Article

Not peer-reviewed version

Assembly Theory - Formalizing Assembly Spaces and Discovering Patterns and Bounds

Wawrzyniec Bieniawski , Piotr Masierak , Andrzej Tomski , Szymon Łukaszyk *

Posted Date: 10 January 2025

doi: 10.20944/preprints202409.1581.v9

Keywords: assembly theory; information theory; complexity measures; information entropy; emergent dimensionality; mathematical physics

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

Article

Assembly Theory - Formalizing Assembly Spaces and Discovering Patterns and Bounds

Wawrzyniec Bieniawski [1], Piotr Masierak [1], Andrzej Tomski [2] and Szymon Łukaszyk [1,*]

1 Łukaszyk Patent Attorneys, ul. Głowackiego 8, 40-052 Katowice, Poland
2 University of Silesia, Institute of Mathematics, Bankowa 14, 40-007 Katowice, Poland
* Correspondence: szymon@patent.pl

Abstract: Assembly theory bridges the gap between evolutionary biology and physics by providing a framework to quantify the generation and selection of novelty in biological systems. We formalize the assembly space as an acyclic digraph of strings with 2-in-regular assembly steps vertices and provide a novel definition of the assembly index. In particular, we show that the upper bound of the assembly index depends quantitatively on the number b of unit-length strings, and the longest length N of a string that has the assembly index of $N-k$ is given by $N_{(N-1)} = b^2 + b + 1$ and by $N_{(N-k)} = b^2 + b + 2k$ for $2 \leq k \leq 9$. We also provide particular forms of such maximum assembly index strings. For $k = 1$, such odd-length strings are nearly balanced, and there are four such different strings if $b = 2$ and seventy-two if $b = 3$. We also show that each k copies of an n-plet contained in a string decrease its assembly index at least by $k(n-1) - a$, where a is the assembly index of this n-plet. We show that the minimum assembly depth satisfies $d_{\text{min}}^{(N)} = \lceil \log_2(N) \rceil$, for all b, and is the assembly depth of a maximum assembly index string. We also provide the general formula for the lengths of the minimum assembly index strings having only one independent assembly step in their assembly spaces. Since these results are also valid for $b = 1$, assembly theory subsumes information theory.

Keywords: assembly theory; information theory; complexity measures; information entropy; emergent dimensionality; mathematical physics

---

1. Introduction

There is an ordered non-physical latent space of patterns that can be studied systematically and are not dependent on the constants of physics [1,2] and the probability is zero that any perceptual system has been shaped by natural selection to represent the true structure of an observer-independent world [3,4], which is, in fact, provably impossible to exist [5]. These patterns underpin not only the geometry of Platonic solids and polytopes in complex dimensions [6] but also the periodic table of elements [7], for example. The presence of these patterns makes the Fresnel equation for the normal incidence of EMR have the same form as the Euclid formula to generate Pythagorean triples [8], connect Pythagorean triples with the relativistic law for the addition of velocities [9,10], make Pythagorean triples define metallic ratios of rational argument [11], and so on.

Assembly Theory (AT), discovered in 2017, provides a structured framework for explaining the evolution of those patterns and understanding complex systems. Remarkably, it does so only by introducing the concepts of an assembly pool and an assembly step leading to a new item by joining a pair of items taken from a set of predefined basic items and items assembled in previous steps [12]. A wealth of results and insights related to AT can be found in the literature [12–22].

Here, we explored this latent space of patterns, extending the results of our previous study [20] on bitstrings to strings $C_k^{(N,b)}$ (we often write them simply as C_k) of length N made of b distinct unit length strings (basic symbols c) and strings (doublets, triplets, quadruplets,...,n-plets) assembled in previous steps. In fact, any embodiment of AT, with basic symbols representing LEGO® blocks, chemical bonds, graphs, monomers, etc. assembled in any space corresponds to the string AT version. This is because

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

in AT an assembly step always consists in joining two parts only, which can be thought of as the left and right fragments of the newly formed string. The ancient Greek verb *symbállein* means putting only two *things* ("symbols") together [23]. Put simply, AT explains and quantifies selection and evolution [18] but it is through the word (aka string or *message*), in particular a nucleotide sequence in the case of $b = 4$, all AT *things* come into existence [24]. In evolutionary biology, natural selection explains the survival and prevalence of certain traits, but it does not address the mechanisms for generating novel phenotypic variants. Traditional physics, while offering predictive power from past initial conditions to future states, lacks a functional perspective necessary to differentiate meaningful novelty from random fluctuations.

The paper is organized as follows. Section 2 introduces definitions and basic theorems used in the paper. Section 3 shows certain relations between the minimum assembly index, assembly depth, depth index, and Shannon entropy of the minimum assembly index bitstrings. Section 4 derives the bounds of the maximum assembly index as a function of a string length and the number of basis symbols. Section 5 presents a method of constructing a string having the maximum assembly index by maximizing the number of independent assembly steps. Finally, Section 6 summarizes and discusses the findings of this study.

2. Rudiments

The Definition 1 and Theorems 1 and 2 were already stated in our previous studies [20,22]. We restate them here for clarity.

Definition 1 (Assembly Space)**.** *An assembly space $\Omega = (C, E, \phi)$ is an acyclic digraph of strings $C = \{C_k\}, k \in \mathbb{N}$, where all $b \in \mathbb{N}$ unit length strings (basic symbol(s)) are inaccessible source vertices and the remaining strings are 2-in-regular assembly steps vertices, E is a set of edges, and $\phi : E \ni e \to C_k \in C$ is an edge labeling map, wherein an assembly step $s > 0$ consists of forming a new string C_z from two, not necessarily different, $s - 1 + b$ strings C_x, C_y by concatenating them with each other, establishing edges $e = (C_x, C_z)$ and $e' = (C_y, C_z)$, and assigning, strings C_x, C_y to edges e', e using the map ϕ as*

$$\begin{aligned} C_z = C_x \circ C_y = strcat(C_x, C_y) \quad &\Leftrightarrow \quad \phi(e) = C_y \wedge \phi(e') = -C_x, \\ C_z = C_y \circ C_x = strcat(C_y, C_x) \quad &\Leftrightarrow \quad \phi(e) = -C_y \wedge \phi(e') = C_x, \end{aligned} \tag{1}$$

where "$\circ$" denotes the string concatenation (strcat) operator.

In other words, the edge labeling map (1) has the following property

$$\forall e = (C_x, C_z) \in E(\Omega),\, \phi(e) = \begin{cases} C_y & \Rightarrow \quad \exists! e' = (C_y, C_z) \in E(\Omega) \,:\, \phi(e') = -C_x \quad \wedge C_z = C_x \circ C_y, \\ -C_y & \Rightarrow \quad \exists! e' = (C_y, C_z) \in E(\Omega) \,:\, \phi(e') = C_x \quad \wedge C_z = C_y \circ C_x, \end{cases} \tag{2}$$

that preserves the commutativity of the assembly step, defines the concatenation order of the strings C_x, C_y in the string C_z being the endpoint of both edges e and e', as - in general - for different strings $C_x \neq C_y \Leftrightarrow C_x \circ C_y \neq C_y \circ C_x$. Although the notion of a *concatenation direction* is pointless for one symbol only, we consider such a degenerate case here.

Although all the Ω vertices are strings, it is convenient to separate this set into a set $B := C \setminus \{C_k^{(N,b)} \in C : N \neq 1\}$ of inaccessible source vertices, and a set $S := C \setminus \{C_k^{(N,b)} \in C : N = 1\}$ of 2-in-regular assembly steps vertices, associating them with labels $\{1, 2, \ldots, |S|\}$. At each assembly step s, the cardinality of the set S of assembly steps vertices increases by one. The relation (1) (based on the map discussed in [25]) is superfluous if the vertices defining the directed edges of Ω are strings, as any edge $e = (C_x, C_z)$ unambiguously resolves to either $e = (C_x, C_x \circ C_y)$ or $e = (C_x, C_y \circ C_x)$. For example, the edge $e = ([010], [0101])$ unambiguously resolves to $e = ([010], [010] \circ [1])$. However, we leave it in Definition 1 for clarity.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

Definition 1 is consistent: all vertices are unique (in any standard graph, all vertices should be unique) and all are strings. Since an assembly step always consists of joining two parts only [12], this can be thought of as the left and right fragments of the newly formed string, and those strings that can be the result of concatenation of two shorter strings are assembly step 2-in-regular vertices, while unit-length strings are inaccessible. Remarkably, the uniqueness of each vertex is a sufficient criterion to establish the admissibility of an assembly step and to introduce the notion of an assembly pool. Vertices (strings) present in the assembly space can not be *assembled again* as new vertices of Ω, as they would not be unique.

Definition 2 (String Assembly Space)**.** *An assembly space Ω_{C_s} of a string C_s is the assembly space 1 containing the vertex C_s and all the vertices leading to the string C_s.*

There can be more than one assembly space of a string reflecting different assembly pathways leading to this string. The Definition 2 of the string assembly space provides a novel definition of the assembly index.

