constant and $N_{\text{BB}} := \pi D_{\text{BB}}^2/\ell_{\text{P}}^2 = \pi d_{\text{BB}}^2$ is the information capacity of the BB surface, i.e., the $\lfloor N_{\text{BB}} \rfloor \in \mathbb{N}$ Planck triangles corresponding to bits of information [17–19,32,34,35], and the fractional part triangle(s) having the area $\{N_{\text{BB}}\}\ell_{\text{P}}^2 = (N_{\text{BB}} - \lfloor N_{\text{BB}} \rfloor)\ell_{\text{P}}^2$ too small to carry a single bit of information [17,18].

Therefore, a balanced string B_k represents a BB surface comprising $N_1 = \lfloor N_{\text{BB}} \rfloor/2$ active Planck triangles (APTs) with binary potential $\varphi_{\text{APT}} = -c^2$ and energy $E_{\text{BB}} = \pm M_{\text{BB}} c^2$ [18].

Theorem 1. *A string having length $N = 4$ is the shortest string having more than one string assembly index 1.*

Proof. The proof is trivial. For $N = 1$ the assembly index $a^{(1)}(C) = 0$, as all basis *objects* have a pathway assembly index of 0 [2] (they are not *assembled*). $N = 2$ provides four available strings with $a^{(2)}(C) = 1$. $N = 3$ provides eight available strings with $a^{(3)}(C) = 2$. Only $N = 4$ provides 16 strings that include four stings with $a^{(4)}(C) = 2$ and twelve strings with $a^{(4)}(C) = 3$ including $\lvert B^{(4)}\rvert = 6$ balanced strings, as shown in Tables 1 and 2. For example, to assemble the string $B_1 = [0101]$, we need to assemble the string [01] and reuse it. Therefore, $a^{(N)}(C_k) = N - 1$ for $0 < N < 4$, $\forall k = \{1, 2, \dots, 2^N\}$ and $\min_k\left(\{a^{(N)}(C_k)\}\right) < N - 1$ for $N \geq 4$, where $\{a^{(N)}(C_k)\}$ denotes a set of assembly indices of all $\lvert 2^N \rvert$ strings. $\square$

---

[1] "$\lfloor x \rfloor$" is the floor function that yields the greatest integer less than or equal to x and "$\lceil x \rceil$" is the ceiling function that yields the least integer greater than or equal to x.

Table 2. $|B^{(4)}| = 6$ balanced strings $B_k^{(4)}$.

k	$B_k^{(4)}$				$a^{(4)}(B_k)$
1	(0	1)	(0	1)	2
2	(1	0)	(1	0)	2
3	0	1	1	0	3
4	1	1	0	0	3
5	1	0	0	1	3
6	0	0	1	1	3

Interestingly, Theorem 1 strengthens the meaning of $N = 4$ as the minimum information capacity that provides a minimum thermodynamic black hole entropy [31–33]. There is no *disorder* or *uncertainty* in an *object* that can be assembled in the same number of steps $s \leq 2$.

The following definition, taking into account the cyclic order of binary strings, is also provided for the sake of clarity.

Definition 3. *A string $R_k^{(N)}$ is a ringed string if a ring formed with this string by joining its beginning with its end is unique among the rings formed from the other ringed strings $R_l^{(N)}, l \neq k$.*

There are at least two and at most N forms of a ringed string $R_k^{(N)}$ that differ in the position of the starting bit. For example for $|B^{(4)}| = 6$ balanced strings, shown in Table 2, two augmented strings with $a^{(4)} = 2$ correspond to each other if we change the starting bit

$$[\ldots 1 \mid 0101 \mid 0101 \mid 01 \ldots] =$$
$$[\ldots 10 \mid 1010 \mid 1010 \mid 1 \ldots]. \tag{8}$$

Similarly, four augmented strings with $a^{(4)} = 3$ correspond to each other

$$[\ldots \mid 0110 \mid 0110 \mid 011 \ldots] =$$
$$[\ldots 0 \mid 1100 \mid 1100 \mid 11 \ldots] =$$
$$[\ldots 01 \mid 1001 \mid 1001 \mid 1 \ldots] =$$
$$[\ldots 011 \mid 0011 \mid 0011 \mid \ldots], \tag{9}$$

after a change in the position of the starting bit. Thus, there are only two balanced ringed strings $E_k^{(4)}$.

The number of ringed strings $|R^{(N)}|$ among all 2^N strings is given by the OEIS sequence A000031. In general (for $N \geq 3$), the number $|R^{(N)}|$ of ringed strings is much lower than the number $|B^{(N)}|$ of balanced strings.

> As asserted by the no-hair theorem [36], BH is characterized only by three parameters: mass, electric charge, and angular momentum.
>
> However, BHs are fundamentally uncharged and non-rotating, since the parameters of any conceivable BH, that is, charged (Reissner-Nordström), rotating (Kerr) and charged rotating (Kerr-Newman), can be arbitrarily altered, provided that the area of a BH surface does not decrease [37] using Penrose processes [38,39] to extract electrostatic and/or rotational energy of a BH [40].
>
> Thus, a BH is defined by a single real number, and no Planck triangle is distinct on a BH surface. We can define neither a beginning nor an end of a balanced ringed string $E_k^{(N_{BH})}$ that represents a given BH.

By neglecting the notion of the beginning and end of a string, we focus on its length and content. In Yoda's language,

"complete, no matter where it begins. A message is".

The numbers of the balanced $|B_k|$, ringed $|R_k|$, and balanced ringed[2] $|E_k|$ strings are shown in Table 3 and Figure 1. The formula for $|E_k|$ remains to be researched.

Table 3. String length N, number of all strings 2^N, number of balanced strings $B^{(N)}$, number of ringed strings $R^{(N)}$, and number of balanced ringed strings $E^{(N)}$.

N	2^N	$|B^{(N)}|$	$|R^{(N)}|$	$|E^{(N)}|$	$|B^{(N)}|/|E^{(N)}|$
1	2	1	2	1	1
2	4	2	3	1	2
3	8	3	4	1	3
4	16	6	6	2	3
5	32	10	8	2	5
6	64	20	14	4	5
7	128	35	20	5	7
8	256	70	36	10	7
9	512	126	60	14	9
10	1024	252	108	26	9.6923...
11	2048	462	188	42	11
12	4096	924	352	80	11.55
13	8192	1716	632	132	13
14	16384	3432	1182	246	13.9512...
15	32768	6435	2192	429	15
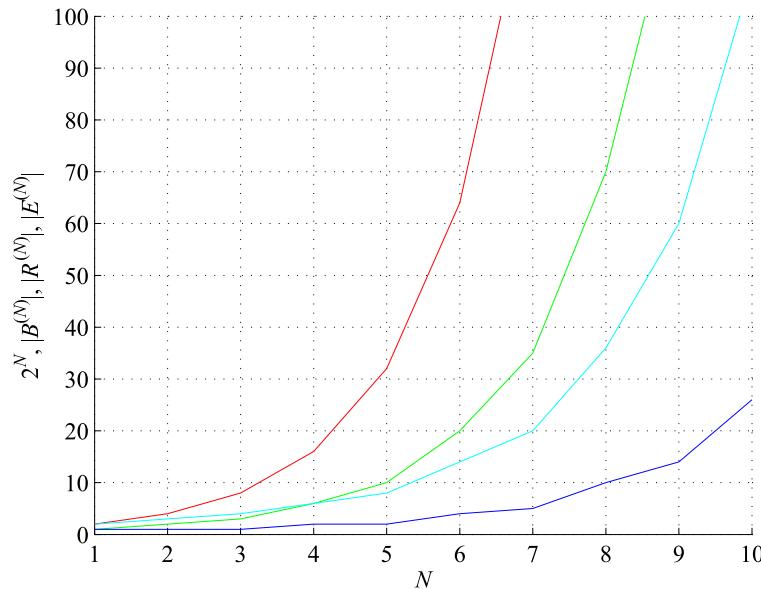
Figure 1. Numbers of all 2^N strings (red), balanced strings $|B^{(N)}|$ (green), ringed strings $|R^{(N)}|$ (cyan), and balanced ringed strings $|E^{(N)}|$ (blue) as a function of the string length N.

We note that, in general, the starting bit is relevant for the assembly index. Thus, different forms of a ringed string may have different assembly indices. For example, for $N = 7$ balanced strings B_{34} and B_{35}, shown in Table A16 have $a^{(7)} = 6$. However, these strings are not ringed, since they correspond to each other and to the balanced strings B_{13}, B_{18}, B_{20}, B_{28}, and B_{30} with $a^{(7)} = 5$. They all have the same triplet of adjoining ones.

---

2    $|E_k|$ is close to OEIS A000014 up to the eleventh term.

Definition 4. *The assembly index of a ringed string $R_k^{(N)}$ is the smallest assembly index among all forms of this string.*

Thus, if different forms of a ringed string have different assembly indices, we assign the smallest assembly index to this string. In other words, we assume that the smallest number of steps

$$a^{(N)}(R_k) = \min_l \left(\{a^{(N)}(R_k)_l\} \right), \quad (R_k)_l \in R_k, \tag{10}$$

where $(R_k)_l$ denotes a particular l^{th} form of a ringed string R_k, is the string assembly index of this ringed string. We assume that if an *object* that can be represented by a ringed string can be assembled in fewer steps, this procedure will be preferred by nature.

The distribution of the assembly indices of the balanced ringed strings E_k is shown in Table 4.

Table 4. Distribution of assembly indices among balanced ringed strings $E^{(N)}$ for $4 \leq N \leq 11$.

N	$|E^{(N)}|$	$a^{(N)} = 2$	$a^{(N)} = 3$	$a^{(N)} = 4$	$a^{(N)} = 5$	$a^{(N)} = 6$	$a^{(N)} = 7$	$a^{(N)} = 8$
4	2	1	1					
5	2		1	1				
6	4		1	2	1			
7	5			2	3			
8	10		1	1	6	2		
9	14			1	4	7	2	
10	26			1	6	9	10	
11	42				2	14	20	6

3. Minimum Assembly Index

In the following, we derive the tight lower bound of the set of different string assembly indices 1.

Theorem 2 (Tight lower bound on the string assembly index). *The smallest string assembly index $a^{(N)}(C_{min})$ as a function of N corresponds to the shortest addition chain for N (OEIS sequence A003313).*

Proof. Strings C_{min} for which $a^{(N)}(C_{min}) = \min_k \left(\{a^{(N)}(C_k)\} \right), \forall k = \{1, 2, \ldots, 2^N\}$ can be formed in subsequent steps s by joining the longest string assembled so far with itself until $N = 2^s$ is reached [1]. Therefore, if $N = 2^s$, then $\min_k \left(\{a_{(2^s)}(C_k)\} \right) = s = \log_2(N)$. Only four strings

$$C_{\text{min}_1}^{(2^s)} = [00\ldots], \quad C_{\text{min}_2}^{(2^s)} = [11\ldots], \quad C_{\text{min}_3}^{(2^s)} = [0101\ldots], \quad \text{and} \quad C_{\text{min}_4}^{(2^s)} = [1010\ldots] \tag{11}$$

have such an assembly index in this case.

An addition chain for $N \in \mathbb{N}$ having the shortest length $s \in \mathbb{N}$ (commonly denoted as $l(N)$) is defined as a sequence $1 = b_0 < b_1 < \cdots < b_s = N$ of integers such that for each $j \geq 1$, $b_j = b_k + b_l$ for $l \leq k < j$. The first step in creating an addition chain for N is always $b_1 = 1 + 1 = 2$ and this corresponds to assembling a doublet $[**]$ from the initial assembly pool P. Thus, the lower bound for s of the addition chain for N, $s \geq \log_2(N)$ is achieved for $N = 2^s$. In our case, this bound is achieved by the strings (11). The second step in creating an addition chain can be $b_2 = 1 + 1 = 2$ or $b_2 = 1 + 2 = 3$.

Thus, finding the shortest addition chain for N corresponds to finding an assembly index of a string containing bits and/or doublets and/or triplets generated by these doublets for $N \neq 2^s$ since due to Theorem 1 only they provide the same assembly indices $\{0, 1, 2\}$. Such strings correspond to linear molecules made of carbons Supplementary Materials, S3.2 in [4]. □

The smallest assembly indices $a_{\text{min}}^{(N)}$ are shown in Table 5 for $1 \leq N \leq 21$. Calculating the minimum length of the addition chain for N, as well as finding the shortest assembly pathway for a chemical molecule, have been shown to be at least as hard as NP-complete [4,41].

Table 5. The lower bound on the binary string assembly index (OEIS A003313).

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
$a_{min}^{(N)}$	0	1	2	2	3	3	4	3	4	4	5	4	5	5	5	4	5	5	6	5	6

Interestingly, if a binary string $C^{(N)}$ were to encode four DNA/RNA nucleobases (for example as, A= 00, C= 01, G= 10, and T= 11), then strings with the smallest assembly indices (as well as strings generated by trivial assembly programs Q according to Definition 6) would not encode all nucleobases. For example, the string $C_{min}^{(10)} = [1001010010]$ with $a_{min}^{(10)} = 4$ does not encode T= 11.

4. Degree of Causation for Minimum Asembly Index Strigs

Using the difference between the general AT lower bound (1) and the smallest assembly index (OEIS A003313) we can define the quantity

$$D_C(N) := 2^{\left(\log_2(N)-a_{min}^{(N)}\right)} = N2^{-a_{min}^{(N)}}, \tag{12}$$

capturing a degree of causation [6] of assembling the strings of length N with the smallest assembly index, as shown in Figure 2. For $N = 2^s$, the degree of causation $D_C(N) = 1$, as all strings (11) can be assembled along a single pathway only; their assembly is entirely causal. However, for $N \neq 2^s$, $D_C(N) < 1$, since some strings $C_{min}^{(N)}$ can be assembled along different pathways. For example, there are two pathways for the string $[001]$: (a) $[00] + 1$ and (b) $0 + [01]$ leaving different subunits ($[00]$ and $[01]$) in their assembly pools and resulting in lower values of $D_C(N)$.
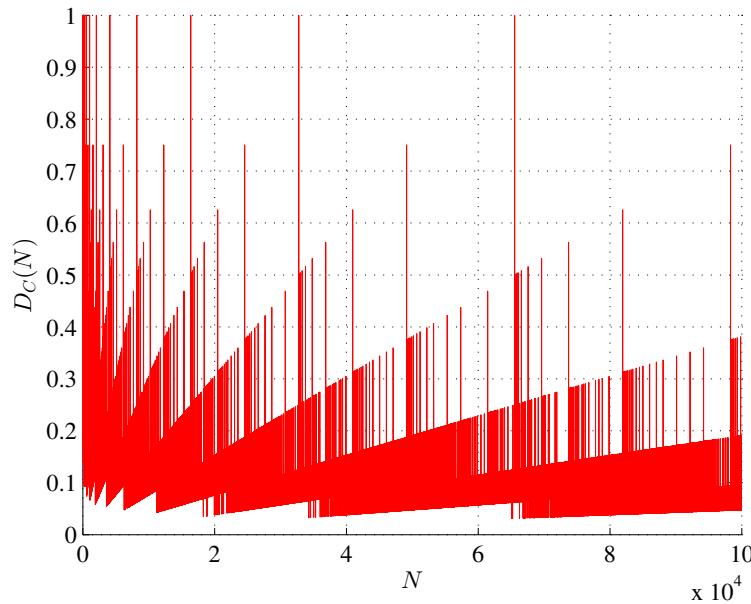
Figure 2. Degree of causation as a function of $1 \leq N \leq 10^5$.