Definition 3 (Assembly Index)**.** *The assembly index (ASI) $a^{(N,b)}(C_s)$ of a string $C_s^{(N,b)}$ is the minimum cardinality $|S(\Omega_{C_s})|$ of the set of the assembly step vertices $S(\Omega_{C_s})$ of all assembly spaces Ω_{C_s} of the string C_s.*

Theorem 1. *For all b a quadruplet is the shortest string that allows for more than one ASI.*

Proof. $N = 2$ provides b^2 available doublets with unit ASI. $N = 3$ provides b^3 available triplets with ASI equal to two. Only $N = 4$ provides b^4 quadruplets that include b^2 quadruplets with ASI equal to two, that is b quadruplets $C_{k,\min}^{(4,b)} = [* * **]$ and $b(b-1)$ quadruplets $C_{l,\min}^{(4,b)} = [* \star * \star]$, while the ASI of the remaining $b^4 - b^2$ quadruplets is three. □

For example, to assemble the quadruplet $C_{k,\min}^{(4,4)} = [0202]$, we need to assemble the doublet [02] and reuse it, while there is nothing available to reuse, in the case of the quadruplet $C_l^{(4,4)} = [0123]$.

Where the symbol value can be arbitrary, we write $*$ assuming that it is the same within the string. If we allow for the 2nd possibility different from $*$, we write $\star$. Thus, $C_k^{(2,b)} = [**]$, for example, is a placeholder for all b strings, while $C_l^{(2,b)} = [*\star]$ a placeholder for all $b(b-1)$ strings.

Theorem 2. *For all b the minimum ASI $a^{(N)}(C_{min})$ as a function of N corresponds to the shortest addition chain for N (OEIS A003313).*

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

$$C_k^{(2^s,b)} = [* * \ldots], \quad C_l^{(2^s,b)} = [* \star * \star \ldots], \tag{3}$$

and the assembly space of each of the strings (3) is unique. At each assembly step, its length doubles.

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 $1 = a_0 < a_1 < \cdots < a_s = N$ of integers such that $\forall\, j \geq 1$, $a_j = a_k + a_l$ for $k \leq l < j$. Hence, $j = 1 \implies k = l = 0$ and the first step in forming an addition chain for N is always $a_1 = 1 + 1 = 2$, which is an equivalent of saying that the ASI of any doublet is one. The second step in forming an addition chain can be $a_2 = 1 + 1 = 2$, $a_2 = 1 + 2 = 3$, or $a_2 = 2 + 2 = 4$. The 1st case does not represent the shortest addition chain but the first step, the 2nd one corresponds to assembling a triplet based on the previously assembled doublet, and the 3rd one corresponds to assembling a minimum ASI quadruplet (3) from this doublet. Maximum ASI quadruplet can be assembled in a third

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

step $a_3 = 3 + 1 = 4$, which corresponds to joining a basic symbol to a triplet. Therefore, four is the smallest number achievable in two ways according to Theorem 1.

Thus, finding the shortest addition chain for N corresponds to finding the ASI of a string containing basic symbols and/or doublets and/or triplets containing these doublets for $N \neq 2^s$ since due to Theorem 1 only they provide the same assembly indices $\{0, 1, 2\}$ with no internal repetitions. □

The assembly spaces of strings $a_{\min}^{(N)}$ of length $N \neq 2^s$ are not unique. For example, a string $C_{\min}^{(5,b)} = [01010]$ can be assembled in three steps from four assembly spaces with $S(\Omega) = \{[01], [010]\}$, $S(\Omega) = \{[01], [0101]\}$, $S(\Omega) = \{[10], [010]\}$, or $S(\Omega) = \{[10], [1010]\}$.

We note in passing that any shortest addition chain for n starts with one, not zero, as zero is the neutral element of addition. For the same reason, two is considered the smallest prime, as one is the neutral element of multiplication. Hence, the fundamental theorem of arithmetic can be thought of as the shortest multiplication chain for n.

Theorem 3. *The strings $C_{min}^{(2^s,b)}$ can contain at most two distinct symbols if $b > 1$. Other minimum ASI strings of length $N \neq 2^s$ can contain at most three distinct symbols if $b > 2$.*

Proof. Minimum ASI strings of length $N = 2^s$ are formed by joining the newly assembled string to itself, where a clear or mixed doublet is assembled in the first step. Minimum ASI strings of other lengths admit a doublet and a triplet containing this doublet and an additional basic symbol.

To formally prove the first part, we can also use mathematical induction on the assembly step s. If $s = 1$, then the minimum ASI strings $C_{\min}^{(2,b)}$ are doublets of the form $[c_1 c_2]$, where $c_1, c_2 \in B(\Omega)$. If $c_1 = c_2$, the string contains one distinct symbol, and if $c_1 \neq c_2$, the string contains two distinct symbols. In both cases, the string has a form (3) and the number of distinct symbols does not exceed two. Now assume that for some $k \in \mathbb{N}$, all minimum ASI strings $C_{\min}^{(2^k,b)}$ contain at most two distinct symbols. We must show that $C_{\min}^{(2^{k+1},b)}$ also contains at most two distinct symbols. We construct $C_{\min}^{(2^{k+1},b)}$ by joining two identical minimum ASI strings $C_{\min}^{(2^k,b)}$

$$C_{\min}^{(2^k,b)} \circ C_{\min}^{(2^k,b)} = C_{\min}^{(2^{k+1},b)}, \tag{4}$$

with each other. By the inductive hypothesis, each $C_{\min}^{(2^k,b)}$ contains at most two distinct symbols. Therefore, their concatenation also contains at most two distinct symbols. By induction, for all $s \in \mathbb{N}$, the minimum ASI string $C_{\min}^{(2^s,b)}$ contains at most two distinct symbols.

We will now show that other minimum ASI strings of length $N \neq 2^s$ can contain at most three distinct symbols if $b > 2$. We provide the construction of minimum ASI strings with three symbols. In the first step $s = 1$, we assemble a doublet $[c_1 c_2]$ where $c_1, c_2 \in B(\Omega)$ and $c_1 \neq c_2$. Next, we join the existing doublet $[c_1 c_2]$ with a new symbol $c_3 \in\in B(\Omega)$ where $c_3 \notin \{c_1, c_2\}$. This forms a triplet $[c_1 c_2 c_3]$, introducing a third distinct symbol and further increasing the ASI by 1. We continue assembling by joining the longest string formed so far with itself or with previously formed strings, maintaining the minimal ASI increase.

Assume *a contrario* that there exists a minimum ASI string $C_{\min}^{(N,b)}$ of length $N \neq 2^s$ that contains four or more distinct symbols. But, incorporating such a fourth symbol is equivalent to assembling a maximum ASI quadruplet, which contradicts the minimality of $C_{\min}^{(N,b)}$ (only a doublet must be assembled from basic symbols and a triplet must be assembled from a basic symbol and a doublet). Thus, Theorem 3 is proven. □

The strings having non-minimum ASI can contain all symbols. For example, the string [26]

$$C_k = [01234012340123401234], \tag{5}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

has ASI $a^{(20,5)}(C_k) = 6 = a^{(20)}_{\text{min}} + 1$ and contains all five basic symbols $B(\Omega) := \{0,1,2,3,4\}$. We conjecture [20] that the problem of constructing a non-minimum ASI string is NP-hard, the problem of determining the ASI of such string is NP, and hence it is an NP-complete problem.

Another quantity quantifying the complexity of a string is the assembly depth (ASD) defined [27] as

$$d_s^{(N_k,b)}(C_k) := \max\left(d^{(N_l,b)}(C_l), d^{(N_m,b)}(C_m)\right) + 1, \tag{6}$$

where $d_0^{(1,b)}(c) := 0$, and $d^{(N_l,b)}(C_l)$ and $d^{(N_m,b)}(C_m)$ are the ASDs of two substrings C_l, C_m of the string C_k that were joined in step s. For $N > 3$, and if there are more assembly pathways with different depths w_j leading to a string, which happens if at least two independent assembly steps are possible, the minimum pathway depth is the ASD of this string. Hence, the ASD captures the notion of an *independent assembly step*.

Theorem 4. *If an assembly space Ω contains strings having the same (non-zero) ASD they were assembled in independent assembly steps.*

Proof. Without loss of generality (w.l.o.g.) assume *a contrario* that Ω contains two strings C_l, C_m having the same ASD, i.e., $d^{(N_l,b)}(C_l) = d^{(N_m,b)}(C_m) \neq 0$, that were not assembled in independent assembly steps, i.e., that C_m was used in the assembly of C_l along with a basic symbol c in some previous step s. Then

$$d_s^{(N_l,b)}(C_l) = \max\left(d^{(N_m,b)}(C_m), d^{(1,b)}(c)\right) + 1 = d^{(N_m,b)}(C_m) + 1 \neq d^{(N_m,b)}(C_m), \tag{7}$$

which contradicts our assumption and completes the proof. □

In other words, if two strings C_l, C_m in Ω have the same ASD, their assembly pathways are unrelated to each other; by the defining equation (6) neither of them could have been used in the assembly pathway of the other.

Corollary 4.1. *If the ASI and ASD of a string are equal to each other, an assembly space of this string cannot contain independent assembly steps.*

Theorem 5. *For all b the maximum length N of any string that can be assembled with the ASD $d_s^{(N)}$ (6) satisfies*

$$N \leq 2^{d_s^{(N)}}. \tag{8}$$

Proof. Assume *a contrario* that $N > 2^{d_s^{(N)}}$. Then for the ASD $d_s^{(N)} = 0$, we have $N > 2^0 = 1$ which is a contradiction as all basic symbols c are unit-length strings and $N = 1$. Similarly, for $d_s^{(N)} = 1$, $N > 2$ is also contradiction in the case of doublets, and so on. This is a consequence of the ASD Definition (6). □

Theorem 6. *For all b the minimum ASD as a function of a string length N, is given by*

$$d_{min}^{(N)} = \lceil \log_2(N) \rceil, \tag{9}$$

where $\lceil x \rceil$ denotes the ceiling function.

Proof. $d_s^{(N)} \geq \log_2(N)$ follows from the relation (8). $d^{(2)}_{\text{min}} = \lceil \log_2(2) \rceil = 1$ satisfies both the definition (6) and our hypothesis (9). Similarly $N = 3$. Using induction on length N, assume that for some

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

$N > 3$, we can assemble a minimum ASD string with ASD (9). We need to show that for $N+1$, we can assemble a string with the ASD satisfying

$$d_{\text{min}}^{(N+1)} = \lceil \log_2(N+1) \rceil. \tag{10}$$

Since, by definition (6), the ASD as a function of N is monotonously nondecreasing and can increase at most by one between N and $N+1$, we have

$$d_{\text{min}}^{(N+1)} = \begin{cases} d_{\text{min}}^{(N)} = \lceil \log_2(N) \rceil \\ d_{\text{min}}^{(N)} + 1 = \lceil \log_2(N) \rceil + 1 \end{cases} = \lceil \log_2(N+1) \rceil, \tag{11}$$

where we used relations (9) and (10). Solving the relation (11) for N yields

$$d_{\text{min}}^{(N+1)} = \begin{cases} d_{\text{min}}^{(N)} = s & \text{if} \quad 2^{s-1} < N < 2^s, \\ d_{\text{min}}^{(N)} + 1 = s+1 & \text{if} \quad N = 2^s, \end{cases} \tag{12}$$

and completes the proof. □

The ASD does not have to be a monotonously nondecreasing function of the assembly step. For example

$$[11] \quad d_1 = 1; \quad [110] \quad d_2 = 2; \quad [01] \quad d_3 = 1; \quad [00] \quad d_4 = 1; \quad [0001] \quad d_5 = 2; \quad [0001110] \quad d_6 = 3. \tag{13}$$

We cannot consider the ASD apart from the ASI. For example, the ASD of a string $C_{\text{max}}^{(7,2)} = [0001110]$ is $d_{a_{\text{max}}}^{(7,2)} = \lceil \log_2(7) \rceil = 3$ even though this string can be assembled in six steps with three larger pathway depths $w_6 \in \{4,5,6\}$ as

$$\begin{array}{llll} [00] \quad d_1 = 1, & [00] \quad w_1 = 1, & [00] \quad w_1 = 1, & [00] \quad w_1 = 1, \\ [01] \quad d_2 = 1, & [01] \quad w_2 = 1, & [01] \quad w_2 = 1, & [000] \quad w_2 = 2, \\ [11] \quad d_3 = 1, & [11] \quad w_3 = 1, & [0001] \quad w_3 = 2, & [0001] \quad w_3 = 3, \\ [110] \quad d_4 = 2, & [0001] \quad w_4 = 2, & [00011] \quad w_4 = 3, & [00011] \quad w_4 = 4, \\ [0001] \quad d_5 = 2, & [000111] \quad w_5 = 3, & [000111] \quad w_5 = 4, & [000111] \quad w_5 = 5, \\ [0001110] \quad d_6 = 3, & [0001110] \quad w_6 = 4, & [0001110] \quad w_6 = 5, & [0001110] \quad w_6 = 6. \end{array} \tag{14}$$

Similarly, the ASD of a string $C_{\text{max}}^{(8,2)} = [00011101]$ is $d_{a_{\text{max}}}^{(8,2)} = \lceil \log_2(8) \rceil = 3$ as

$$\begin{array}{llll} [00] \quad d_1 = 1, & [00] \quad w_1 = 1, & [00] \quad w_1 = 1, & [01] \quad w_1 = 1, \\ [01] \quad d_2 = 1, & [01] \quad w_2 = 1, & [01] \quad w_2 = 1, & [001] \quad w_2 = 2, \\ [11] \quad d_3 = 1, & [11] \quad w_3 = 1, & [0001] \quad w_3 = 2, & [0001] \quad w_3 = 3, \\ [0001] \quad d_4 = 2, & [0001] \quad w_4 = 2, & [00011] \quad w_4 = 3, & [00011] \quad w_4 = 4, \\ [1101] \quad d_5 = 2, & [000111] \quad w_5 = 3, & [000111] \quad w_5 = 4, & [000111] \quad w_5 = 5, \\ [00011101] \quad d_6 = 3, & [00011101] \quad w_6 = 4, & [00011101] \quad w_6 = 5, & [00011101] \quad w_6 = 6. \end{array} \tag{15}$$

However, the non-maximum and non-minimum ASI string $C_k^{(8,2)} = [01001011]$ has only two doublets that can be assembled in independent steps. Hence, its ASD cannot be decreased to $\lceil \log_2(8) \rceil = 3$

$$\begin{array}{ll} [01] \quad d_1 = 1, & [01] \quad w_1 = 1, \\ [11] \quad d_2 = 1, & [010] \quad w_2 = 2, \\ [010] \quad d_3 = 2, & [010010] \quad w_3 = 3, \\ [010010] \quad d_4 = 3, & [0100101] \quad w_4 = 4, \\ [01001011] \quad d_5 = 4, & [01001011] \quad w_5 = 5. \end{array} \tag{16}$$

In general, the Ω that contains a 2^d-plet having the ASD d can also contain $\{2^{d-1}+1, 2^{d-1}+2, \ldots, 2^d - 1\}$-plets having the ASD d and based on the shorter n-plets of length $n < 2^{d-1}+1$.

Theorem 7. *For all b the ASD of any maximum ASI string $C_{max}^{(N,b)}$, corresponds to the minimum ASD (9) of Theorem 6, that is*

$$d_{a_{max}}^{(N,b)} = \lceil \log_2(N) \rceil, \tag{17}$$

Proof. Using the property of the ceiling function $n = \lceil x \rceil \Leftrightarrow n-1 < x \leq n$ valid for $n \in \mathbb{N}, x \in \mathbb{R}$, we have

$$d_{a_{\max}}^{(N,b)} = \lceil \log_2(N) \rceil \quad \Leftrightarrow \quad d_{a_{\max}}^{(N,b)} - 1 < \log_2(N) \leq d_{a_{\max}}^{(N,b)}, \tag{18}$$

The non-strict inequality (18) corresponds to the non-strict inequality (8) valid for any N and any ASD. Therefore, we need to prove that the strict inequality $d_{a_{\max}}^{(N,b)} < \log_2(N) + 1$ holds for all $C_{\max}$ strings. Assume, for contradiction, that there exists a maximum ASI string $C_{\max}^{(N,b)}$ such that

$$d_{a_{\max}}^{(N,b)} \geq \log_2(N) + 1 = \log_2(2N) \quad \Longrightarrow \quad 2^{d_{a_{\max}}^{(N,b)}} \geq 2N \quad \Longrightarrow \quad N \leq 2^{d_{a_{\max}}^{(N,b)}-1}. \tag{19}$$

But this relation does not hold for the maximum ASI string $C_{\max}^{(N,b)}$. □

For example, as shown in Figure 1(c,d), the string $C_{\max}^{(15,2)} = [010101000011100]$ has the ASI $a_{\max}^{(15,2)} = 10$ and the ASD $d_{a_{\max}}^{(15,2)} = 4$, while the string $C_{\min}^{(15,2)} = [010010100101001]$ has smaller ASI $a_{\min}^{(15)} = 5$ but larger ASD $d_{a_{\min}}^{(15,2)} = 5$. On the other hand, the ASD of the maximum ASI string $C_{(N-5)}^{(16,2)}$ (E11) and the minimum ASI string (3), shown in Figure 1(a,b), is the same.