Equation (12) features regularities that for certain values of N can be used to determine the smallest assembly index (i.e. the shortest addition chain for N) as $a_{min}^{(N)} = \log_2(N) - \log_2(D_C(N))$. For example, for each $N = 2^s, s \in \mathbb{N} \Leftrightarrow a_{min}^{(2^s)} = s$ and for each $\hat{N}$ being the sum of two powers of 2 (OEIS A048645)

$$\hat{N} := 2^s + 1 \cdot 2^l, \quad l = 0, 1, \ldots, s-1 \quad \Leftrightarrow \quad a_{min}^{(\hat{N})} = s+1, \quad \lim_{s \to \infty} \left(\min D_C(\hat{N})\right) = \frac{1}{2}, \tag{13}$$

while for the remaining $\tilde{N}$ not being the sum of two powers of 2 (OEIS A072823)

$$2^s < \tilde{N} < 2^{s+1}, \quad \tilde{N} \neq \hat{N} \quad \Leftrightarrow \quad a_{\min}^{(\tilde{N})} = s + k, \; k \geq 2, \quad \lim_{s \to \infty} \left(\max D_C(\tilde{N}) \right) = \frac{15}{32}, \tag{14}$$

where $k = 2$ for $\tilde{N} = \{7, 11, 13 - 15, 19, 21 - 23, 25 - 28, \dots\}$, while some $\tilde{N}$'s are exceptions to this general rule. For example, $k = 3$ for $\tilde{N} = \{29, 31, 47, 53, 55, 57 - 59, 61 - 63, \dots\}$, $k = 4$ for $\tilde{N} = \{127, 191, 235, 237, 239, 247, 251, 253, 254, \dots\}$, etc. The first exception, $k = 3$ is for $\tilde{N}_{13} = 29$. The first double exception, $k = 4$ is for $\tilde{N}_{63} = 127$. However, in particular, for

$$\tilde{N}_3 = 2^s + 3 \cdot 2^l, \quad l = 0, 1, \dots, s - 2 \quad \Leftrightarrow \quad a_{\min}^{(\tilde{N}_3)} = s + 2, \quad \text{and} \tag{15}$$

$$\tilde{N}_{7,s-3} = 2^s + 7 \cdot 2^{s-3} = \{15, 30, 60, \dots\}, \quad s \geq 3 \quad \Leftrightarrow \quad a_{\min}^{(\tilde{N}_{7,s-3})} = s + 2. \tag{16}$$

Table 6. List of the shortest addition chain sequence generating factors for $1 \leq s \leq 5$.

s	2^s	The shortest addition chain sequence generating factors														
1	2	1_1														
2	4	1_2	3_1													
3	8	1_3	3_2	5_1	7_1											
4	16	1_4	3_3	5_2	7_2	9_1	11_1	13_1	15_1							
5	32	1_5	3_4	5_3	7_3	9_2	11_2	13_2	15_2	17_1	19_1	21_1	23_1	25_1	27_1	29_1	31_1

The shortest addition chain sequence generating factors for $1 \leq s \leq 5$ are listed in Table 7. The numbers of the form m_k mean that they generate sequences $N = 2^s + m \cdot 2^l$, where $l = 0, 1, \cdots, k - 1$, while the numbers in red indicate that certain numbers within the sequences they generate are exceptions to the general $a_{\min}^{(\tilde{N})} = s + 2$ rule. For example, if $s = 4$ then

$$\begin{aligned}
\hat{N} &= 2^4 + 1 \cdot 2^l = \{17, 18, 20, 24\}, \quad a_{\min}^{(\hat{N})} = 4 + 1 = 5, \\
\tilde{N}_3 &= 2^4 + 3 \cdot 2^l = \{19, 22, 28\}, \quad a_{\min}^{(\tilde{N}_3)} = 4 + 2 = 6, \\
\tilde{N}_5 &= 2^4 + 5 \cdot 2^l = \{21, 26\}, \quad a_{\min}^{(\tilde{N}_5)} = 4 + 2 = 6, \\
\tilde{N}_7 &= 2^4 + 7 \cdot 2^l = \{23, 30\}, \quad a_{\min}^{(\tilde{N}_7)} = 4 + 2 = 6, \\
\tilde{N}_9 &= 2^4 + 9 \cdot 2^l = 25, \quad a_{\min}^{(\tilde{N}_9)} = 4 + 2 = 6, \\
\tilde{N}_{11} &= 2^4 + 11 \cdot 2^l = 27, \quad a_{\min}^{(\tilde{N}_{11})} = 4 + 2 = 6, \\
\tilde{N}_{13} &= 2^4 + 13 \cdot 2^l = 29, \quad a_{\min}^{(\tilde{N}_{13})} = 4 + 3 = 7, \\
\tilde{N}_{15} &= 2^4 + 15 \cdot 2^l = 31, \quad a_{\min}^{(\tilde{N}_{15})} = 4 + 3 = 7,
\end{aligned} \tag{17}$$

where the last two values $a_{\min}^{(\tilde{N})}$ are higher than those given by the general rule. The exceptional $\tilde{N}$ values bear similarity to the atomic numbers of chemical elements that violate the Aufbau rule that predicts the electron configurations of most of the elements. Only about twenty chemical elements within $24 \leq Z \leq 103$ (with only two non-doubleton sets of consecutive ones) violate the Aufbau rule.

There are two major classes of orbits of dynamical systems: periodic and chaotic ones. For the shortest addition chains, $\hat{N}$ (OEIS A048645) and at least $\tilde{N}_3$ (15) and $\tilde{N}_{7,s-3}$ (16) represent periodic orbits, while the remaining numbers, which are not the sums of two powers of 2, chaotic ones. Only living systems have been found to be capable of producing abundant molecules with an assembly index greater than an experimentally determined value of 15 steps. The cut-off between 13 and 15 is sharp, which means that molecules made by random processes cannot have assembly indices exceeding 13 steps [3,16]. In particular, $N = 15$ is the length of the shortest addition chain for N which is smaller than the number of multiplications to compute N^{th} power by the Chandah-sutra method

(OEIS A014701). Furthermore, $N_{A014701} > N_{A003313}$ (OEIS A371894) for numbers that are not sums of two powers of 2 (OEIS A072823) and are not given by equation (15) but equation (16) provides their subset.

5. Maximum Assembly Index

In the following, we conjecture the form of the upper bound of the set of different binary string assembly indices. In general, of all strings C_k having a given assembly index, shown in Tables 1 and A6–A13 (Appendix D), most have $N_1 = \lfloor N/2 \rfloor$, though we have found a few exceptions, mostly for non-maximal assembly indices, namely for $a^{(8)} = 4$ ($4 < 8$) and for $a^{(8)} = 6$ ($24 < 26$), for $a^{(10)} = 4$ ($2 < 5$) and for $a^{(10)} = 5$ ($32 < 33$), and for $a^{(12)} = 4$ ($2 < 3$). These observations allow us to restrict the search space of possible strings with the largest assembly indices to balanced strings only: with the exception of $N = 8$, of all strings $C_k^{(N)}$ having a largest assembly index, most are balanced. We can further restrict the search space to ringed strings (Definition 3). If a string $C_{\min}$ for which $a^{(N)}(C_{\min}) = \min_k \left(\{a^{(N)}(C_k)\} \right)$ is constructed from repeating patterns, then a string $C_{\max}$ for which $a^{(N)}(C_{\max}) = \max_k \left(\{a^{(N)}(C_k)\} \right)$ must be the most patternless. The string assembly index must be bounded from above and $a^{(N)}(C_{\max})$ must be a monotonically nondecreasing function of N that can increase at most by one between N and $N + 1$. Certain heuristic rules apply in our binary case. For example,

- for $N = 7$ we cannot avoid two doublets (e.g. $2 \times [00]$) within a ringed string $E_{28}^{(7)} = [0011100]$ and thus $a^{(7)}(C_{\max}) = 5 < 6$,
- for $N = 8$ we cannot avoid two pairs of doublets (e.g. $2 \times [00]$ and $2 \times [11]$) within a ringed string $E_7^{(8)} = [00001111]$ and thus $a^{(8)}(C_{\max}) = 5 < 6$,
- for $N = 12$ we cannot avoid three pairs of doublets (e.g. $2 \times [00]$, $2 \times [10]$, and $2 \times [11]$) within a ringed string $E_k^{(12)} = [111000101100]$ and thus $a^{(12)}(C_{\max}) = 8 < 9$,
- for $N = 14$ we cannot avoid two pairs of doublets and one doublet three times (e.g. $2 \times [00]$, $2 \times [11]$, and $3 \times [01]$, and thus $a^{(14)}(C_{\max}) = 9 < 10$,
- etc.

Table 7 shows the exemplary balanced strings $B_{\max}$ having the largest assembly indices that we assembled (cf. also Appendix B). To determine the assembly index $a^{(18)} = 11$ of the string

$$E_k^{(18)} = [1(001)(11)(110)(110)(00)(001)0], \tag{18}$$

for example, we look for the longest patterns that appear at least twice within the string, and we look for the largest number of these patterns. Here, we find that each of the two triplets $[001]$ and $[110]$ appear twice in $E_k^{(18)}$ and are based on the doublets $[00]$ and $[11]$ also appearing in $E_k^{(18)}$. Thus, we start with the assembly pool $\{1, 0, [00], [001], [11], [110]\}$ made in four steps and join the elements of the pool in the following seven steps to arrive at $a^{(18)}(E_k) = 11$. On the other hand, another form of this balanced ringed string

$$E_l^{(18)} = [(01)(11)(110)(110)00(001)(01)0], \tag{19}$$

has $a^{(18)}(E_l) = 12$.

Table 7. Exemplary balanced strings $B_{\max}^{(N)}$ that have a largest assembly index. Conjectured ($a_{\text{conj}}^{(N)}$) form of the largest assembly index and its factual values for ringed ($a_{\text{rng}}^{(N)}$) and non-ringed ($a_{\text{nrng}}^{(N)}$) strings (red if below the conjectured value, green if above).

N	$B_{\max}^{(N)}$																				$a_{\text{conj}}^{(N)}$	$a_{\text{rng}}^{(N)}$	$a_{\text{nrng}}^{(N)}$
1	0																				0	0	0
2	1	0																			1	1	1
3	0	0	1																		2	2	2
4	0	0	1	1																	3	3	3
5	0	0	0	1	1																4	4	4
6	0	0	0	1	1	1															5	5	5
7	0	0	1	1	1	0	0														5	5	6
8	0	0	0	1	0	1	1	1													6	6	6
9	0	0	0	0	1	1	1	0	1												7	7	7
10	0	0	0	0	1	1	1	1	0	1											7	7	8
11	0	0	0	0	0	1	0	1	1	1	1										8	8	8
12	1	1	1	0	0	0	1	0	1	1	0	0									9	8	8
13	0	0	0	0	0	0	1	0	1	1	1	1	1								9	9	9
14	0	0	0	0	0	1	0	1	0	1	1	1	1	1							9	9	9
15	0	0	0	0	0	1	0	1	0	1	1	1	1	1	0						10	10	10
16	1	0	0	0	0	0	0	1	0	1	0	1	1	1	1	1					11	10	10
17	0	0	0	0	0	0	1	0	1	0	1	1	1	1	1	1	0				11	11	11
18	1	0	0	1	1	1	1	1	0	1	1	0	0	0	0	0	1	0			11	11	12
19	1	0	0	0	0	1	0	1	0	1	0	0	1	1	1	1	1	0	1		12	11	12
20	1	0	1	0	0	1	1	1	1	1	0	1	1	0	0	0	0	0	1	0	13	12	13

These results allow us to formulate the following conjecture.

Conjecture 1 (Tight upper bound on a string assembly index). *With exceptions for small N the largest string assembly index $a^{(N)}(C_{max})$ of a binary string as a function of N is given by a sequence formed by $\{+1, +1, k \times 0, +1, +1, k \times 0\}$ strings for $k \in \mathbb{N}_0$, where +1 denotes increasing $a^{(N)}(C_{max})$ by one, and 0 denotes maintaining it at the same level, and $a^{(0)} = -1$.*

However, at this moment, we cannot state whether this conjecture applies to ringed or non-ringed strings. The assembly indices for $N < 3$ are the same for a given N, whereas the assembly indices for $4 \le N \le 10$ were discussed above and are calculated in Appendix D for balanced and balanced ringed strings.

The conjectured sequence is shown in Figures 3 and 4 starting with $a^{(0)} = -1$ (we note in passing that $n = -1$ is a dimension of the void, the empty set $\varnothing$, or (-1)-simplex). Subsequent terms are given by $\{0, 1, 2, 3, 4, 5, 5, 6, 7, 7, 8, 9, 9, 9, 10, \dots\}$, which is periodic for $N = k(k+3)$ and flattens at $a^{(N)}(C_{\max}) = 4k - 3$, and $a^{(N)}(C_{\max}) = 4k - 1$, $k \in \mathbb{N}, k > 1$.

This sequence can be generated using the following procedure

```
step=1;                    % step flag
run =1;                    % run flag
flat=0;                    % flat counter

Nk = 0;
aub= -1;                   % the upper bound
while Nk < N
    if step < 3
        Nk = Nk+1;         % next Nk
        aub= aub + 1;      % increment the bound
    else                   % step==3
        for k=1:flat
            if flat > 0
                Nk = Nk+1; % next Nk
```

```
            end
        end
        run = run+1;          % increment run flag
        if run > 2
            run  = 1;         % reset run flag
            flat = flat+1; % increment flat counter
        end
    end
    step = step+1;            % increment step flag
    if step > 3
        step=1;               % reset step flag
    end
end
```

We note the similarity of this bound to the monotonically nondecreasing Shannon entropy of chemical elements, including observable ones [11]. Perhaps the exceptions in the sequence of Conjecture 1 vanish as N increases.