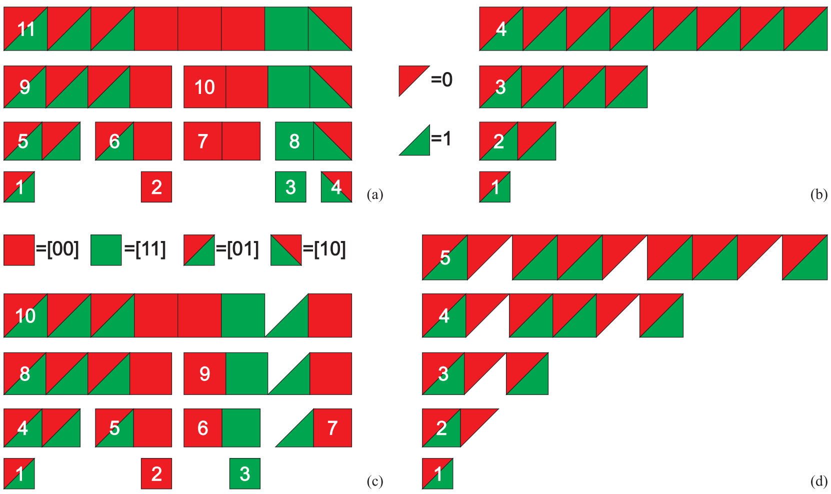

Figure 1. Assembly steps vertices $S(\Omega)$ of assembly spaces of bitstrings $C_{\max}^{(N,2)}$ (a, c) and $C_{\min}^{(N,2)}$ (b, d) for $N = 2^s = 16$ (a, b) and $N = 15 \neq 2^s$ (c, d), where the assembly index is a number in a string (final string for (a, c)) and the assembly depth corresponds to the level. For $N = 2^s$, $d_{a_{\max}}^{(2^s,b)} = d_{a_{\min}}^{(2^s,b)} = s$. In general, for $N \neq 2^s$, the assembly depth $d_{a_{\max}}^{(N,b)} < d_{a_{\min}}^{(N,b)}$. The distributions of n-plets in $C_{\max}^{(N,2)}$ strings is shown in Table 2.

Here, we introduce the following definition, which - as we shall see - is also related to the independent assembly step.

Definition 4 (Depth Index)**.** *We call the number of steps $\hat{a}_{min}^{(N)}$ to reach 1 starting from $N := N_0$ and assigning*

$$N_{s+1} = \begin{cases} N_s - 1 & \text{if} \quad N_s \text{ is odd}, \\ N_s - 2 & \text{if} \quad N_s = 2^s + 2, s \in \mathbb{N}, \\ N_s/2 & \text{otherwise} \end{cases} \tag{20}$$

the depth index (DPI).

The relation (20) yields the same number of steps as the Chandah-sutra method (cf. OEIS A014701) and, unlike the minimum ASI, is an analytical function of N. For example, $\hat{a}_{\min}^{(2^s)} = s$ and $\hat{a}_{\min}^{(2^s-1)} = 2(s-1)$.

We assume that initially a new string of length N is formed in an assembly space based on a basic symbol and a string of length $N-1$. Subsequently, this string assembly space evolves to reduce the cardinality $|S(\Omega_{C_s})|$ of the set of the assembly step vertices until it equals the ASI of this string, that is until $|S(\Omega_{C_s})| = a^{(N,b)}(C_s)$. This assumption is supported by physics. It was shown [28] by equating the binary entropy variation on a holographic sphere with the Bekenstein–Hawking entropy that a black hole can be thought of as a patternless bitstring of length $N_{\text{BH}} \in \mathbb{R}$ having the Hamming weight of $N_1 = \lfloor N_{\text{BH}}/2 \rfloor$ active Planck triangles and - in general - containing at least one fractional triangle having an area smaller than the Planck area and therefore too small to carry a single bit of information. It was further shown [29] that a black hole represents a pure binary quantum state (qubit) in an equal superposition, that is, the only quantum state attaining three known bounds for the quantum orthogonalization interval, and generates entropy variation spheres through the solid angle correspondence that can be thought of as strings of length $N_{\text{VS}} \in \mathbb{R}$, wherein $N_{\text{BH}} \leq N_{\text{VS}} \leq 4N_{\text{BH}}$. Finally, it was shown [20] based on a simple model of binputation that elegant [30] binary assembling programs (i.e., bitstrings) assemble minimum ASI bitstrings of lengths expressible as a product of Fibonacci numbers (OEIS A065108), wherein some binary assembling programs of at least four bits can also assemble non-minimum ASI strings or are not elegant. Hence, we assume that the assembly spaces evolve by reconfiguring the network of edges to decrease the ASD of newly assembled strings, possibly finding shorter pathways for these strings, and if only such a decrease would not result in ASI increase ($N = 15$ shown in Figure 1(d) is the shortest length, where $5 = d_{a_{\min}}^{(15)} > \lceil \log_2(15) \rceil = 4$).

The concepts of assembly space, string assembly space, assembly index and depth, as well as the evolution of assembly spaces are illustrated in Figure 2. The assembly depth naturally divides the lengths of strings into sections $2^{d-1} < N \leq 2^d$.

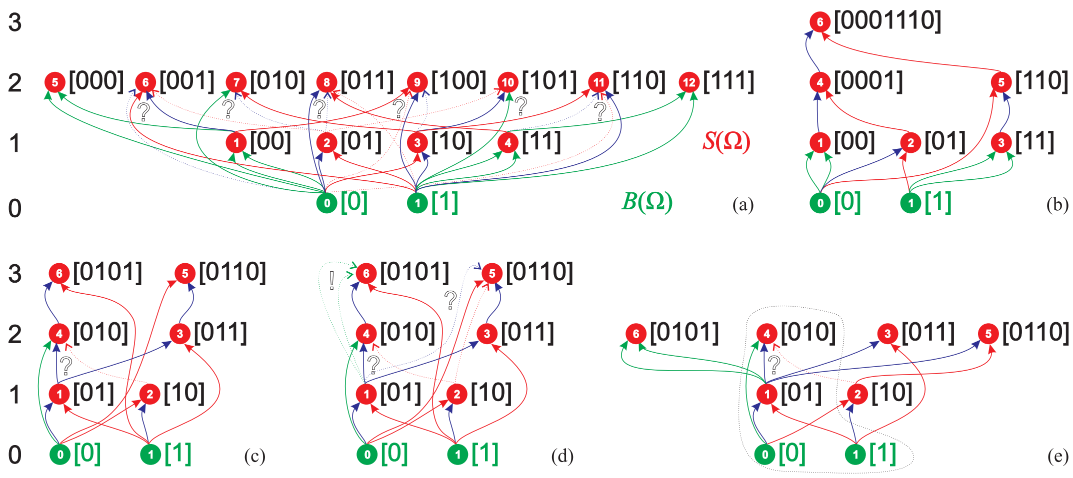

Figure 2. Assembly space Ω, assembly index, and assembly depth. The assembly space of all eight binary triplets with all pathways (a). Via ϕ mapping blue edge provides the 1st string, red edge provides the 2nd string in the assembly step, and the order is irrelevant for two green edges or green edge provides the 1st or 2nd string in dependence of the color of the complementary edge. Dotted edges and question marks indicate alternative pathways (e.g., $\phi([0],[010]) = [10] \wedge \phi([10],[010]) = -[0]$ or $\phi([01],[010]) = [0] \wedge \phi([0],[010]) = -[01]$), showing that the string [01], for example, is unnecessary to construct any triplet. The assembly space of the bitstring $C_6^{(7,2)} = [0001110]$ showing that its assembly index $a^{(7,2)}(C_6) = 6$ (b). The evolution of assembly spaces of strings [0101] and [0110] (c-e). Strings [0101] and [0110] are initially assembled from triplets and basic symbols, increasing the assembly depth (c). New pathways increasing the number of independent assembly steps are found (d), and the edges of Ω are reconfigured, decreasing the number of assembly steps of the string [0101] from three to two steps and the assembly depth of both quadruplets from three to two. Five assembly spaces of the bitstrings [0101], [010] (two alternatives, one encircled), [011], and [0110] (e).