6. Binputation

So far we have assembled binary strings "manually". Now we shall automatize this process using other binary strings as assembling programs.

Definition 5. *The binary assembling program Q_B is a binary string of length s_Q that acts on the assembly pool P and outputs the assembled strings, adding them to the pool.*

Definition 6. *The trivial assembling program Q is a binary assembling program with consecutive bits denoting the following commands:*

 0 $\Leftrightarrow$ take the last element from P, join it with itself, and output,
 1 $\Leftrightarrow$ take the last two elements from P, join them with each other, and output.

As the assembly pool P is a distinct set to which strings are added in subsequent assembly steps, only these two commands are applicable to the initial assembly pool $P = \{1, 0\}$ containing only two bits.

Theorem 3. *If a string $C_{min}^{(N)}$ can be assembled by an elegant trivial program of length $s_Q = a^{(N)}(C_{min})$ then N is expressible as a product of Fibonacci numbers (OEIS A065108) and the length s_Q of any trivial program Q is not shorter than the assembly index of the string that this trivial assembling program assembles.*

Proof. An elegant program is the shortest program that produces a given output [42,43]. Furthermore, no program P shorter than an elegant program Q can find this elegant program Q [42]. If it could, it could also generate the Q's output. But if P is shorter than Q, then Q would not be elegant, which leads to a contradiction.

The 1st bit of the trivial assembling program Q is irrelevant as $Q = 0$ assembles $C_{min_1}^{(2)} = [00]$ and $Q = 1$ assembles $C_{min_4}^{(2)} = [10]$, so $Q = *$ assembles $C_{min_{1,4}}^{(2)} = [*0]$. Then the programs $Q = *0\ldots 0$ assemble the 2^{s_Q}-bit strings $C_{min_{1,4}}^{(2^s)} = [*0 * 0\ldots]$ having the assembly index $a_{min}^{(2^{s_Q})} = s_Q$, while strings $C_{min_{2,3}}^{2^{s_Q}}$ with the smallest assembly index $a_{min}^{(2^{s_Q})} = s_Q$ can be assembled with the same two programs starting with the reversed assembly pool $P = \{0, 1\}$.

The remaining $2^{s_Q-1} - 2$ programs will assemble some of the shorter strings with the assembly index $a_{min}^{(N)} = s_Q$. In general, all programs Q assemble strings having lengths expressible as a product of Fibonacci numbers (OEIS A065108) as shown in Table A2 (Appendix C), wherein out of 2^{s_Q-1} programs (cf. Tables A5 and A2):

- 2^{s_Q-2} programs $Q = *0\star\ldots$ assemble even length balanced strings $B = [*0*0\ldots]$ having natural binary entropies (3) $H(C) = \{0, 1\}$, including strings $C^{(2^{s_Q})}_{\min_{1,4}}$ (11),
- 2^{s_Q-3} programs $Q = *10\star\ldots$ assemble $[0*00*0\ldots]$ strings having lengths divisible by three and entropies $H(C) \approx \{0, 0.9183\}$,
- 2^{s_Q-4} programs $Q = *110\star\ldots$ assemble $[*00*0*00*0\ldots]$ strings having lengths divisible by five and entropies $H(C) \approx \{0, 0.9710\}$,
- 2^{s_Q-5} programs $Q = *1110\star\ldots$ assemble $[0*0*00*0\ldots]$ strings having lengths divisible by eight, entropies $H(C) \approx \{0, 0.9544\}$, and assembly indices $a^{(N)} = s_Q - 1$ if $* = 1$,
- $\ldots$,
- the program $Q = *1\ldots0$ joins two shortest strings assembled in a previous step into a string of length being twice the Fibonacci sequence (OEIS A055389), and finally
- the program $Q = *1\ldots1$ assembles the shortest string that has length belonging to the set of Fibonacci numbers.

Thus, for $* = 1$, binary assembling programs Q assemble subsequent $2^{s_Q-1} = 2^{s_Q-2} + 2^{s_Q-3} + \cdots + 2^0 + 1$ Fibonacci words and their concatenations having entropies (3) with ratios (4)

$$p_{1,m} = \frac{F_m}{F_{m+2}} \quad \text{and} \quad p_{0,m} = \frac{F_{m+1}}{F_{m+2}}, \tag{20}$$

where $m = \{1, 2, \ldots s_Q\}$, and F is the Fibonacci sequence starting from 1. Ratios (20) rapidly converge to

$$\lim_{s_Q \to \infty} p_{0,m} = \varphi - 1 \approx 0.618033989 \quad \text{and} \quad \lim_{s_Q \to \infty} p_{1,m} = 2 - \varphi \approx 0.381966011 \tag{21}$$

where φ is the golden ratio. Therefore, $\lim_{s_Q \to \infty} H_m \approx 0.9594$ is the binary entropy of the Fibonacci word limit. The Fibonacci sequence can be expressed through the golden ratio, which corresponds to the smallest Pythagorean triple $\{-3, 4, 5\}$ [29,44].

However, for $s_Q \geq 4$, some of the programs are no longer elegant if $* = 0$ and some of the assembled strings are not $C_{\min}$ if $* = 1$.

For $s_Q \geq 4$, $Q = 111100\ldots$ assembles a string

$$C^{(2^{s_Q-1})}_{\text{non-min}} = [01010010\ldots] \tag{22}$$

with an assembly index $a^{(2^{s_Q-1})} = s_Q$ which is not the minimum for this length of the string. For example, the 4-bit program $Q = *111$ assembles the string $C^{(8)} = [0*0*00*0]$, but if $* = 0$ this string can be assembled by a shorter 3-bit program $Q = *00$, and if $* = 1$ this string does not have the smallest assembly index $a^{(8)}(C_{\min}) = 3$ but $a^{(8)}(C_{\text{non-min}}) = 4$.

For $s_Q = \{4, 7\}$ and $s_Q \geq 10$ and for the shortest string assembled by the program Q the program Q is not elegant for $* = 0$ and the shortest string assembled by the program is not $C_{\min}$ for $* = 1$.

However, the length s_Q of any program Q is not shorter than the assembly index of the string that this program assembles. □

Theorem 3 would be violated if in Definition 6 we specified the command "0" e.g. as "take the last element from the assembly pool, join it with itself, join with what you have already assembled (say at "the right"), and output". Then the 2-bit program "00" would produce $[000000]$ with the assembly index $a^{(6)} = 3$. However, such a one-step command would violate the axioms of assembly theory, since it would perform two assembly steps in one program step. An elegant program to output the gigabyte binary string of all zeros would take a few bits of code and would have a low Kolmogorov complexity [45]. However, such a string would be *outputted*, not *assembled*. Furthermore, the length of such a program that outputs the string $[0\ldots]$ would be shorter than the length of the program that outputs the string $[10\ldots]$, while in AT, the lengths of these programs must be the same if the strings

have the same assembly indices. Definitions 5, 6 and Theorem 3 are about *binputation*[3], about binary strings assembling other binary strings.

The trivial assembly programs Q and the strings they assemble are listed in Tables 8 and A3–A5 (Appendix C) for one version of the assembly pool and for $1 \leq s_Q \leq 6$.

Table 8. 3-bit elegant programs assembling strings with $a^{(N)} = 3$.

Q	$C(s=1)$	$C(s=2)$	$C(s_Q=3)$	N
$*11$	$*0$	$0*0$	$*00*0$	5
$*10$	$*0$	$0*0$	$0*00*0$	6
$*01$	$*0$	$*0*0$	$*0*0*0$	6
$*00$	$*0$	$*0*0$	$*0*0*0*0$	8

We note in passing that there are other mathematical results on binary strings and the Fibonacci sequence. For example, it was shown [46] that having two concentric circles with radii $\{F_n, F_{n+2}\}$ and drawing two pairs of parallel lines orthogonal to each other and tangent to the inner circle, one obtains an octagon defined by the points of intersection of those lines with the outer circle, which comes very close to the regular octagon with $n \to \infty$. Furthermore, each of these octagons defines a Sturmian binary word (a cutting sequence for lines of irrational slope) except in the case of $n = 5$ [46].

Perhaps the smallest assembly index given by Theorem 2 and the strings of Theorem 3 are related to the Collatz conjecture, as the lengths of the strings (11) for $N = 2^{2k}$ correspond to the numbers to which the Collatz conjecture converges, from $N = (2^{2k} - 1)/3, k \in \mathbb{N}$ (OEIS A002450).

Theorem 3 is also related to Gödel's incompleteness theorems and the halting problem. N cases of the halting problem correspond only to $\log_2(N)$, not to N bits of information [47] and therefore, complexity is more fundamental to incompleteness than self-reference of Gödel's sentence [48]. Any formal axiomatic system only enables provable theorems to be proved. If a theorem can be proved by an automatic theorem prover, the prover will halt after proving this theorem. Thus, proving a theorem equals halting. If we assume that the axioms of the trivial program given by Definition 6 define the formal axiomatic system, then the strings having lengths expressible as a product of Fibonacci numbers assembled by this program would represent provable theorems.

Using the binary assembling program Q_B that were to use strings other than the last or two last strings in the assembly pool we would have to index the strings in the pool. However, at the beginning of the assembly process, we cannot predict in advance how many strings will enter the assembly pool. Thus, we do not know how many bits will be needed to encode the indices of the strings in the pool. Therefore, we state the following conjecture.

Conjecture 2. *There is no binary assembling program (Definition 5) that has a length shorter than the length of the string having the largest assembly index that could assemble this string.*

7. Discussion

The bounds of Theorem 2 and Conjecture 1 are shown in Tables 5 and 7 and illustrated in Figures 3 and 4. No binary string can be assembled in a smaller number of steps than given by a lower bound of Theorem 2. On the other hand, some strings cannot be assembled in a smaller number of steps than given by an upper bound (which for large N, as we suppose, has the form presented in Conjecture 1). Since one bit is the smallest amount and the quantum of information, the lower bound and the upper bound of the string assembly index define the allowed region of the assembly indices.

---

[3]   As an analog to chemputation, where assembly theory is applied in digital chemistry.
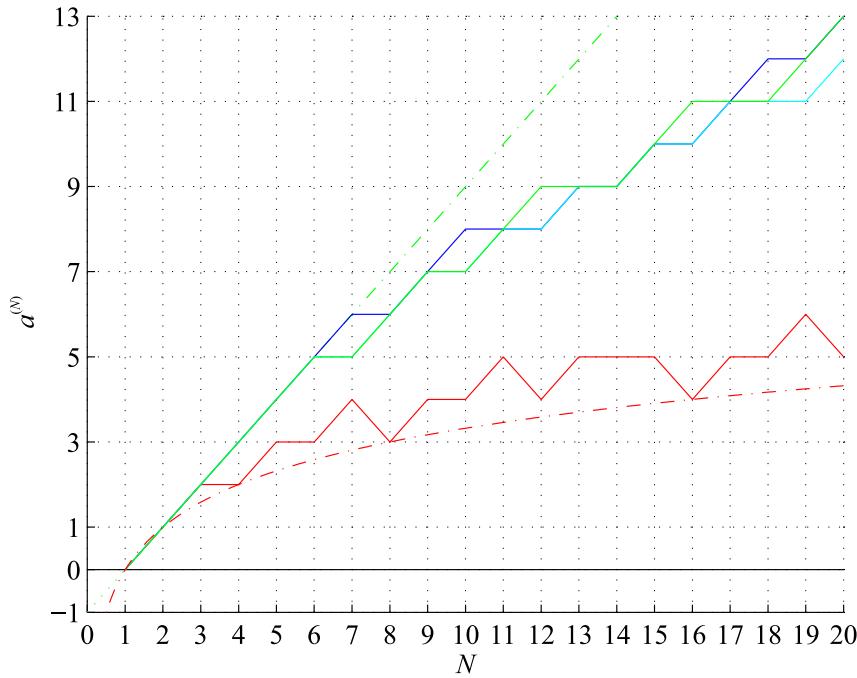
Figure 3. Lower bound on the binary string assembly index 2 (red) and $\log_2(N)$ (red, dash-dot), conjectured upper bound on the binary string assembly index **??** (green), factual values of the string assembly index (blue) and the ringed string assembly index (cyan) and $N-1$ (green, dash-dot), for the string length $0 \le N \le 20$.
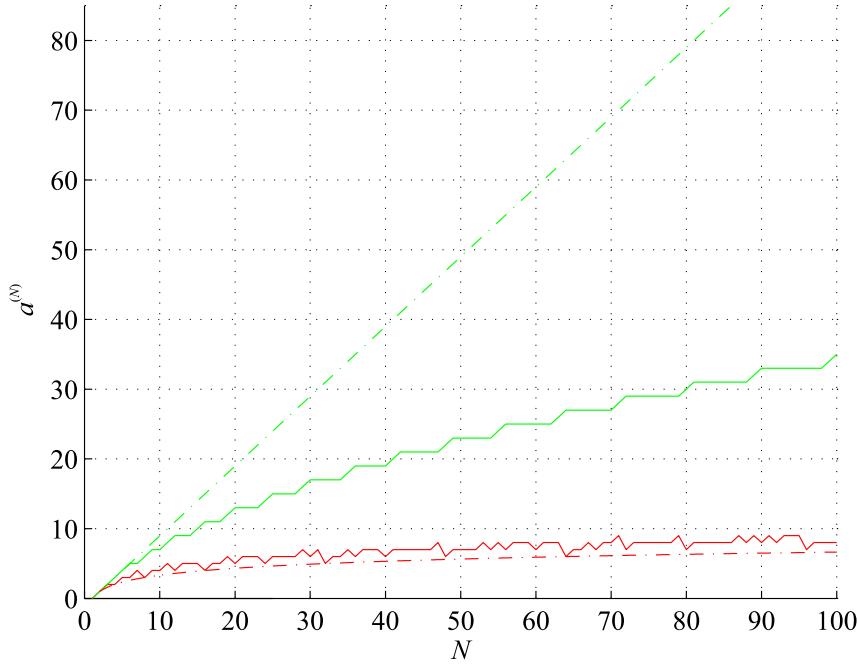
Figure 4. Lower bound on the binary string assembly index (red) and $\log_2(N)$ (red, dash-dot), conjectured upper bound on the string assembly index (green) and $N-1$ (green, dash-dot), for the binary string length $0 \le N \le 100$.