Theorem 8. *A string containing the same three doublets has the same ASI as a string containing two pairs of the same doublets, provided that both strings have the same distributions of other repetitions and have the same lengths.*

Proof. W.l.o.g., consider the following two strings of the same length $N+8$ with $** \neq 01$ and the same distributions of other repetitions (if there are any other repetitions)

$$C_k = [\ldots 01 \ldots 01 \ldots 01 \cdots * \star \ldots], \quad C_l = [\ldots 01 \ldots 01 \ldots 22 \ldots 22 \ldots]. \tag{21}$$

Assembling a doublet takes one assembly step. Each appending of a doublet to an assembled string counts as another assembly step. Hence, in a general case (i.e., for strings C_k, C_l containing also other symbols), the string C_k requires six additional assembly steps, the same as the string C_l, which completes the proof. □

Theorem 9. *A string containing the same three doublets has the same ASI as a string containing the same two triplets, provided that both strings have the same distributions of other repetitions.*

Proof. W.l.o.g. consider the following two strings of the same length $N+6$ with the same distributions of other repetitions

$$C_k = [\ldots 01 \ldots 01 \ldots 01 \ldots], \quad C_l = [\ldots 010 \ldots 010 \ldots]. \tag{22}$$

The assembly of a triplet takes two steps. Hence, in the general case, the string C_k requires four additional assembly steps, the same as the string C_l, which completes the proof. □

Theorem 10. *A string containing the same two triplets has the same ASI as a string containing two pairs of the same doublets, provided that both strings have the same distributions of other repetitions and have the same lengths.*

Proof. The proof comes from Theorems 8 and 9. □

Theorem 11. *A string containing the same two quadruplets of the minimum ASI has the same ASI as a string containing the same three triplets, provided that both strings have the same distributions of other repetitions and have the same lengths.*

Proof. W.l.o.g. consider the following two strings of the same length $N+9$ with the same distributions of other repetitions

$$C_k = [\ldots 0101 \ldots 0101 \cdots \star \ldots], \quad C_l = [\ldots 010 \ldots 010 \ldots 010 \ldots]. \tag{23}$$

The assembly of such a quadruplet takes two steps. Hence, in a general case, the string C_k requires five additional assembly steps, the same as the string C_l, which completes the proof. □

Theorem 12. *A string containing the same two quadruplets of the maximum ASI has the same ASI as a string containing a doublet and the same two triplets based on this doublet, provided that both strings have the same distributions of other repetitions.*

Proof. W.l.o.g. consider the following two strings of the same length $N+8$ with the same distributions of other repetitions

$$C_k = [\ldots 0001 \ldots 0001 \ldots], \quad C_l = [\ldots 110 \ldots 10 \ldots 110 \ldots]. \tag{24}$$

The assembly of such a quadruplet takes three steps. Hence, in a general case, the string C_k requires five additional assembly steps, the same as the string C_l, which completes the proof. □

Theorem 13. *A string containing the same two doublets and the same two triplets not based on this doublet has the same ASI as a string containing a doublet and the same two triplets based on this doublet, provided that both strings have the same distributions of other repetitions and have the same lengths.*

Proof. W.l.o.g. consider the following two strings of the same length $N + 10$ with the same distributions of other repetitions

$$C_k = [\ldots 110 \ldots 00 \ldots 110 \ldots 00 \ldots], \quad C_l = [\ldots 110 \ldots 10 \ldots 110 \cdots * \star \ldots], \tag{25}$$

where $*\star \notin \{11, 10\}$. In a general case, the string C_k requires seven additional assembly steps, the same as the string C_l, which completes the proof. □

In general, Theorems 1-13 show that

- k copies of a doublet in a string decrease the ASI of this string at least by $k - 1$;
- k copies of a triplet in a string decrease the ASI of this string at least by $2k - 2$;
- k copies of a minimum ASI quadruplet in a string decrease the ASI of this string at least by $3k - 2$;
- k copies of a maximum ASI quadruplet in a string decrease the ASI of this string at least by $3k - 3$;

where, the phrase "at least" is meant to indicate that other repetitions, such as e.g. doublets forming multiple quadruplets, etc. can further decrease the ASI of the string. This observation allows us to state the following theorem.

Theorem 14. *Each k_r copies of an n_r-plet $C_r^{(n_r,b)}$ contained in a string $C_m^{(N,b)}$ decrease its ASI at least by $k_r(n_r - 1) - a^{(n_r,b)}(C_r)$. That is*

$$a^{(N,b)}(C_m) \leq N - 1 - \sum_{r=1}^{R} \left[k_r(n_r - 1) - a^{(n_r,b)}(C_r) \right], \tag{26}$$

where R is the total number of repeated n_r-plets.

Proof. W.l.o.g. consider the following string

$$C_m^{(N,b)} = [\ldots [c_1 c_2 \ldots c_n] \ldots [c_1 c_2 \ldots c_n] \ldots], \tag{27}$$

containing two copies of an n-plet $C_l^{(n,b)} = [c_1 c_2 \ldots c_n]$. The n-plet $C_l^{(n,b)}$ can be assembled in at least $a^{(n,b)}(C_l)$ steps and appended to the assembled string C_m in one step. Consider that the ASI of the n-plet $C_l^{(n,b)}$ is $a^{(n,b)}(C_l) = n - 1$, i.e. the n-plet does not have any repetitions that can be reused. Then one copy of this n-plet - as expected - does not decrease the ASI of the string $C_m^{(N,b)}$, as $1(n - 1) - (n - 1) = 0$, while more copies k decrease it by $(n - 1)(k - 1)$. On the other hand, if $a^{(n,b)}(C_l) < n - 1$ then even a single copy of this n-plet will decrease the ASI of C_m. □

For example, due to the presence of three copies of a 5-plet [01001], each with $a^{(5,6)}([01001]) = 3$, in a string

$$C_k^{(24,6)} = [12|01001|21|01001|235|01001|52], \tag{28}$$

its ASI amounts to $a^{(24,6)}(C_k) = 24 - 1 - (3 \cdot (5 - 1) - 3) = 14$. The relation (26) provides the upper bound on ASI as it does not describe a situation in which n-plet for $n > 2$ is assembled based on a doublet also present in one copy in the string. For example, the string $a^{(14,9)}([56101781014301]) = 10$, while $14 - 1 - (2(3 - 1) - 2) = 11$. We note that the maximum ASI decrease is provided by 2^s-plets of the minimum ASI and amounts to $k(n - 1) - \log_2(n) = k(2^s - 1) - s$.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

3. Minimum Assembly Depth, Assembly Depth and Entropy of a Minimum Assembly Index, Minimum Assembly Index, and Depth Index

The minimum ASD as a function of the length of a string $d_{\min}^{(N)}$ (9), the ASD of a minimum ASI string $d_{a_{\min}}^{(N)}$ (which we call here the *minimum ASI ASD*), the minimum ASI as a function of the length of a string $a_{\min}^{(N)}$ (OEIS A003313), and DPI $\hat{a}_{\min}^{(N)}$ (OEIS A014701) define four distinct sets illustrated in Figure 4, wherein $d_{\min}^{(N)} \leq d_{a_{\min}}^{(N)} \leq a_{\min}^{(N)} \leq \hat{a}_{\min}^{(N)}$. We observed certain salient regularities among them.

Theorem 15. *If a minimum ASI string has length $N := 2^s$, $s \in \mathbb{N}_0$, then the minimum ASD, minimum ASI ASD, minimum ASI, and DPI are equal to s.*

Proof. To prove that the minimum ASI ASD equals minimum ASI, we use mathematical induction on the length N of the string. For the base case ($N = 2^0 = 1$), the string consists of a single basic symbol $c \in P_0^{(b)}$. Hence, its ASI is $a_{\min}^{(1)} := 0$ and its ASD $d_{a_{\min}}^{(1)} := 0$. Therefore, $d_{a_{\min}}^{(1)} = a_{\min}^{(1)} = 0$. Assume now that for all strings of length 2^s less than N, the ASD equals the minimum ASI, that is

$$d_{a_{\min}}^{(2^s)} = a_{\min}^{(2^s)} \quad \forall 2^s < N. \tag{29}$$

For some integer s, we construct the minimum ASI string as follows. First, we assemble a doublet from two basic symbols:

$$c_1 \circ c_2 = C^{(2,b)}, \quad c_1, c_2 \in P_0^{(b)}. \tag{30}$$

Its ASI is $a_{\min}^{(2)} = 1$ and its ASD is $d_{a_{\min}}^{(2)} = 1$. Then for each $s \geq 2$ we have $C^{(2^{s-1},b)}$ with the ASI $a_{\min}^{(2^{s-1})} = s - 1$ and the ASD $d_{a_{\min}}^{(2^{s-1})} = s - 1$ and we construct $C^{(2^s,b)}$ by joining two copies of $C^{(2^{s-1},b)}$

$$C^{(2^{s-1},b)} \circ C^{(2^{s-1},b)} = C^{(2^s,b)}. \tag{31}$$

The ASI of the string $C^{(2^s,b)}$ is equal to

$$a_{\min}^{(2^s)} = a_{\min}^{(2^{s-1})} + 1 = (s-1) + 1 = s, \tag{32}$$

and, similarly, its ASD is equal to

$$d_{a_{\min}}^{(2^s)} := \max\left(d_{a_{\min}}^{(2^{s-1})}, d_{a_{\min}}^{(2^{s-1})}\right) + 1 = (s-1) + 1 = s. \tag{33}$$

Therefore, $a_{\min}^{(2^s)} = d_{a_{\min}}^{(2^s)} = s$. At any step, we assemble strings (3), and no two assembly steps can be independent, which follows from Theorem 2. The equation (12) establishes that $N = 2^s$ is the largest N for which $d_{\min}^{(N)} = s$. This proves $d_{\min}^{(2^s)} = d_{a_{\min}}^{(2^s)} = a_{\min}^{(2^s)} = s$. Finally, the even part of the definition of the DPI 4 is the only defining part of this definition iff $N = 2^s$. Hence, $d_{\min}^{(2^s)} = d_{a_{\min}}^{(N)} = a_{\min}^{(2^s)} = \hat{a}_{\min}^{(2^s)} = s$. □

Theorem 15 can be generalized as follows.