We found it much easier to determine an assembly index of a given binary string $C_k^{(N)}$ than to assemble a string so that it would have a largest assembly index. Similarly, a trivial string with the smallest assembly index for N can have the form $C_{\min_{1-4}}^{(N)} = [* \star \ldots]$ (11) or the form of a Fibonacci

word generated by the trivial assembling program (Definition 6). Therefore, we state the following conjecture.

Conjecture 3. *The problem of determining the assembly index of any binary string $C_k^{(N)}$ and the problem of assembling a non-trivial string so that it would have a smallest assembly index (Theorem 2) for N are NP-complete. The problem of assembling the string so that it would have a largest assembly index for N is NP-hard.*

A proof of conjecture 3 would also be the proof of the following conjecture.

Conjecture 4. $P \neq NP$

Every computable problem and every computable solution can be encoded as a finite binary string. Here, determining whether the assembly index of a given string has its known maximal value corresponds to checking the solution to a problem for correctness, whereas assembling such a string corresponds to solving the problem. Thus, AT would solve the P versus NP problem in theoretical computer science. There is ample pragmatic justification for adding $P \neq NP$ as a new axiom [47].

Consider the SARS-CoV-2 genome sequence containing 29903 nucleobases $\{A, C, G, T\}$, its initial version MN908947[4] collected in December 2019 in Wuhan (CN) and its sample OL351370[5] collected in Egypt nearly two years after the Wuhan outbreak, on October 23, 2021. In the MN version, the nucleobases are distributed as $|A| = 8954$, $|C| = 5492$, $|G| = 5863$, and $|T| = 9594$ and in the OL version as $|A| = 8954$, $|C| = 5470$, $|G| = 5856$, and $|T| = 9623$, following Chargaff's parity rules with the same count of adenines. We can convert these sequences into binary strings by assigning two bits per nucleobase. For such $N = 59806$, not being the sum of two powers of 2, with the degree of causation [6] given by equation (14), the assembly index is bounded by

$$21 \leq a^{(59806)}(C_k) \ll 971. \tag{23}$$

We increased the lower bound (23), given by Theorem 2, by one since strings with the smallest assembly indices cannot encode all nucleobases. The upper bound (23) was estimated by finding the smallest k that satisfies $k(k + 3) \geq N$ and using the relation $a^{(N)}(C_{\max}) = 4\lceil k \rceil - 1$ of Conjecture 1. We do not know the actual assembly indices of the MN and OL sequences. Their determination is an NP-complete problem, as we conjecture. There are twelve possible assignments of two bits per one nucleobase with twelve different Hamming weights and six different Shannon entropies (3)

$$\begin{aligned}
N_1\left(C_{MN}^{(59806)}\right) &= \{25801, 26172, 26441, 26812, 29263, 29532, 30274, 30543, 32994, 33365, 33634, 34005\}, \\
N_1\left(C_{OL}^{(59806)}\right) &= \{25750, 26136, 26419, 26805, 29234, 29517, 30289, 30572, 33001, 33387, 33670, 34056\}, \\
H\left(C_{MN}^{(59806)}\right) &= \{0.9864, 0.9887, 0.9903, 0.9923, 0.9997, 0.99989\}, \\
H\left(C_{OL}^{(59806)}\right) &= \{0.9860, 0.9885, 0.9902, 0.9922, 0.9996, 0.99988\}.
\end{aligned} \tag{24}$$

All sequences (24) are almost balanced ($N/2 = 29903$). However, the later OL versions are less balanced, producing lower Shannon entropies and showcasing the existence of an entopic force that governs genetic mutations [25]. We conjecture that the assembly index of the OL sequence is higher than that of the MN one. The evolution of information tends to increase the assembly index.

The bounds given by Theorem 2 and Conjecture 1, and the general bounds (1), and (2) on the assembly index are shown also in Figure 5 (adopted from [1] and modified). According to the authors

---

of [1], the green region is illustrative of the location in the complexity space where life might reasonably be found. Yellow region below can be thought of as being potentially naturally occurring, and red region above being so complex that even living systems might have been unlikely to create them. This is because they represent structures with limited internal structure and symmetries, which would require vast amounts of effort to faithfully reproduce [1].
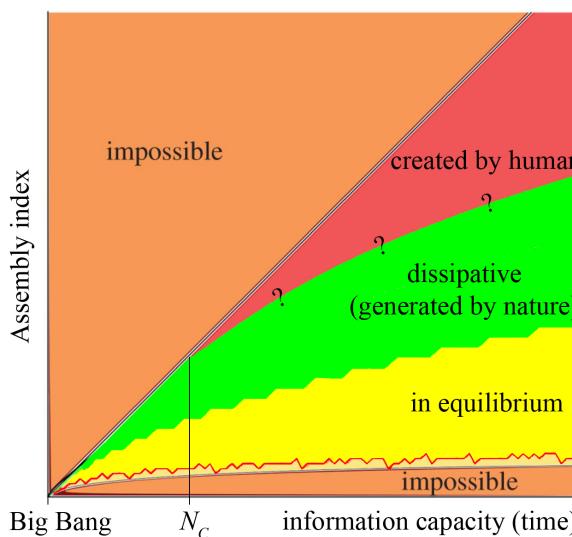
Figure 5. An illustrative graph of complexity against information capacity: orange regions are impossible, as they are above or below the assembly bounds; yellow region contains structures optimally assembled (in equilibrium); green region contains dissipative structures; and red region is the region of human creativity (figure not to scale).

We disagree with this statement.

Perceivable information about any *object* can be encoded by a binary string [14,15]. This does not imply that a binary string defines an *object*. Information that defines a chemical compound, a virus, a computer program, etc. can be encoded by a binary string. However, a dissipative structure [21] such as a living biological cell (or its conglomerate such as a human, for example) cannot be represented by a binary string (even if its genome can). This information can only be perceived (so this is not an *object defining* information). Each of us is given to ourselves as a mystery [49].

Furthermore, a binary string itself is neither dissipative nor creative. It is its assembly process that can be dissipative or creative. The [perceivable] universe is not big enough to contain the future; it is deterministic going back in time and non-deterministic going forward in time [50]. But we know [2,17–28] that it has evolved to the present since the Big Bang. Evolution is about assembling new information and optimizing its assembly process until it reaches the assembly index. Once the new information is assembled (by a dissipative structure operating far from thermodynamic equilibrium, or created by humans), it enters the realm of the 2nd law of thermodynamics, and nature seeks how to optimize its assembly pathway.

Only humans are gifted with creativity. Any creation is required to be shaped by the unique personality of the creator to such an extent that it is statistically one-time in nature [51]; it is an imprint of the author's personality. There is a certain minimum amount of information N_C required to establish a creation, as shown in Figure 5. Sixteen possibilities provided by the minimum thermodynamic entropy [31–33] bifurcate the assembly pathways (cf. Theorem 1) but none of these possibilities can be considered a *creation*. However, the boundary between the green region of dissipative structures [21] and the red region of human creativity remains to be discovered.

At first, the newly assembled information corresponds to the discovery by groping [20]. However, its assembly pathway does not attain its most economical or efficient form all at once. For a certain period of time, its evolution gropes about within itself. The try-out follows the try-out, not being

finally adopted. Then finally perfection comes within sight, and from that moment the rhythm of change slows down [20]. The new information, having reached the limit of its potentialities, enters the phase of conquest. Stronger now than its less perfected neighbours, the new information multiplies and consolidates. When the assembly index is reached, new information attains equilibrium, and its evolution terminates. It becomes stable.

> "Thanks to its characteristic additive power, living matter (unlike the matter of the physicists) finds itself 'ballasted' with complications and instability. It falls, or rather rises, towards forms that are more and more improbable. Without orthogenesis life would only have spread; with it there is an ascent of life that is invincible." [20]

The total entropy of the universe S is constant and is the sum of the information entropy S_{info} and the physical entropy S_{phys}. Therefore, over time [28]

$$\frac{dS_{info}}{dt} + \frac{dS_{phys}}{dt} = 0. \tag{25}$$

The time corresponds to an increasing information capacity. Bit by bit: $dt = (N+1) - N = 1$

BB having the energy given by mass-energy equivalence

$$E_{BB} = \frac{k}{2}M_{BB}c^2 = \frac{k}{2}m_{BB}E_P = \frac{d_{BB}}{4}E_P = \frac{1}{4}\sqrt{\frac{N_{BB}}{\pi}}E_P,$$

$$2 \le k \le k_{max} = \frac{2\alpha^2}{\sqrt{\alpha^4 - \alpha_2^4}} \approx 6.7933 \tag{26}$$

where $M_{BB} := m_{BB}m_P$, $m_{BB} \in \mathbb{R}$ denote the BB mass, and E_P, m_P denote the Planck energy and mass, $\alpha \approx 1/137.036$ is the fine-structure constant and $\alpha_2 \approx -1/140.178$ is the 2nd fine-structure constant related to α by $(\alpha + \alpha_2)/(\alpha\alpha_2) = -\pi$, and k is the BB size-to-mass ratio (STM) [19] ($k = 2$ if BB is BH).

It was shown [18] based on the Mandelstam-Tamm [52], Margolus–Levitin [53], and Levitin-Toffoli [54] theorems on the quantum orthogonalization interval that BBs generate (or rather *assemble*) a pattern forming nonequilibrium shell (VS) through the solid-angle correspondence, as shown in Figure 6. The BB entropic work

$$W_{BB} = T_{BB}S_{BB} = T_{BB}\frac{1}{4}k_B N_{BB} = T_{BB}\frac{1}{4}k_B \pi d_{BB}^2$$

$$= \frac{E_P d_{BB}}{4k}\left(1 \pm i\sqrt{\frac{k^2}{4} - 1}\right), \tag{27}$$

is the work done by all APTs of a BB. It is the product of the BB entropy [31–33] and the general, complex BB temperature

$$T_{BB} = \frac{T_P}{k\pi d_{BB}}\left(1 \pm i\sqrt{\frac{k^2}{4} - 1}\right), \tag{28}$$

which in modulus and for a BH ($k = 2$) reduces [19] to Hawking temperature

$$T_{BH} = \frac{\hbar c^3}{8\pi G M_{BH} k_B} = \frac{T_P}{2\pi d_{BH}}, \tag{29}$$

where $\hbar = h/(2\pi)$ is the reduced Planck constant, G is the gravitational constant, and T_{P} is the Planck temperature. In particular [19]

$$T_{\mathrm{BB}}(k_{\max}) = \frac{T_{\mathrm{P}}}{2\pi d_{\mathrm{BB}}\alpha^2}\left(\sqrt{\alpha^4 - \alpha_2^4} \pm i\alpha_2^2\right), \tag{30}$$

$$T_{\mathrm{BB}}(k_{\mathrm{eq}}) = \frac{T_{\mathrm{P}}}{2\pi d_{\mathrm{BB}}}\frac{\alpha^2 \pm i\alpha_2^2}{\sqrt{\alpha^4 + \alpha_2^4}}, \tag{31}$$

where

$$k_{\mathrm{eq}} = \frac{2}{\alpha^2}\sqrt{\alpha^4 + \alpha_2^4} \approx 2.7665. \tag{32}$$

is the energy equilibrium STM.
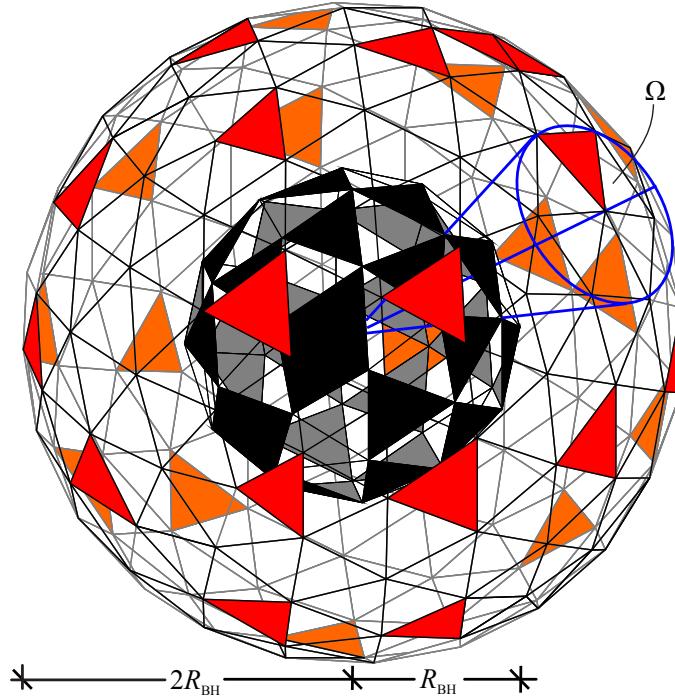
Figure 6. A black body object as a generator of an entropy variation shell (VS) through the solid angle Ω correspondence.