Theorem 16. *The minimum ASD, minimum ASI ASD, minimum ASI, and DPI of a minimum ASI string are equal to $s \in \mathbb{N}$ iff $\hat{N}_1 := 2^{s-1} + 2^l$, $l = 0, 1, \ldots, s-1$, $s \geq 1$ or, in other words*

$$\hat{N}_1 := 2^{s-1} + 2^l, \quad l = 0, 1, \ldots, s-1, \quad s \geq 1 \quad \Leftrightarrow \quad d_{min}^{(\hat{N}_1)} = d_{a_{min}}^{(\hat{N}_1)} = a_{min}^{(\hat{N}_1)} = \hat{a}_{min}^{(\hat{N}_1)} = s. \tag{34}$$

Proof. The strings (34) (OEIS A173786 or A048645) are the generalization of the strings of length $N = 2^{s-1} + 2^{s-1} = 2^s$ of the previous Theorem 15. For other lengths of the strings (34), the base case for $s = 2, l = 0$ describes the assembly of a triplet, by joining a symbol to a doublet made in the first step, so that both the ASI and the ASD of this triplet increase by one. And so on. For any s we can join

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

a symbol to a string of length $N = 2^{s-1}$ assembled in $s-1$ steps or join two such strings, as shown in Figure 3(a).

To see that $\hat{a}_{\min}^{(\hat{N}_1)} = s$ (34) holds for $\hat{N}_1 \neq 2^s$ note that there is only one odd part of the definition of the DPI 4 that *restores* $N = 2^s$. For example, we reach one starting from $\hat{N}_1 = 20$ in five steps through $\{20, 10, 5, 4, 2, 1\}$. □

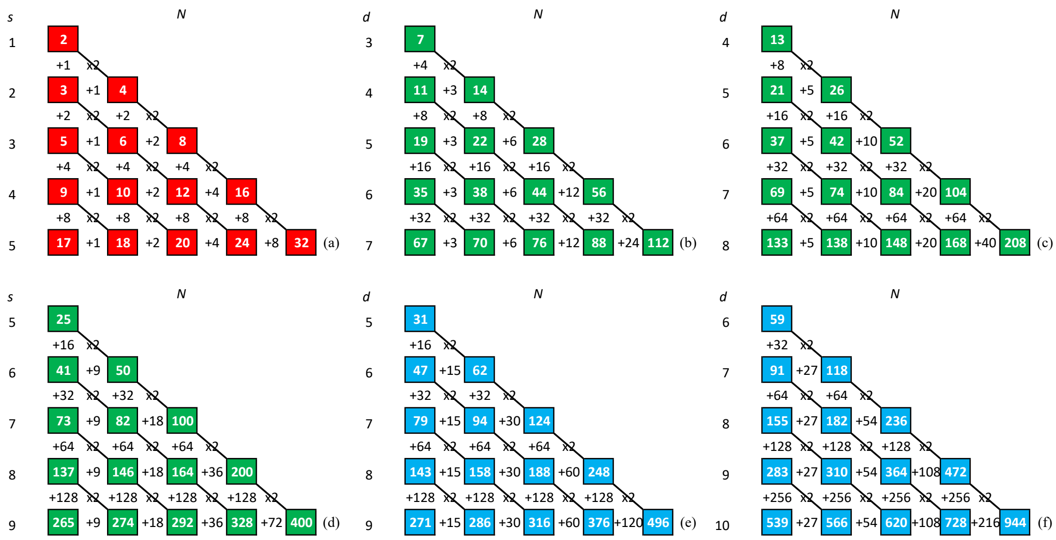

Figure 3. Lengths of all strings having the property of $d_{\min}^{(\hat{N}_1)} = a_{\min}^{(\hat{N}_1)} = d_{a_{\min}}^{(\hat{N}_1)} = \hat{a}_{\min}^{(\hat{N}_1)} = s$ (a). Lengths $\tilde{N}_3$, $\tilde{N}_5$, $\tilde{N}_9$ of certain strings having the property of $a_{\min}^{(\tilde{N}_*)} = d_{\min}^{(\tilde{N}_*)} + 1$ (b-d). Lengths $\tilde{N}_{15}$, $\tilde{N}_{27}$ of certain strings having the property of $a_{\min}^{(\tilde{N}_*)} = d_{\min}^{(\tilde{N}_*)} + 2$ (e,f).

The assembly spaces of other minimum ASI strings can contain independent assembly steps. The first such case occurs for $N = 7$, where, for example, the $S(\Omega)$

$$\begin{aligned} [01] \quad & d_1 = 1, \\ [001], [0101] \quad & d_2 = d_3 = 2, \\ [0010101] \quad & d_4 = 3 \end{aligned} \tag{35}$$

results in a string having ma $a_{\min}^{(7)} = 4$ and $d_{a_{\min}}^{(7)} = \lceil \log_2(7) \rceil = 3$, since both [001] and [0101] were assembled from the doublet [01] in two independent assembly steps at the same depth $d_2 = d_3 = 2$, which is congruent with Theorem 4.

Theorem 17. *The minimum ASI strings [20] (strings (15)) of lengths*

$$\tilde{N}_3 := 2^{d-1} + 3 \cdot 2^l, \quad l = 0, 1, \ldots, d-3, \quad d \geq 3 \quad \Leftrightarrow \quad a_{min}^{(\tilde{N}_3)} = d + 1 = \lceil \log_2(\tilde{N}_3) \rceil + 1, \tag{36}$$

have only one independent assembly step in their assembly spaces, and excluding this step, they are assembled by joining the longest string assembled so far with itself. Therefore, their ASI is one greater than the minimum ASD (9).

Proof. We begin at $d = 3$ by assembling a $C_{\min}^{(7)}$ using a quadruplet and a triplet assembled independently (e.g., using an assembly space (35)) with $a_{\min}^{(7)} = 4$ and $d_{a_{\min}}^{(7)} = \lceil \log_2(7) \rceil = 3$. For $d = 4$, the string (36) $C_{\min}^{(11)}$ can be assembled by joining the string $C_{\min}^{(8)}$ assembled in three steps and the triplet, while the string $C_{\min}^{(14)}$ by joining two strings $C_{\min}^{(7)}$ made in the previous step. For any d, the shortest string (36) $C_{\min}^{(\tilde{N}_3)}$ can be assembled by joining the string $C_{\min}^{(2^{d-1})}$ (3) assembled in $d-1$ steps and the

 doi:10.20944/preprints202409.1581.v9

triplet, while the remaining strings $C_{\min}^{(\tilde{N}_3)}$ - by joining two strings made in a previous step $d-1$, as shown in Figure 3(b). □

Theorem 18. *The minimum ASI strings of lengths*

$$\tilde{N}_5 := 2^{d-1} + 5 \cdot 2^l, \quad l = 0, 1, \ldots, d-4, \quad d \geq 4 \quad \Leftrightarrow \quad a_{min}^{(\tilde{N}_5)} = d + 1 = \lceil \log_2(\tilde{N}_5) \rceil + 1, \tag{37}$$

have only one independent assembly step in their assembly spaces, and excluding this step, they are assembled by joining the longest string assembled so far with itself. Therefore, their ASI is one greater than the minimum ASD (9).

Proof. We begin at $d = 4$ by assembling a $C_{\min}^{(13)}$ through $\{2, 4, (5, 8), 13\}$ with $a_{\min}^{(13)} = d_{\min}^{(13)} + 1 = 5$. For any d, the shortest string (37) $C_{\min}^{(\tilde{N}_5)}$ can be assembled by joining the string $C_{\min}^{(2^{d-1})}$ (3) assembled in $d-1$ steps with the 5-plet assembled in the independent assembly step, while the remaining strings $C_{\min}^{(\tilde{N}_5)}$ - by joining two strings made in a previous step $d-1$, as shown in Fig 3(c). □

Theorem 19. *The minimum ASI strings of lengths*

$$\tilde{N}_9 := 2^{d-1} + 9 \cdot 2^l, \quad l = 0, 1, \ldots, d-5, \quad d \geq 5 \quad \Leftrightarrow \quad a_{min}^{(\tilde{N}_9)} = d + 1 = \lceil \log_2(\tilde{N}_9) \rceil + 1, \tag{38}$$

have only one independent assembly step in their assembly spaces, and excluding this step, they are assembled by joining the longest string assembled so far with itself. Therefore, their ASI is one greater than the minimum ASD (9).