A VS has the information capacity bounded by

$$\begin{aligned}N_{\mathrm{BB}} \leq N_{\mathrm{VS}} \leq 4N_{\mathrm{BB}},\\ N_{\mathrm{VS}} := lN_{\mathrm{BB}}, \quad 1 \leq l \leq 4,\end{aligned} \tag{33}$$

where l is a VS defining factor. The number of APTs is bounded by

$$\left\lfloor\frac{1}{4}N_{\mathrm{BB}}\right\rfloor \leq N_1 \leq \left\lfloor\frac{1}{2}N_{\mathrm{BB}}\right\rfloor, \tag{34}$$

as shown in Figure 7, and thus its binary potential $\delta\varphi_{\mathrm{VS}} = -N_1 c^2/N_{VS}$ [17,18] is bounded by

$$-\frac{1}{2}c^2 \leq \delta\varphi_{\mathrm{VS}} < \begin{cases}\left(\frac{1}{N_{\mathrm{BB}}} - \frac{1}{4}\right)c^2 > 0, & N_{\mathrm{BB}} < 4 \\ \left(\frac{1}{4N_{\mathrm{BB}}} - \frac{1}{16}\right)c^2 < 0, & N_{\mathrm{BB}} > 4\end{cases}. \tag{35}$$

and the theoretical probability $p_1 := N_1/N_{VS}$ for a triangle on a VS to be an active Planck triangle is also bounded [18] by

$$\frac{1}{16} - \frac{1}{4N_{BB}} < p_1 \leq \frac{1}{2},\tag{36}$$

which is satisfied by the ratio $p_{1,m}$ (20) of the trivial assembling program 6 for $N_{BB} > 2$. On the other hand, the entropy variation [17,55] $\delta S/k_B = -c\delta\varphi$ so that for $N_{BB} < 4$ the lower bound (36) is negative and the upper bound (35) is positive ($N_1 \leq 1$ in this range). The Planck triangle of VS is located *somewhere* on the VS surface defined by a solid angle

$$\Omega = \frac{\ell_P^2}{R_{BB}^2} = \frac{4\pi\ell_P^2}{4\pi R_{BB}^2} = \frac{4\pi}{N_{BB}},\tag{37}$$

that corresponds to the BB Planck triangle.
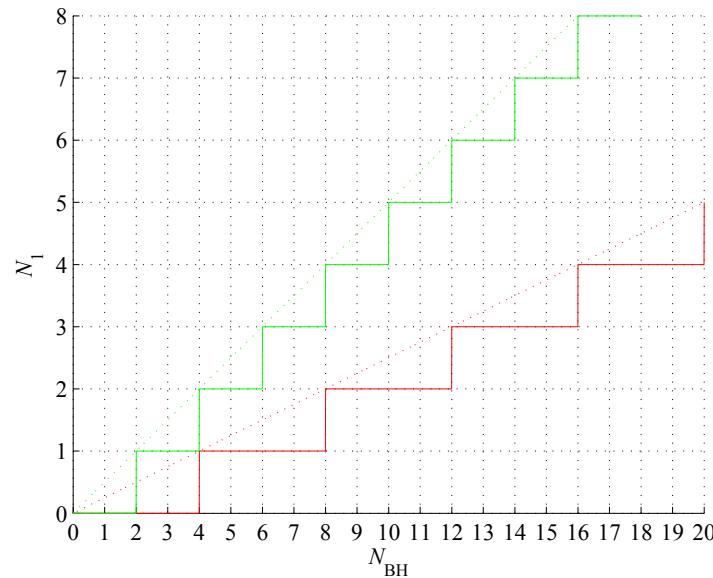
Figure 7. Lower (red) and upper (green) bound on the number of APTs N_1 on a VS as a function of the information capacity of the generating BB [18].

The BB information capacity is dictated by its diameter and the BB energy (26) as a function of its diameter is the same for all BBs (it is independent on k). However, the BB mass and density

$$\rho_{BB} = \frac{M_{BB}}{V_{BB}} = \frac{3}{kN_{BB}}\rho_P,\tag{38}$$

are not.

Based on the orbiting condition $V_O^2 \leq V_R^2 \leq V_E^2$, where $V_O = \sqrt{GM_C/R_{avg}}$ is the orbital, and $V_E = \sqrt{2GM_C/R_{avg}}$ is the escape speed of an orbiting *object*, R_{avg} is the average distance from the center of the central *object* to the center of the orbiting *object*, and M_C is the mass of the central *object*, the bounds

$$N_{BB} \leq 4v_R^4 N_{VS} \leq 4N_{BB},\tag{39}$$

containing the velocity term $V_R = v_R c$, $v_R \in \{\mathbb{R}, \mathbb{I}\}$ were also derived [18]. Plugging N_{VS} from the bounds (33) into the bounds (39) we arrive at

$$\frac{1}{4l} \leq v_R^4 \leq \frac{1}{l},\tag{40}$$

which is satisfied by real and imaginary (but not complex) velocities (for example, for $l = 1$ by $-1 \leq v_R \leq -1/\sqrt{2}$, $1/\sqrt{2} \leq v_R \leq 1$, $-i \leq v_R \leq -i/\sqrt{2}$, and $i/\sqrt{2} \leq v_R \leq i$). Taking the square root of the bounds (40), using $v_{LL}^2 + v_{RR}^2 = 1$, $v_R \in \{\mathbb{R}, \mathbb{I}\}$ [18], and squaring again, we arrive at

$$\frac{l - 2\sqrt{l} + 1}{l} \leq v_L^4 \leq \frac{4l - 4\sqrt{l} + 1}{4l}. \tag{41}$$

The bounds (40) and (41), shown in Figure 8, meet at $v = 1/\sqrt{2}$, where de Broglie and Compton wavelengths of mass M are the same

$$\lambda_{\mathrm{dB}} = \frac{h}{p} = h\frac{\sqrt{1 - \frac{V^2}{c^2}}}{MV} = \lambda_{\mathrm{C}} = \frac{h}{Mc} \Leftrightarrow \frac{V}{c} = \frac{1}{\sqrt{2}}, \tag{42}$$

where p is the relativistic momentum. The same is the ratio of orbital to escape speed: $\frac{V_O}{V_E} = \frac{1}{\sqrt{2}}$.
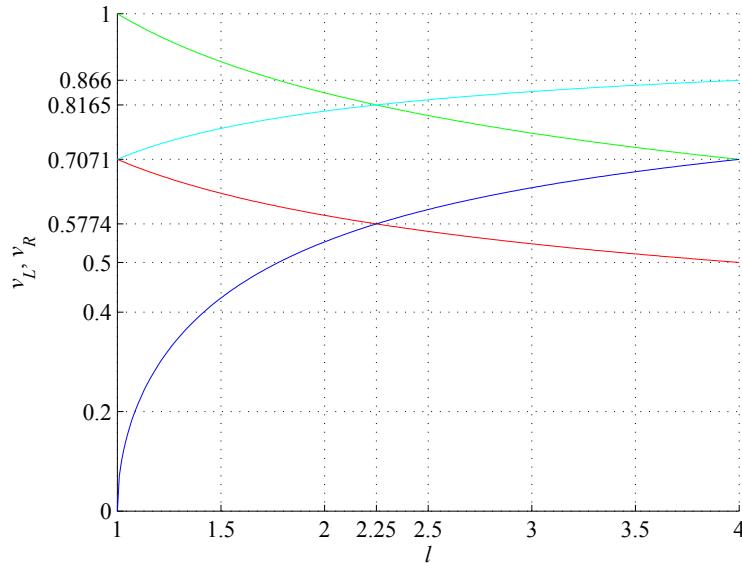
Figure 8. Lower (red) and upper (green) bounds on v_R and lower (blue) and upper (cyan) bounds on v_L as a function of l defining VS. Characteristic velocities are $\{0, \sqrt{1/4}, \sqrt{1/3}, \sqrt{2/4}, \sqrt{2/3}, \sqrt{3/4}, 1\}$, $v_L, v_R \in \mathbb{R}_+$.

Furthermore, the bounds (40) and (41) do not overlap only for $l = \{1, 4\}$. Therefore, $1 < l < 4$ defines the dissipativity or the assembly range. Furthermore, the intersection of the bounds (40) and (41) is the common region for both velocities. If v_L is within this region, then v_R is as well. We note that the average orbital velocity of each orbiting *object* only slightly exceeds its orbital speed V_O. This implies that the average VS defining factor $l_{\mathrm{avg}} \gtrsim 1$ in (33) for a VS orbiting *object* (cf. Appendix A).

BBs define a perfect thermodynamic equilibrium, and the bounds (33) and (34) show that nature uses optimally assembled information (cf. Conjecture **??**) to assemble new information. Figure 9 shows the bounds on the string assembly indices and Figure 10 shows the BB temperature (29), energy (26), and entropic work (27) for $0 \leq N_{\mathrm{BB}} \leq 5$. $k_{\mathrm{B}}|T_{\mathrm{BB}}|/E_{\mathrm{BB}} = 2/N_{\mathrm{BB}}$ is a rational number for natural N_{BB}. Furthermore, $\log_2(N_{\mathrm{BB}}) > N_{\mathrm{BB}} - 1$ for $N_{\mathrm{BB}} \in \mathbb{R}$ and $1 < N_{\mathrm{BB}} < 2$.

Let us examine this process starting from the Big Bang during the Planck epoch and shortly thereafter, and for continuous $N_{\mathrm{BB}} \in \mathbb{R}$ (i.e., including fractional Planck triangle(s)).
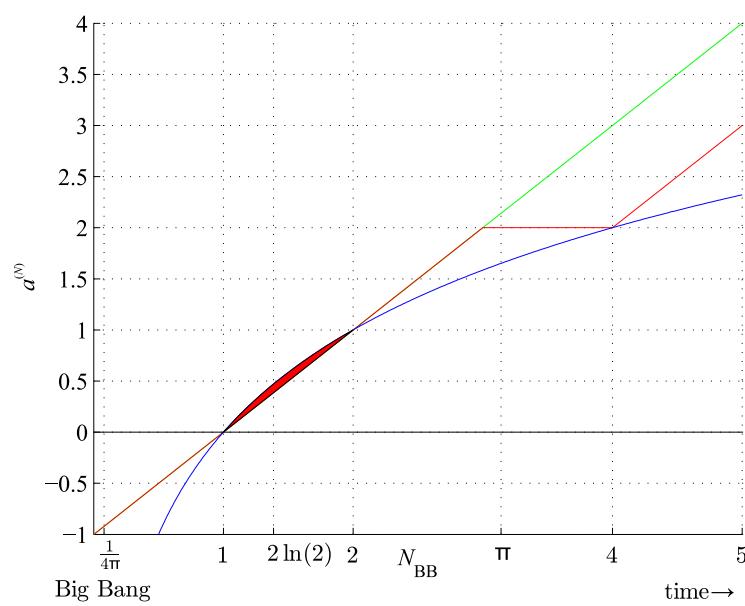
Figure 9. Lower (red) and upper (green) bounds on the binary string assembly index of length N_{BB} and $\log_2(N_{BB})$ (blue), for $0 \le N_{BB} \le 5$.
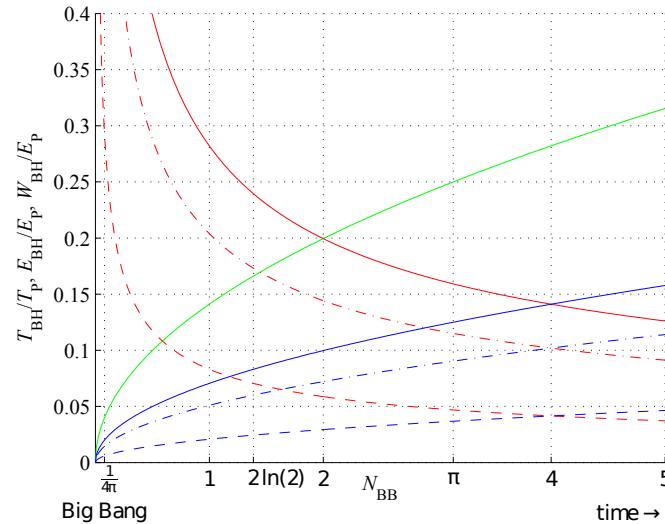
Figure 10. Black body object energy E_{BB} (green); temperature T_{BH} (red), $\mathrm{Re}[T_{BB}(k_{eq})]$ (red, dash-dot), $\mathrm{Re}[T_{BB}(k_{max})]$ (red, dash); and work W_{BH} (blue), $\mathrm{Re}[W_{BB}(k_{eq})]$ (blue, dash-dot), $\mathrm{Re}[W_{BB}(k_{max})]$ (blue, dash),as a function of its information capacity N_{BB} in terms of Planck units, for $0 \le N_{BB} \le 5$.

$N_{BB} = 0$

There is nothing to talk about. It is a mystery.
$0 < N_{BB} < 1$

The Big Bang has occurred, forming the 1st BB. At $N_{BB}(k_{max}) = (\alpha^4 - \alpha_2^4)/(4\pi\alpha^4) \approx 0.0069$ the BB temperature (29) and subsequently at $N_{BH} = 1/(4\pi) \approx 0.0796$ the BH temperature (29) become equal to the Planck temperature, but any BB in this range is still too small to carry a single bit of information and cannot be triangulated. However, independent BBs merge [18,19] summing their entropies and increasing the information capacity.

$N_{BB} = 1$

The first bit (a degree of freedom [18]) becomes available and APTs on BBs begin to fluctuate providing the initial assembly pool $P = \{1, 0\}$. The BH energy reaches the limit of the equipartition theorem for one bit ($E_{BH} = \frac{1}{2} k_B T_{BH}$). However, the bounds (34) make them unable to generate any APTs on a VS ($N_1 = 0$).

$1 < N_{BB} < 2$

This is the only range in which the lower AT bound (1) is greater than the upper AT bound (2).

The BH temperature (29) exceeds its energy (26) ($\frac{1}{2} k_B T_{BH} < E_{BH} < k_B T_{BH}$) [18]. At $N_{BH} = 2\ln(2)$ the BH energy (26) is equal to the Landauer limit $E_{BH} = k_B T_{BH} \ln(2) \approx 1.3863$ [56]. Shortly thereafter, at $N_{BH} = 1.5$, the BH density reaches the level of the Planck density For a BB [19] Still $N_1 = 0$. Merging BBs expand fractional Planck triangle(s) to form the 2nd bit.

$N_{BB} = 2$

The first nonvanishing $N_1 = 1$ becomes available on a VS generated by a BB. The BH temperature (29) is equal to its energy (26) ($k_B T_{BH} = E_{BH} = E_P / (2\sqrt{2\pi})$).

$2 < N_{BB} < 3$

At $N_{BB} = 4\ln(2)$ the BH entropic work (27) is equal to the Landauer limit ($k_B T_{BH} = W_{BH} = E_P / (4\sqrt{\pi})$). At $N_{BB} > 2.4507$ the density of the least dense BB ($k_{max} \approx 6.7933$) decreases below the modulus of its temperature. $N_1 = \{0, 1\}$.

$N_{BB} = 3$

$3 < N_{BB} < 4$

With $N_{BB} > 3$ BBs can finally be triangulated. Yet, containing only one APT ($N_1 = \{0, 1\}$), they are not ergodic [18].

At $N_{BH} > \pi$ the BH surface gravity $g_{BH} = 1/d_{BH}$ decreases below the Planck acceleration and the tangential acceleration [17,18] becomes real ($a_L \in \mathbb{R}$).

$N_{BB} = 4$

The BB assembly index bifurcates, minimal thermodynamic entropy [32] is reached, and the relation (34) provides the second bit on a VS ($N_1 = 2$). At this moment BB can be assembled in a different number of steps and nature seeks to minimize this number following the dynamics induced by the relation (25). The BH temperature (29) is equal to its entropic work (27) ($k_B T_{BH} = W_{BH}$).

$4 < N_{BB} < 6$

The BH temperature (29) finally decreases below the entropic work (27) limit and $N_1 \geq 2$.

$N_{BB} = 6$

A BB reaches the upper bound on ringed assembly index.