Proof. We begin at $d = 5$ by assembling a $C_{\min}^{(25)}$ with $a_{\min}^{(25)} = d_{\min}^{(25)} + 1 = 6$. For any d, we assemble the shortest strings (38) as

$$\begin{array}{llllllll} \{2, & 4, & 8, & (9, & 16), & 25\}, & & \\ \{\ldots & & & & & 32, & 41\}, & \\ \{\ldots & & & & & & 64, & 73\}, \\ \{\ldots & & & & & & & 128, \quad 137\}, \\ \ldots & & & & & & & \end{array} \tag{39}$$

with one independent assembly step $(9, 16)$ to assemble the string of length $N = 2^{d-1}$ and joining 9-plet at the last step, while the remaining strings $C_{\min}^{(\tilde{N}_9)}$ - by joining two strings made in a previous step $d-1$, as shown in Figure 3(d). □

Theorems 17-19 allow for the following generalization.

Theorem 20. *The minimum ASI strings of lengths*

$$\tilde{N}_{2^n+1} := 2^{d-1} + \left(2^{k-4} + 1\right)2^l, \quad k \geq 5, \ d \geq k-2, \ l = 0, 1, \ldots, d-(k-2), \quad \Leftrightarrow \quad a_{min}^{(\tilde{N}_{2^n+1})} = d + 1 = \lceil \log_2(\tilde{N}_{2^n+1}) \rceil + 1, \tag{40}$$

have only one independent assembly step in their assembly spaces, and excluding this step, they are assembled by joining the longest string assembled so far with itself. Therefore, their ASI is one greater than the minimum ASD (9).

Proof. The lengths of the strings (40) are listed in rows in Table 1 starting after the length of the substring assembled in an independent assembly step marked green. Hence, the first row contains the lengths of strings of Theorem 17 shown on the diagonal of Figure 3(b), and so on. □

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

Table 1. Certain lengths of minimum ASI strings, which are defined by the ASI and the minimum ASI ASD for $2 \leq s \leq 7$.

s	$a_{\min}^{(\tilde{N}_*)} = 1$	$a_{\min}^{(\tilde{N}_*)} = 2$	$a_{\min}^{(\tilde{N}_*)} = 3$	$a_{\min}^{(\tilde{N}_*)} = 4$	$a_{\min}^{(\tilde{N}_*)} = 5$	$a_{\min}^{(\tilde{N}_*)} = 6$	$a_{\min}^{(\tilde{N}_*)} = 7$	$a_{\min}^{(\tilde{N}_*)} = 8$	$a_{\min}^{(\tilde{N}_*)} = 9$	…	$\tilde{N}_{2^n+1}$
2	2	4	3	7	14	28	56	112	224	…	$\tilde{N}_3$
					15	30	60	120	240	480	$\tilde{N}_{3,a}$
						23	46	92	184	368	$\tilde{N}_{3,b}$
3	2	4	8	3	11	22	44	88	176	…	$\tilde{N}_3$
						27	54	108	216	432	$\tilde{N}_{3,a}$
							43	86	172	344	$\tilde{N}_{3,b}$
	2	4	8	5	13	26	52	104	208	…	$\tilde{N}_5$
						45	90	180	360	…	$\tilde{N}_{5,b}$
4	2	4	8	16	3	19	38	76	152	…	$\tilde{N}_3$
							51	102	204	408	$\tilde{N}_{3,a}$
								83	166	332	$\tilde{N}_{3,b}$
	2	4	8	16	5	21	42	84	168	…	$\tilde{N}_5$
							85	170	340	…	$\tilde{N}_{5,b}$
	2	4	8	16	9	25	50	100	200	…	$\tilde{N}_9$
5	2	4	8	16	32	3	35	70	140	…	$\tilde{N}_3$
								99	198	396	$\tilde{N}_{3,a}$
									163	326	$\tilde{N}_{3,b}$
	2	4	8	16	32	5	37	74	148	…	$\tilde{N}_5$
								165	330	…	$\tilde{N}_{5,b}$
	2	4	8	16	32	9	41	82	164	…	$\tilde{N}_9$
	2	4	8	16	32	17	49	98	196	…	$\tilde{N}_{17}$
6	2	4	8	16	32	64	3	67	134	…	$\tilde{N}_3$
									195	390	$\tilde{N}_{3,a}$
										323	$\tilde{N}_{3,b}$
	2	4	8	16	32	64	5	69	138	276	$\tilde{N}_5$
									325	650	$\tilde{N}_{5,b}$
	2	4	8	16	32	64	9	73	146	…	$\tilde{N}_9$
	2	4	8	16	32	64	17	81	162	…	$\tilde{N}_{17}$
	2	4	8	16	32	64	33	97	194	…	$\tilde{N}_{33}$
7	2	4	8	16	32	64	128	3	131	…	$\tilde{N}_3$
										387	$\tilde{N}_{3,a}$
	2	4	8	16	32	64	128	5	133	266	$\tilde{N}_5$
										645	$\tilde{N}_{5,b}$
	2	4	8	16	32	64	128	9	137	…	$\tilde{N}_9$
	2	4	8	16	32	64	128	17	145	…	$\tilde{N}_{17}$
	2	4	8	16	32	64	128	33	161	…	$\tilde{N}_{33}$
	2	4	8	16	32	64	128	65	193	…	$\tilde{N}_{65}$

Theorem 21. *The minimum ASI strings [20] of lengths*

$$\tilde{N}_7 := 2^{d-1} + 7 \cdot 2^{d-4} \in \{15, 30, 60, \dots\}, \quad d \geq 4 \quad \Leftrightarrow \quad a_{min}^{(\tilde{N}_7)} = d_{a_{min}}^{(\tilde{N}_7)} = d_{min}^{(\tilde{N}_7)} + 1 = \hat{a}_{min}^{(\tilde{N}_1)} - 1 = \lceil \log_2(\tilde{N}_7) \rceil + 1, \tag{41}$$

are assembled by joining the longest string assembled so far with itself. Their ASI and ASD are the same, one greater than the minimum ASD (9) and one smaller than the DPI.

Proof. The equality of ASI and ASD of the strings (41) follows from the proof of Theorem 16. Furthermore,

$$2^{d-1} < 2^{d-1} + 7 \cdot 2^{d-4} < 2^d \quad \Longrightarrow \quad 0 < 7 \cdot 2^{d-4} < 2^{d-1} \quad \Longrightarrow \quad 0 < 7 < 8 \quad \forall d, \tag{42}$$

shows that $d_{a_{\min}}^{(\tilde{N}_7)} = \lceil \log_2(\tilde{N}_7) \rceil + 1$. Finally, $\hat{a}_{\min}^{(\tilde{N}_1)} = \lceil \log_2(\tilde{N}_7) \rceil + 2$ follows from the DPI Definition 4: six steps are required to reach one starting from fifteen and additional steps for thirty, sixty, etc., which completes the proof. □

Theorem 21 seems to allow for the following generalization, which we have validated numerically based on the OEIS A003313 sequence for $N \leq 10^5$.

Conjecture 22. *For d, l, and $\tilde{N}_{2^n+1}$ defined by the relation (40), the following holds*

$$\tilde{N}_{2^n+1,a} := \tilde{N}_{2^n+1} + 2^d = 3 \cdot 2^{d-1} + \left(2^{k-4} + 1\right) 2^l \wedge k = 5 \quad \Leftrightarrow \quad a_{min}^{(\tilde{N}_{2^n+1,a})} = d + 2 = \lceil \log_2(\tilde{N}_{2^n+1,a}) \rceil + 1, \tag{43a}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

$$\tilde{N}_{2^n+1,b} := \tilde{N}_{2^n+1} + 2^{d+1} = 5 \cdot 2^{d-1} + \left(2^{k-4}+1\right)2^l \wedge k \in \{5,6\} \quad \Leftrightarrow \quad a_{min}^{(\tilde{N}_{2^n+1,b})} = d+3 = \lceil \log_2(\tilde{N}_{2^n+1,b}) \rceil + 1, \tag{43b}$$

The lengths of the strings (43a) and (43b) are listed in rows in Table 1.

Furthermore, we have numerically validated the following conjecture.