$6 < N_{BB} < 7$

The imaginary Planck time appears at the BH surface [17] heralding the end of the Planck epoch. After crossing this threshold, the VSs begin to operate with $1 \leq N_1 \leq 3$ on $2\pi < N_{VS} \leq 8\pi$, and the first dissipative structures can be assembled.

Nature enters a directed exploration phase ($\alpha < 1$) and selectivity emerges, limiting the discovery of new objects [6].

$N_{BB} = 7$

A BB reaches the upper bound on non-ringed assembly index.

. . .

$N_{BB} > 12$

At $N_{BB} = 4\pi$ a first precise diameter relation can be established between the vertices of the BB surface. Furthermore, for $N_{BB} = 4\pi$, the solid angle (37) equals one steradian.

> ...
> $N_{\text{BB}} > N_C$
> The onset of human creativity.

8. Conclusions

The results reported here can be applied in the fields of cryptography, data compression methods, stream ciphers, approximation algorithms [57], reinforcement learning algorithms [58], information-theoretically secure algorithms, etc. Another possible application of the results of this study could be molecular physics and crystallography. Overall, the results reported here support the AT, the Bekenstein's minimum thermodynamic entropy [31–33], the holographic principle [34], entropic gravity [35], emergent dimensionality [8,11,17–19,24,26,27,29], the second law of infodynamics [25,28], and invite further research.

Author Contributions: WB: Conjecture concerning the diversification of strings in Theorem **??**; partitioning conjecture for $N \geq 16$ resulting in the flattening of the string assembly index upper bound curve; observation that the string assembly index upper bound curve should be monotonically nondecreasing; prior-art search; numerous clarity corrections and improvements; SŁ: The remaining part of the study.

Data Availability Statement: The public repository for the code written in MATLAB computational environment is given under the link https://github.com/szluk/Evolution_of_Information (accessed on November 23, 2023).

Acknowledgments: The authors thank Piotr Masierak for his research on the general strategy for determining the string assembly indices and creating the $E_k^{(18)}$ string (*Peter's rock*), Andrzej Tomski for clarity corrections, and Mariola Bala for noting that "this is logical". SŁ thanks his wife Magdalena Bartocha for her unwavering support and motivation and his partner and friend, Renata Sobajda, for her prayers.

Abbreviations

The following abbreviations are used in this manuscript:

AT	assembly theory;
BH	black hole;
BB	black-body object (BH, white dwarf, neutron star);
VS	nonequilibrium shell;
APT	active Planck triangle;
N	length of a binary string;
N_0	number of 0's in the binary string;
N_1	binary Hamming weight of the binary string;
$C_k^{(N)}$	binary string of length N;
$B_k^{(N)}$	balanced string of length N;
$R_k^{(N)}$	ringed string of length N;
$E_k^{(N)}$	balanced ringed string of length N;
$|C^{(N)}|$	number of binary strings of length N (2^N);
$|B^{(N)}|$	number of balanced strings of length N (OEIS A001405);
$|R^{(N)}|$	number of ringed strings (OEIS A000031);
$|E^{(N)}|$	number of balanced ringed strings;
$a^{(N)}$	assembly index of a string of length N;
$P = \{1,0\}$	initial assembly pool;
s	assembly step;
Q	binary assembling program;
s_Q	length of the binary assembling program;
F	Fibonacci sequence.

Appendix A. Orbital Velocities and the VS Defining Factor

Table A1 shows the orbital speed V_O and escape speed V_E of some celestial objects, their minimal $V_{\min}$ and maximal $V_{\max}$ velocities[6]. The former lie below the orbital speed limits. The average VS defining factor $l_{\text{avg}} = (l_{\max} - l_{\min})/2$, where $l_{\min/\max} = 3(V_{\min/\max} - V_O)/(V_E - V_O) + 1$ was determined by linear interpolation.

Table A1. Exemplary orbital speeds and velocities, and the average VS defining factor l_{avg}.

Object	V_O [km/s]	$V_{\min}$ [km/s]	$V_{\max}$ [km/s]	V_E [km/s]	l_{avg}
Mercury	47.88	38.86	58.98	67.71	1.158
Venus	35.02	34.79	35.26	49.53	1.000
Earth	29.79	29.29	30.29	42.13	1.000
Mars	24.13	21.97	26.50	34.13	1.030
Jupiter	13.06	12.44	13.72	18.47	1.011
Saturn	9.62	9.09	10.18	13.61	1.009
Uranus	6.8	6.49	7.11	9.61	1.000
Neptune	5.43	5.37	5.50	7.68	1.001
Pluto	4.74	3.71	6.10	6.70	1.247
The Moon	1.02	0.96	1.08	14.40	1.011

Appendix B. Exemplary Strings with Maximal Assembly Indices

For the exemplary balanced ringed strings $E_{\max}$, shown in Table 7:

- all forms of $E_k^{(4)} = [0011]$ have $a^{(4)} = 3$,
- all forms of $E_6^{(5)} = [00011]$ have $a^{(5)} = 4$,
- all forms of $E_{16}^{(6)} = [000111]$ have $a^{(6)} = 5$,
- the form $E_{28}^{(7)} = [0011100]$ has $a^{(7)} = 5$ but the form $E_{34}^{(7)} = [0001110]$ has $a^{(7)} = 6$,
- all forms of $E_{45}^{(8)} = [00010111]$ have $a^{(8)} = 6$,
- all forms of $E_{13}^{(9)} = [000011101]$ have $a^{(9)} = 7$,
- the form $E_{22}^{(10)} = [0000111101]$ has $a^{(10)} = 7$ but the form $E_l^{(10)} = [0111101000]$ has $a^{(10)} = 8$,
- all forms of $E_7^{(11)} = [00000101111]$ have $a^{(11)} = 8$,
- all forms of $E_9^{(12)} = [111000101100]$ have $a^{(12)} = 8$,
- all forms of $E_8^{(13)} = [0000001011111]$ have $a^{(13)} = 9$,
- all forms of $E_k^{(14)} = [00000101011111]$ have $a^{(14)} = 9$,
- all forms of $E_k^{(15)} = [000001010111110]$ have $a^{(15)} = 10$,
- all forms of $E_k^{(16)} = [1000000101011111]$ have $a^{(16)} = 10$,
- all forms of $E_k^{(17)} = [00000010101111110]$ have $a^{(17)} = 11$,
- all forms of $E_k^{(18)} = [000000101010111111]$ have $a^{(18)} = 11$,
- some forms of $E_k^{(19)} = [1000010101001111101]$ have $a^{(19)} = 12$,
- some forms of $E_k^{(20)} = [10100111110110000010]$ have $a^{(20)} = 13$.

Appendix C. Trivial Assembling Programs

Table A2 shows the lengths of the strings assembled by the trivial assembling program introduced in Section 6 for $1 \leq s_Q \leq 7$. The table is divided into sections corresponding to sets of assembled strings having the same form but different lengths. For example, thirty two 7-bit programs in the bottom section assemble strings $C = [*1 * 1 \dots]$. The boxed symbols denote program commands, not the string lengths.

Table A2. Lengths of the strings assembled by trivial assembly programs Qs (OEIS A065108).

$a^{(N)}$	0	1	2	3	3,4	4,5	5,6	6,7
s_Q		1	2	3	4	5	6	7
1	1	2	3	5	8‡	13	21	34‡
							[★]	42
						[0]	26	39
							[★]	52
					[0]	16‡	24‡	40‡
							[★]	48‡
						[★]	32‡	48‡
							[★]	64‡
				[0]	10	15	25	40‡
							[★]	50
						[★]	30	45
							[★]	60
					[★]	20	30	50
							[★]	60
						[★]	40	60
							[★]	80
			[0]	6	9	15	24†	39
							[★]	48†
						[★]	30	45
							[★]	60
					[★]	18	27	45
							[★]	54
						[★]	36	54
							[★]	72
				[★]	12	18	30	48†
							[★]	60
						[★]	36	54
							[★]	72
					[★]	24	36	60
							[★]	72
						[★]	48	72
							[★]	96
		[0]	4	6	10	16†	26	42
							[★]	52
						[★]	32†	48†
							[★]	64†
					[★]	20	30	50
							[★]	60
						[★]	40	60
							[★]	80
				[★]	12	18	30	48†
							[★]	60
						[★]	36	54
							[★]	72
					[★]	24	36	60
							[★]	72
						[★]	48	72
							[★]	96
			[★]	8	12	20	32†	52
							[★]	64†
						[★]	40	60
							[★]	80
					[★]	24	36	60
							[★]	72
						[★]	48	72
							[★]	96
				[★]	16	24	40	64†
							[★]	80
						[★]	48	72
							[★]	96
					[★]	32	48	80
							[★]	96
						[★]	64	96
							[★]	128

Table A3. 4-bit programs assembling strings with $a^{(N)} = \{3,4\}$.

Q	$C(s=3)$	$C(s_Q = 4)$	N
$*111$	$*00*0$	$0*0*00*0$	$8^{\ddagger}$
$*110$	$*00*0$	$*00*0\dots$	10
$*101$	$0*0\dots$	$0*0\dots$	9
$*100$	$0*0\dots$	$0*0\dots$	12
$*011$	$*0\dots$	$*0\dots$	10
$*010$	$*0\dots$	$*0\dots$	12
$*001$	$*0\dots$	$*0\dots$	12
$*000$	$*0\dots$	$*0\dots$	16

‡. This program is not elegant if $* = 0$ and the assembled string is not C_{min} if $* = 1$.

Table A4. 5-bit programs assembling strings with $a^{(N)} = \{4,5\}$.

Q	$C(s=4)$	$C(s_Q = 5)$	N
$*1111$	$0*0*00*0$	$*00*00*0*00*0$	13
$*1110$	$0*0*00*0$	$0*0*00*0\dots$	$16^{\ddagger}$
$*1101$	$*00*0\dots$	$*00*0\dots$	15
$*1100$	$*00*0\dots$	$*00*0\dots$	20
$*1011$	$0*0\dots$	$0*0\dots$	15
$*1010$	$0*0\dots$	$0*0\dots$	18
$*1001$	$0*0\dots$	$0*0\dots$	18
$*1000$	$0*0\dots$	$0*0\dots$	24
$*0111$	$*0\dots$	$*0\dots$	$16^{\dagger}$
$*0110$	$*0\dots$	$*0\dots$	20
$*0101$	$*0\dots$	$*0\dots$	18
$*0100$	$*0\dots$	$*0\dots$	24
$*0011$	$*0\dots$	$*0\dots$	20
$*0010$	$*0\dots$	$*0\dots$	24
$*0001$	$*0\dots$	$*0\dots$	24
$*0000$	$*0\dots$	$*0\dots$	32

†. This program is not elegant (the same string can be assembled using the shorter 4-bit program $*000$). ‡. This program is not elegant if $* = 0$ and the assembled string is not C_{min} if $* = 1$.

Table A5. 6-bit programs assembling strings with $a^{(N)} = \{5,6\}$.

Q	$C(s_Q = 6)$	N
$*11111$	$0*0*00*0*00*00*0*00*0$	21
$*11110$	$*00*00*0*00*0\dots$	26
$*1110\star$	$0*0*00*0\dots$	$24^{\ddagger}, 32^{\ddagger}$
$*110\star\star$	$*00*0\dots$	$25, 30, 40$
$*10\star\star\star$	$0*0\dots$	$24^{\dagger}, \dots, 48$
$*0\star\star\star\star$	$*0\dots$	$26, 32^{\dagger}, \dots, 64$

†. This program is not elegant. ‡. This program is not elegant if $* = 0$ and the assembled string is not C_{min} if $* = 1$.

Appendix D. Binary Strings and Their Assembly Indices

Table A2 show the lengths of the strings assembled by programs Fs having the smallest assembly indices. Tables A6-A13 show distributions of the assembly indices for $5 \leq N \leq 12$. Tables A14-A18 show balanced strings $B^{(N)}$ and their assembly indices for $5 \leq N \leq 8$. Tables A19-A24 show the balanced ringed strings $E^{(N)}$ and their assembly indices for $5 \leq N \leq 10$. Tables A25-A27 show selected balanced ringed strings $E^{(N)}$ and their assembly indices for $11 \leq N \leq 13$.

Table A6. Distribution of the assembly indices for $N = 5$.

$a^{(5)}(C)$	$\lvert a^{(5)}(C)\rvert$	N_1					
		0	1	2	3	4	5
3	18	1	3	5	5	3	1
4	14		2	5	5	2	
	32	1	5	10	10	5	1

Table A7. Distribution of the assembly indices for $N = 6$.

$a^{(6)}(C)$	$\lvert a^{(6)}(C)\rvert$	N_1						
		0	1	2	3	4	5	6
3	10	1		3	2	3		1
4	44		6	10	12	10	6	
5	10			2	6	2		
	64	1	6	15	20	15	6	1

Table A8. Distribution of the assembly indices for $N = 7$.

$a^{(7)}(C)$	$|a^{(7)}(C)|$	N_1							
		0	1	2	3	4	5	6	7
4	50	1	5	7	12	12	7	5	1
5	74		2	14	21	21	14	2	
6	4				2	2			
	128	1	7	21	35	35	21	7	1

Table A9. Distribution of the assembly indices for $N = 8$.

$a^{(8)}(C)$	$|a^{(8)}(C)|$	N_1								
		0	1	2	3	4	5	6	7	8
3	4	1				2				1
4	38			9	8	4	8	9		
5	132		8	17	22	40	22	17	8	
6	82			2	26	24	26	2		
	256	1	8	28	56	70	56	28	8	1

Table A10. Distribution of the assembly indices for $N = 9$.

$a^{(9)}(C)$	$|a^{(9)}(C)|$	N_1									
		0	1	2	3	4	5	6	7	8	9
4	24	1	3		3	5	5	3		3	1
5	184		4	17	35	36	36	35	17	4	
6	248		2	19	42	61	61	42	19	2	
7	56				4	24	24	4			
	512	1	9	36	84	126	126	84	36	9	1

Table A11. Distribution of the assembly indices for $N = 10$.

$a^{(10)}(C)$	$|a^{(10)}(C)|$	N_1										
		0	1	2	3	4	5	6	7	8	9	10
4	20	1		3		5	2	5		3		1
5	198		8	22	20	33	32	33	20	22	8	
6	502		2	18	68	108	110	108	68	18	2	
7	288			2	32	62	96	62	32	2		
8	16					2	12	2				
	1024	1	10	45	120	210	252	210	120	45	10	

Table A12. Distribution of the assembly indices for $N = 11$.

$a^{(11)}(C)$	$|a^{(11)}(C)|$	N_1											
		0	1	2	3	4	5	6	7	8	9	10	11
5	184	1	7	14	23	18	29	29	18	23	14	7	1
6	686		4	32	69	104	134	134	104	69	32	4	
7	970			9	69	178	229	229	178	69	9		
8	208				4	30	70	70	30	4			
	2048	1	11	55	165	330	462	462	330	165	55	11	

Table A13. Distribution of the assembly indices for $N = 12$.

$a^{(12)}$	$|a^{(12)}|$	N_1												
		0	1	2	3	4	5	6	7	8	9	10	11	12
4	10	1				3		2		3				1
5	94			13	4	10	12	16	12	10	4	13		
6	1034		12	42	94	141	130	196	130	141	94	42	12	
7	1688			11	106	196	354	354	354	196	106	11		
8	1180				16	143	282	298	282	143	16			
9	90					2	14	58	14	2				
	4096	1	12	66	220	495	792	924	792	495	220	66	12	1

Table A14. $|B^{(5)}| = 10$ balanced strings.

k	$B_k^{(5)}$					$a^{(5)}(B_k)$
1	0	(0	1)	(0	1)	3
2	(0	1)	0	(0	1)	3
3	(0	1)	(0	1)	0	3
4	(1	0)	0	(1	0)	3
5	(1	0)	(1	0)	0	3
6	0	0	0	1	1	4
7	0	0	1	1	0	4
8	0	1	1	0	0	4
9	1	0	0	0	1	4
10	1	1	0	0	0	4

Table A15. $|B^{(6)}| = 20$ balanced strings.

k	$B_k^{(6)}$						$a^{(6)}(B_k)$
1	(0	1)	(0	1)	(0	1)	3
2	(1	0)	(1	0)	(1	0)	3
3	0	(0	1)	(0	1)	1	4
4	0	(0	1)	1	(0	1)	4
5	(0	1)	0	(0	1)	1	4
6	(0	1)	(0	1)	1	0	4
7	(0	1)	1	0	(0	1)	4
8	(0	1)	1	(0	1)	0	4
9	(1	0)	0	(1	0)	1	4
10	(1	0)	0	1	(1	0)	4
11	(1	0)	(1	0)	0	1	4
12	(1	0)	1	(1	0)	0	4
13	1	(1	0)	0	(1	0)	4
14	1	(1	0)	(1	0)	0	4
15	0	0	1	1	1	0	5
16	0	0	0	1	1	1	5
17	0	1	1	1	0	0	5
18	1	0	0	0	1	1	5
19	1	1	0	0	0	1	5
20	1	1	1	0	0	0	5

Table A16. $|B^{(7)}| = 35$ balanced strings.

k	$B_k^{(7)}$							$a^{(7)}(B_k)$
1	0	(0	1)	(0	1)	(0	1)	4
2	(0	1)	(0	1)	(0	1)	0	4
3	(1	0)	(1	0)	(1	0)	0	4
4	(0	1)	(0	1)	0	(0	1)	4
5	(1	0)	(1	0)	0	(1	0)	4
6	(0	1)	0	(0	1)	(0	1)	4
7	(1	0)	0	(1	0)	(1	0)	4
8	(1	0	0)	(1	0	0)	1	4
9	(1	0	0)	1	(1	0	0)	4
10	1	(1	0	0)	(1	0	0)	4
11	(0	0	1)	1	(0	0	1)	4
12	(0	0	1)	(0	0	1)	1	4
13	1	(0	0)	(0	0)	1	1	5
14	1	0	0	(0	1)	(0	1)	5
15	(1	0)	0	0	1	(1	0)	5
16	(1	0)	(1	0)	0	0	1	5
17	(1	0)	1	(1	0)	0	0	5
18	1	1	(0	0)	(0	0)	1	5
19	1	(1	0)	(1	0)	0	0	5
20	1	1	1	(0	0)	(0	0)	5
21	(0	1)	(0	1)	1	0	0	5
22	(0	1)	1	0	0	(0	1)	5
23	(0	1)	1	0	(0	1)	0	5
24	(0	1)	1	(0	1)	0	0	5
25	(0	1)	0	(0	1)	1	0	5
26	0	(0	1)	(0	1)	1	0	5
27	0	(0	1)	1	(0	1)	0	5
28	(0	0)	1	1	1	(0	0)	5
29	(0	1)	0	0	(0	1)	1	5
30	(0	0)	(0	0)	1	1	1	5
31	0	0	(0	1)	(0	1)	1	5
32	0	0	(0	1)	1	(0	1)	5
33	1	(1	0)	0	0	(1	0)	5
34	0	0	0	1	1	1	0	6
35	0	1	1	1	0	0	0	6

Table A17. $|B^{(8)}| = 70$ balanced strings (1st part).

k	$B^{(8)}_k$								$a^{(8)}(B_k)$
1	((0	1)	(0	1))	((0	1)	(0	1))	3
2	((1	0)	(1	0))	((1	0)	(1	0))	3
3	((0	0)	(1	1))	((0	0)	(1	1))	4
4	((0	1)	(1	0))	((0	1)	(1	0))	4
5	((1	0)	(0	1))	((1	0)	(0	1))	4
6	((1	1)	(0	0))	((1	1)	(0	0))	4
7	(0	0)	(0	0)	(1	1)	(1	1)	5
8	(0	0	1)	(0	0	1)	1	1	5
9	0	(0	1)	(0	1)	(0	1)	1	5
10	0	(0	1)	(0	1)	1	(0	1)	5
11	0	(0	1)	1	(0	1)	(0	1)	5
12	(0	0	1)	1	1	(0	0	1)	5
13	(0	0)	(1	1)	(1	1)	(0	0)	5
14	(0	1)	0	(0	1)	(0	1)	1	5
15	(0	1)	0	(0	1)	1	(0	1)	5
16	(0	1)	(0	1)	0	(0	1)	1	5
17	(0	1)	(0	1)	(0	1)	1	0	5
18	(0	1)	(0	1)	1	0	(0	1)	5
19	(0	1)	(0	1)	1	(0	1)	0	5
20	(0	1	1)	0	0	(0	1	1)	5
21	(0	1)	1	0	(0	1)	(0	1)	5
22	(0	1)	1	(0	1)	0	(0	1)	5
23	(0	1)	1	(0	1)	(0	1)	0	5
24	(0	1	1)	(0	1	1)	0	0	5
25	(1	0	0)	(1	0	0)	1	1	5
26	1	0	(0	1)	(0	1)	(0	1)	5
27	(1	0)	0	(1	0)	1	(1	0)	5
28	(1	0)	0	1	(1	0)	(1	0)	5
29	(1	0	0)	1	1	(1	0	0)	5
30	(1	0	1)	0	0	(1	0	1)	5
31	(1	0)	(1	0)	0	1	(1	0)	5
32	(1	0)	(1	0)	(1	0)	0	1	5
33	(1	0)	(1	0)	1	(1	0)	0	5
34	(1	0)	1	(1	0)	0	(1	0)	5
35	(1	0)	1	(1	0)	(1	0)	0	5
36	(1	1)	(0	0)	(0	0)	(1	1)	5
37	(1	1	0)	0	0	(1	1	0)	5
38	1	1	(0	0	1)	(0	0	1)	5
39	1	1	0	0	1	0	1	0	5
40	1	(1	0)	(1	0)	0	(1	0)	5
41	(1	1	0)	(1	1	0)	0	0	5
42	1	1	0	1	0	1	0	0	5
43	1	1	(1	0	0)	(1	0	0)	5
44	(1	1)	(1	1)	(0	0)	(0	0)	5
45	0	0	(0	1	1)	(0	1	1)	5
46	0	(0	1	1)	(0	1	1)	0	5

Table A18. $|B^{(8)}| = 70$ balanced strings (2nd part).

k	$B^{(8)}_k$								$a^{(8)}(B_k)$
47	0	0	(0	1)	(0	1)	1	1	6
48	0	0	(0	1)	1	1	(0	1)	6
49	0	0	0	(1	1)	(1	1)	0	6
50	0	(0	1)	(0	1)	1	1	0	6
51	0	0	1	1	(1	0)	(1	0)	6
52	(0	1)	0	0	(0	1)	1	1	6
53	(0	1)	0	(0	1)	1	1	0	6
54	(0	1)	(0	1)	1	1	0	0	6
55	(0	1)	1	1	0	0	(0	1)	6
56	(0	1)	1	1	0	(0	1)	0	6
57	(0	1)	1	1	(0	1)	0	0	6
58	0	(1	1)	(1	1)	0	0	0	6
59	1	(0	0)	(0	0)	1	1	1	6
60	(1	0)	0	0	(1	0)	1	1	6
61	1	0	0	(0	1)	1	(0	1)	6
62	(1	0)	0	0	1	1	(1	0)	6
63	(1	0)	(1	0)	0	0	1	1	6
64	(1	0)	1	(1	0)	0	0	1	6
65	(1	0)	1	1	(1	0)	0	0	6
66	1	(1	0)	0	0	(1	0)	1	6
67	1	1	(0	1)	0	0	(0	1)	6
68	1	1	1	(0	0)	(0	0)	1	6
69	1	1	(1	0)	0	0	(1	0)	6
70	1	1	(1	0)	(1	0)	0	0	6

Table A19. $|E^{(5)}| = 2$ balanced ringed strings.

k		$E_k^{(5)}$				$a^{(5)}(E_k)$
1	0	(0	1)	(0	1)	3
6	0	0	0	1	1	4

Table A20. $|E^{(6)}| = 4$ balanced ringed strings.

k			$E_k^{(6)}$				$a^{(6)}(E_k)$
1	(0	1)	(0	1)	(0	1)	3
3	0	(0	1)	(0	1)	1	4
4	0	(0	1)	1	(0	1)	4
16	0	0	0	1	1	1	5

Table A21. $|E^{(7)}| = 5$ balanced ringed strings.

k				$E_k^{(7)}$				$a^{(7)}(E_k)$
1	0	(0	1)	(0	1)	(0	1)	4
12	(0	0	1)	(0	0	1)	1	4
30	(0	0)	(0	0)	1	1	1	5
31	0	0	(0	1)	(0	1)	1	5
32	0	0	(0	1)	1	(0	1)	5

Table A22. $|E^{(8)}| = 10$ balanced ringed strings.

k				$E_k^{(8)}$					$a^{(8)}(E_k)$
1	((0	1)	(0	1))	((0	1)	(0	1))	3
3	(0	0)	(1	1)	(0	0)	(1	1)	4
7	(0	0)	(0	0)	(1	1)	(1	1)	5
8	0	(0	1)	0	(0	1)	1	1	5
9	0	(0	1)	(0	1)	(0	1)	1	5
10	0	(0	1)	(0	1)	1	(0	1)	5
11	0	(0	1)	1	(0	1)	(0	1)	5
46	0	0	(0	1	1)	(0	1	1)	5
45	0	0	(0	1)	(0	1)	1	1	6
47	0	0	(0	1)	1	1	(0	1)	6

Table A23. Selected balanced ringed strings $|E^{(9)}| = 14$.

k					$E_k^{(9)}$					$a^{(9)}(E_k)$
1	0	((0	1)	(0	1))	((0	1)	(0	1))	4
2	0	((0	0)	(1	1))	((0	0)	(1	1))	5
3	(0	(0	1))	(0	1)	(0	0	1)	1	5
4	(0	(0	1))	(0	0	1)	1	(0	1)	5
5	(0	(0	1))	(0	0	1)	(0	1)	1	5
6	0	(0	0	1)	1	1	(0	0	1)	6
7	0	0	(0	1)	1	(0	1)	(0	1)	6
8	0	0	(0	1)	(0	1)	1	(0	1)	6
9	0	0	(0	1)	(0	1)	(0	1)	1	6
10	0	(0	0	1)	(0	0	1)	1	1	6
11	(0	0)	(0	0)	(1	1)	0	(1	1)	6
12	0	(0	0)	(0	0)	(1	1)	(1	1)	6
13	(0	0)	(0	0)	1	1	1	0	1	7
14	(0	0)	(0	0)	1	0	1	1	1	7

Table A24. $|E^{(10)}| = 26$ balanced ringed strings.

k	$E_k^{(10)}$										$a^{(10)}(E_k)$
1	((0	1)	(0	1))	((0	1)	(0	1))	(0	1)	4
2	0	((0	1)	(0	1))	((0	1)	(0	1))	1	5
3	(0	1)	(1	(0	1)	0)	(1	(0	1)	0)	5
4	(0	(0	1)	1	(0	0	1	1)	(0	1)	5
5	0	((0	1)	0	1)	1	(0	1	0	1)	5
6	0	((1	0)	1	0)	1	(1	0	1	0)	5
7	(0	1)	((0	1)	1	0)	(0	1	1	0)	5
8	(0	(0	1))	(0	1)	(0	0	1)	1	1	6
9	(0	(0	1))	(0	0	1)	1	1	(0	1)	6
10	(0	(0	1))	(0	0	1)	1	(0	1)	1	6
11	(0	(0	1))	(0	0	1)	(0	1)	1	1	6
14	0	(0	0	1	1)	1	(0	0	1	1)	6
15	0	0	((0	1)	1)	(0	1	1)	(0	1)	6
16	0	0	((0	1)	1)	(0	1)	(0	1	1)	6
17	0	0	(0	1	1)	(0	0	1	1)	1	6
19	0	0	(0	1)	((0	1)	1)	(0	1	1)	6
12	(0	0)	0	(1	1)	(1	1)	(0	0)	1	7
13	0	0	(0	1)	1	1	(0	1)	(0	1)	7
18	0	0	(0	1)	(0	1)	1	1	(0	1)	7
20	0	0	(0	1)	(0	1)	(0	1)	1	1	7
21	(0	0)	0	1	(0	0)	(1	1)	(1	1)	7
22	(0	0)	(0	0)	(1	1)	(1	1)	0	1	7
23	(0	0)	(0	0)	(1	1)	1	0	(1	1)	7
24	(0	0)	(0	0)	(1	1)	0	(1	1)	1	7
25	(0	0)	(0	0)	1	0	(1	1)	(1	1)	7
26	(0	0)	(0	0)	0	1	(1	1)	(1	1)	7

Table A25. Selected balanced ringed strings $E^{(11)}$.

k	$E_k^{(11)}$											$a^{(11)}(E_k)$
1	0	(0	1)	((0	1))	(0	1))	(0	1	0	1)	5
2	(0	(0	1)	(0	1))	(0	0	1	0	1)	1	5
3	(0	0)	((0	0)	1	1)	(1	0	0	1	1)	6
4	(0	0	1))	(0	1)	(0	1)	(0	0	1)	1	6
5	(0	0)	(0	0)	(0	0)	(1	1)	(1	1)	1	7
6	(0	0)	(1	1	0)	1	(0	0)	(1	1	0)	7
7	(0	0)	(0	0)	(0	1)	(0	1)	1	1	1	8

Table A26. Selected balanced ringed strings $E^{(12)}$.

k	$E_k^{(12)}$												$a^{(12)}(E_k)$
1	((0	1)	(0	1))	(0	1	0	1)	(0	1	0	1)	4
2	(0	(0	1)	1	(0	1))	(0	0	1	1	0	1)	5
3	((0	1)	1	(0	(0	1)))	((0	1)	1	(0	0	1))	5
4	(0	(0	1)	1)	(0	0	1	1)	(0	1)	(0	1)	6
5	((0	1)	0	(0	1))	(0	1	0	0	1)	1	1	6
6	(0	0	1)	(0	0	1)	(0	0	1)	1	1	1	7
7	(0	0)	(0	0)	(0	0)	(1	1)	(1	1)	(1	1)	7
8	(0	0)	(0	0)	(1	1)	(1	1)	1	(0	0)	1	8
9	(0	0)	(1	0)	(1	1)	(0	0)	(1	1)	(1	0)	8
10	(1	1)	(1	1)	(0	1)	(0	1)	(0	0)	(0	0)	8
11	(1	1)	(1	1)	(0	0)	(0	0)	(1	0)	(1	0)	8

Table A27. Selected balanced ringed strings $E^{(13)}$.

k	$E_k^{(13)}$													$a^{(13)}(E_k)$
1	0	((0	1)	(0	1))	(0	1	0	1)	(0	1	0	1)	5
2	0	((1	0)	0	1	(1	0))	(1	0	0	1	1	0)	6
3	(0	((0	1)	(0	1))	(0	0	1	0	1)	(0	1)	1	6
4	(0	0)	((0	0)	(1	1))	(0	0	1	1)		(1	1)	7
5	(0	0)	((0	0)	(1	1))	(0	0	1	1)	0	(1	1)	7
6	(0	0)	(0	0)	(0	0)	0	(1	1)	(1	1)	(1	1)	8
7	(0	0	(0	1))	(0	0	0	1)	(0	1)	1	1	1	8
8	(0	0)	(0	0)	(0	0)	1	0	(1	1)	(1	1)	1	9

References

1. Marshall, S.M.; Murray, A.R.G.; Cronin, L. A probabilistic framework for identifying biosignatures using Pathway Complexity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2017**, *375*, 20160342. doi:10.1098/rsta.2016.0342.

2. Murray, A.; Marshall, S.; Cronin, L. Defining Pathway Assembly and Exploring its Applications, 2018. arXiv:1804.06972 [cs, math].

3. Marshall, S.M.; Mathis, C.; Carrick, E.; Keenan, G.; Cooper, G.J.T.; Graham, H.; Craven, M.; Gromski, P.S.; Moore, D.G.; Walker, S.I.; Cronin, L. Identifying molecules as biosignatures with assembly theory and mass spectrometry. *Nature Communications* **2021**, *12*, 3033. doi:10.1038/s41467-021-23258-x.

4. Liu, Y.; Mathis, C.; Bajczyk, M.D.; Marshall, S.M.; Wilbraham, L.; Cronin, L. Exploring and mapping chemical space with molecular assembly trees. *Science Advances* **2021**, *7*, eabj2465. doi:10.1126/sciadv.abj2465.

5. Marshall, S.M.; Moore, D.G.; Murray, A.R.G.; Walker, S.I.; Cronin, L. Formalising the Pathways to Life Using Assembly Spaces. *Entropy* **2022**, *24*, 884. doi:10.3390/e24070884.

6. Sharma, A.; Czégel, D.; Lachmann, M.; Kempes, C.P.; Walker, S.I.; Cronin, L. Assembly theory explains and quantifies selection and evolution. *Nature* **2023**, *622*, 321–328. doi:10.1038/s41586-023-06600-9.

7. Jirasek, M.; Sharma, A.; Bame, J.R.; Mehr, S.H.M.; Bell, N.; Marshall, S.M.; Mathis, C.; Macleod, A.; Cooper, G.J.T.; Swart, M.; Mollfulleda, R.; Cronin, L. Determining Molecular Complexity using Assembly Theory and Spectroscopy, 2023. arXiv:2302.13753 [physics, q-bio].

8. Łukaszyk, S. A No-go Theorem for Superposed Actions (Making Schrödinger's Cat Quantum Nonlocal). In *New Frontiers in Physical Science Research Vol. 3*; Purenovic, D.J., Ed.; Book Publisher International (a part of SCIENCEDOMAIN International), 2022; pp. 137–151. doi:10.9734/bpi/nfpsr/v3/17106D.

9. Qian, K.; Wang, K.; Chen, L.; Hou, Z.; Krenn, M.; Zhu, S.; Ma, X.s. Multiphoton non-local quantum interference controlled by an undetected photon. *Nature Communications* **2023**, *14*, 1480. doi:10.1038/s41467-023-37228-y.

10. Xue, P.; Xiao, L.; Ruffolo, G.; Mazzari, A.; Temistocles, T.; Cunha, M.T.; Rabelo, R. Synchronous Observation of Bell Nonlocality and State-Dependent Contextuality. *Physical Review Letters* **2023**, *130*, 040201. doi:10.1103/PhysRevLett.130.040201.

11. Łukaszyk, S. Shannon Entropy of Chemical Elements. *European Journal of Applied Sciences* **2024**, *11*, 443–458. doi:10.14738/aivp.116.16194.

12. Tran, D.M.; Nguyen, V.D.; Ho, L.B.; Nguyen, H.Q. Increased success probability in Hardy's nonlocality: Theory and demonstration. *Phys. Rev. A* **2023**, *107*, 042210. doi:10.1103/PhysRevA.107.042210.

13. Colciaghi, P.; Li, Y.; Treutlein, P.; Zibold, T. Einstein-Podolsky-Rosen Experiment with Two Bose-Einstein Condensates. *Phys. Rev. X* **2023**, *13*, 021031. doi:10.1103/PhysRevX.13.021031.

14. Watanabe, S. *Knowing and Guessing: A Quantitative Study of Inference and Information*; Wiley, 1969.

15. Watanabe, S. Epistemological Relativity. *Annals of the Japan Association for Philosophy of Science* **1986**, *7*, 1–14. doi:10.4288/jafpos1956.7.1.

16. Walker, S.; Cronin, L. Time is an object (Not a backdrop, an illusion or an emergent phenomenon, time has a physical size that can be measured in laboratories), 2023.

17. Łukaszyk, S., Black Hole Horizons as Patternless Binary Messages and Markers of Dimensionality. In *Future Relativity, Gravitation, Cosmology*; Nova Science Publishers, 2023; chapter 15, pp. 317–374. doi:10.52305/RLIT5885.

18. Łukaszyk, S. Life as the Explanation of the Measurement Problem. *Journal of Physics: Conference Series* **2024**, *2701*, 012124. doi:10.1088/1742-6596/2701/1/012124.

19. Łukaszyk, S. The Imaginary Universe. preprint, PHYSICAL SCIENCES, 2023.

20. de Chardin, P.T. *The Phenomenon of Man*; Harper, New York, 1959.

21. Prigogine, I.; Stengers, I. *Order out of Chaos: Man's New Dialogue with Nature*; Bantam Books, 1984.

22. Melamede, R. Dissipative Structures and the Origins of Life. Unifying Themes in Complex Systems IV; Minai, A.A.; Bar-Yam, Y., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2008; pp. 80–87.

23. Vedral, V. *Decoding Reality: The Universe as Quantum Information*; Oxford University Press, 2010. doi:10.1093/oso/9780198815433.001.0001.

24. Łukaszyk, S. Four Cubes, 2020. doi:10.48550/ARXIV.2007.03782.

25. Vopson, M.M.; Lepadatu, S. Second law of information dynamics. *AIP Advances* **2022**, *12*, 075310. doi:10.1063/5.0100358.

26. Łukaszyk, S. Novel Recurrence Relations for Volumes and Surfaces of n-Balls, Regular n-Simplices, and n-Orthoplices in Real Dimensions. *Mathematics* **2022**, *10*. doi:10.3390/math10132212.

27. Łukaszyk, S.; Tomski, A. Omnidimensional Convex Polytopes. *Symmetry* **2023**, *15*. doi:10.3390/sym15030755.

28. Vopson, M.M. The second law of infodynamics and its implications for the simulated universe hypothesis. *AIP Advances* **2023**, *13*, 105308. doi:10.1063/5.0173278.

29. Łukaszyk, S. Metallic Ratios and Angles of a Real Argument. *IPI Letters* **2024**, pp. 26–33. doi:10.59973/ipil.55.

30. Shannon, C.E. A Mathematical Theory of Communication. *Bell System Technical Journal* **1948**, *27*, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

31. Bekenstein, J.D. Black holes and the second law. *Lettere Al Nuovo Cimento Series 2* **1972**, *4*, 737–740. doi:10.1007/BF02757029.

32. Bekenstein, J.D. Black Holes and Entropy. *Phys. Rev. D* **1973**, *7*, 2333–2346. doi:10.1103/PhysRevD.7.2333.

33. Hawking, S.W. Particle creation by black holes. *Communications In Mathematical Physics* **1975**, *43*, 199–220. doi:10.1007/BF02345020.

34. Hooft, G.t. Dimensional Reduction in Quantum Gravity, 1993. doi:10.48550/ARXIV.GR-QC/9310026.

35. Verlinde, E. On the origin of gravity and the laws of Newton. *Journal of High Energy Physics* **2011**, *2011*, 29. doi:10.1007/JHEP04(2011)029.

36. Misner, C.W.; Thorne, K.S.; Wheeler, J.A. *Gravitation*; W. H. Freeman: San Francisco, 1973.

37. Gould, A. Classical derivation of black-hole entropy. *Physical Review D* **1987**, *35*, 449–454. doi:10.1103/PhysRevD.35.449.

38. Penrose, R.; Floyd, R.M. Extraction of Rotational Energy from a Black Hole. *Nature Physical Science* **1971**, *229*, 177–179. doi:10.1038/physci229177a0.

39. Christodoulou, D.; Ruffini, R. Reversible Transformations of a Charged Black Hole. *Physical Review D* **1971**, *4*, 3552–3555. doi:10.1103/PhysRevD.4.3552.

40. Stuchlík, Z.; Kološ, M.; Tursunov, A. Penrose Process: Its Variants and Astrophysical Applications. *Universe* **2021**, *7*, 416. doi:10.3390/universe7110416.

41. Downey, P.; Leong, B.; Sethi, R. Computing Sequences with Addition Chains. *SIAM Journal on Computing* **1981**, *10*, 638–646, [https://doi.org/10.1137/0210047]. doi:10.1137/0210047.

42. Chaitin, G.J. Randomness and Mathematical Proof. *Scientific American* **1975**, *232*, 47–52. doi:10.1038/scientificamerican0575-47.

43. Chaitin, G.J. *The unknowable*; Springer series in discrete mathematics and theoretical computer science, Springer: Singapore ; New York, 1999.

44. Rajput, C. Metallic Ratios in Primitive Pythagorean Triples : Metallic Means embedded in Pythagorean Triangles and other Right Triangles. *JOURNAL OF ADVANCES IN MATHEMATICS* **2021**, *20*, 312–344. doi:10.24297/jam.v20i.9088.

45. Kolmogorov, A. On tables of random numbers. *Theoretical Computer Science* **1998**, *207*, 387–395. doi:10.1016/S0304-3975(98)00075-9.

46. Caldarola, F.; d'Atri, G.; Maiolo, M.; Pirillo, G. New algebraic and geometric constructs arising from Fibonacci numbers: In honor of Masami Ito. *Soft Computing* **2020**, *24*, 17497–17508. doi:10.1007/s00500-020-05256-1.

47. Chaitin, G. *From Philosophy to Program Size: Key Ideas and Methods : Lecture Notes on Algorithmic Information Theory from the 8th Estonian Winter School in Computer Science, EWSCS'03 : [2-7 March]*; Institute of Cybernetics at Tallinn Technical University, 2003.

48. Chaitin, G.J. Computational complexity and Gödel's incompleteness theorem. *ACM SIGACT News* **1971**, pp. 11–12. doi:10.1145/1247066.1247068.

49. Zatorski, W. *W życiu chodzi o życie*, wydanie i ed.; Tyniec Wydawnictwo Benedyktynów: Kraków, 2020. OCLC: 1243002797.

50. Cronin, Leroy. Lee Cronin: Controversial Nature Paper on Evolution of Life and Universe | Lex Fridman Podcast #404, 2023. Accessed: 2023-12-18.

51. Barta, J.; Markiewicz, R., Eds. *Prawo autorskie: przepisy, orzecznictwo, umowy miedzynarodowe*, wyd. 4., rozsz. i zaktualizowane ed.; Dom Wydawniczy ABC: Warszawa, 2002.

52. Mandelstam, L.; Tamm, I. The Uncertainty Relation Between Energy and Time in Non-relativistic Quantum Mechanics. *J. Phys. (USSR)* **1945**, *9*, 249—-254.

53. Margolus, N.; Levitin, L.B. The maximum speed of dynamical evolution. *Physica D: Nonlinear Phenomena* **1998**, *120*, 188–195. doi:10.1016/S0167-2789(98)00054-2.

54. Levitin, L.B.; Toffoli, T. Fundamental Limit on the Rate of Quantum Dynamics: The Unified Bound Is Tight. *Physical Review Letters* **2009**, *103*, 160502. doi:10.1103/PhysRevLett.103.160502.

55. Hossenfelder, S. Comments on and Comments on Comments on Verlinde's paper "On the Origin of Gravity and the Laws of Newton", 2010. arXiv:1003.1015 [gr-qc].

56. Landauer, R. Irreversibility and Heat Generation in the Computing Process. *IBM Journal of Research and Development* **1961**, *5*, 183–191. doi:10.1147/rd.53.0183.

57. Łukaszyk, S. A new concept of probability metric and its applications in approximation of scattered data sets. *Computational Mechanics* **2004**, *33*, 299–304. doi:10.1007/s00466-003-0532-2.

58. Castro, P.S.; Kastner, T.; Panangaden, P.; Rowland, M. MICo: Improved representations via sampling-based state similarity for Markov decision processes. *Advances in Neural Information Processing Systems* **2021**, *34*, 30113–30126. doi:10.48550/ARXIV.2106.08229.