Conjecture 23. *The minimum ASI strings of lengths*

$$\tilde{N}_{15} := 2^{d-1} + 15 \cdot 2^l, \quad l = 0,1,\ldots,d-5, \quad d \geq 5, \tag{44a}$$

$$\tilde{N}_{27} := 2^{d-1} + 27 \cdot 2^l, \quad l = 0,1,\ldots,d-6, \quad d \geq 6, \tag{44b}$$

$$\tilde{N}_{50.9} := 50 \cdot 2^{d-6} + 9 \cdot 2^l, \quad l = 0,1,\ldots,d-6, \quad d \geq 6, \tag{44c}$$

have the property of

$$a_{min}^{(\tilde{N}_*)} = d + 2 = \lceil \log_2(\tilde{N}_*) \rceil + 2. \tag{44d}$$

The shortest strings of length $\tilde{N}_{15}$ (44a) can be assembled with the pathways

$$\begin{array}{llllllll} \{2, & 4, & (5, & 8), & 13, & 26, & 31\} & \\ \{\ldots & & & & & & 39, & 47\}, \\ \{\ldots & & & & & & & 78, \quad 79\}, \\ \ldots \end{array} \tag{45}$$

shown in Figure 3(e); the shortest strings of length $\tilde{N}_{27}$ (44b) can be assembled with the pathways

$$\begin{array}{llllllllll} \{2, & (3, & 4), & 7, & 14, & 28, & 31, & 59\} & & \\ \{\ldots & & & & 14 & 28, & 56, & 84, & 91\} & \\ \{\ldots & & & & 11 & 18, & 36, & 72, & 144, & 155\} \\ \ldots \end{array} \tag{46}$$

shown in Figure 3(f); and for any d, the shortest strings of length $\tilde{N}_{50.9}$ (44c) can be assembled as

$$\begin{array}{llllllllll} \{2, & 4, & 8, & (9, & 16), & 25, & 50, & 59\}, & & \\ \{\ldots & & & & & & & 100, & 109\}, & \\ \{\ldots & & & & & & & & 200, & 209\} \\ \ldots \end{array} \tag{47}$$

The remaining strings of length $\tilde{N}_{15}$, $\tilde{N}_{27}$, and $\tilde{N}_{50.9}$ (44) can be assembled by joining two strings made in a previous step $d-1$.

Strings of lengths (34), (36), and (41), revealed in [20] based on the degree of causation, showed that there are certain regularities among the minimum ASI strings. Here, we extended these results to strings of lengths (40), (43), and (44).

In general, Theorems 16-21 (in particular Theorem 21) and Conjectures 23 and 22 show a peculiar interdependence among the minimum ASD (9), minimum ASI ASD, minimum ASI, and DPI, as shown in Figure 4. In particular, they show that

- the Ω of minimum ASI strings having ASI equal to DPI cannot contain strings assembled in independent assembly steps,
- the Ωs of other minimum ASI strings can contain at least two such strings, and therefore
- the assembly space of a maximum ASI string will tend to maximize the number of strings assembled in independent assembly steps in the Ω, taking into account the saturation of the Ω as it cannot contain more than b^n distinct n-plets, and hence to minimize the possible ASD.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

We note that the difference between the DPI and minimum ASI is, in general, larger than one. The pathways of the minimum ASI strings maximizing the number of independent assembly steps are listed in Table A1 for $N \leq 65$.

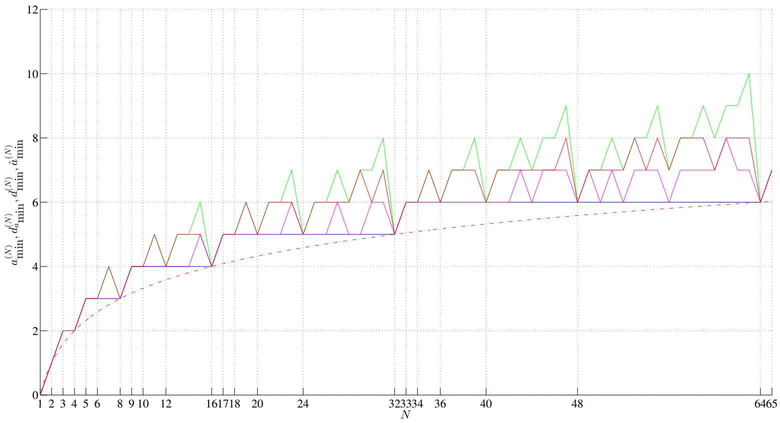

Figure 4. The minimum assembly depth ($\lceil \log_2(N) \rceil$, blue), the assembly depth of the minimum assembly index string (magenta), the minimum assembly index (OEIS A003313, red; $\log_2(N)$, red, dash-dot), and depth index (OEIS A014701, green) for $1 < N \leq 65$.

We have also examined the Shannon entropy

$$H\left(C_{\min}^{(N)}\right) = -p_0 \log_2(p_0) - p_1 \log_2(p_1), \tag{48}$$

of the most balanced minimum ASI bitstrings, where $p_0 = N_0/N$ and $p_1 = N_1/N$ are fractions of the respective symbols $\{0, 1\}$ within the string (N_1 is the Hamming weight). By Theorem 2, the minimum ASI as a function of the length of a string does not depend on b. However, we have chosen the most balanced bitstrings. For $b = 1$, the Shannon entropy vanishes, and the bit is the smallest amount and the quantum of information. Furthermore, by Theorem 3, a string of length $N = 2^s$ can contain at most two distinct symbols (if $b > 1$), and in the case it contains two distinct symbols, it is necessarily the most balanced. The choice of the most balanced bitstrings is also supported by physics [28,29].

Furthermore, following [29], we assumed the Hamming weight $N_1 = \lfloor N/2 \rfloor$. Hence, we first assembled the triplet is [010] or [001] rather than [011], [110], etc., and for the same reason, we preferred the pathway $\{2, 3, 5, 10, 15\}$ (cf. Figure 1(d)) over $\{2, 3, 6, 12, 15\}$, for example, as the string assembled using the former pathway is more balanced ($N_1 = 6$) than the one assembled using the latter one ($N_1 = 5$). Similarly, we preferred the pathway providing a more balanced string over the pathway providing independent assembly steps (cf. Table A1). $N = 14$ is the first exception. $C_{\min}^{(14)}$ assembled in five steps along the pathway $\{2, (3, 4), 7, 14\}$ with the independent assembly steps 3 and 4 has the hamming weight $N_1 = 6$ as compared to $C_{\min}^{(14)}$ assembled in five steps along the pathway $\{2, 4, 8, 12, 14\}$ with no independent assembly steps and the hamming weight $N_1 = N/2 = 7$. The results are listed in Table A1 for $N \leq 65$.

As shown in Figure 5, the Shannon entropy (48) of the most balanced minimum ASI bitstrings rapidly converges to one with exceptions for lengths $N \in \{15, 23, 27, 39, 43, 45, 51, 59, 63, \dots\}$ substantially corresponding to lengths at which DPI is larger than the minimum ASI (cf. Figure 4), which highlights the interdependence among the minimum ASI and DPI.

Theorem 24. *The minimum ASI bitstrings assembled along the pathway given by the DPI 4 and beginning with $C_{min}^{(2)} = [*\star]$ are balanced bitstrings if N is even or nearly balanced bitstrings ($N_0 = N_1 + 1$) if N is odd.*

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

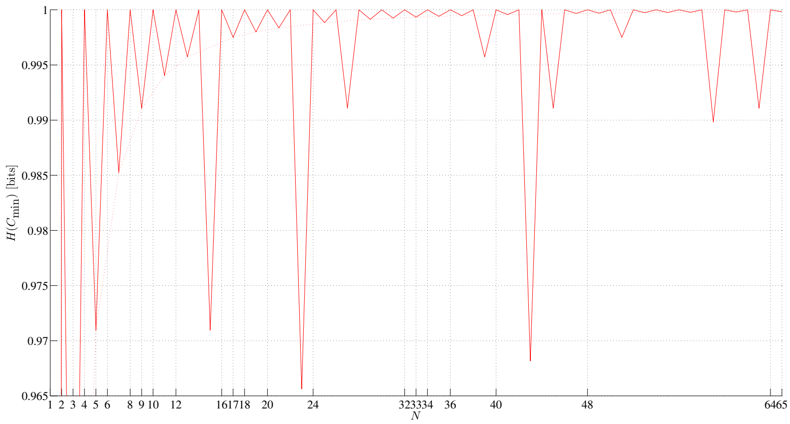

Figure 5. Shannon entropy of the most balanced bitstrings having the minimum assembly index for $1 < N \leq 65$.

Proof. By Theorems 2 and 3, a minimum ASI string of length $N = 2^s$ assembled beginning with $C^{(2)}_{\min} = [\ast\star]$ is a balanced bitstring. To assemble a longer string of other lengths, we assign $N_{s+1} = N_s + 1$ or $N_{s+1} = N_s + 2$. However, the Definition 4 removes the longest string of an odd length $N = 2^s + 1$ from the sequence if only it is not the first one in the sequence. Strings longer than this string of length $N = 2^s + 1$ are assembled by joining the longest string assembled so far with itself ($N_{s+1} = 2N_s$) or by joining a basic symbol chosen to preserve the ballance of the string ($N_{s+1} = N_s + 1$). □

In other words, the Definition 4 removes the *imbalance propagation*. For example, an imbalanced pathway $\{2, 4, 5, 10, 20\}$ ($N_1 = 8$) becomes a balanced pathway $\{2, 4, 8, 10, 20\}$ ($N_1 = 10 = N/2$).

4. Maximum Assembly Index Strings

The seven-bit string is the longest string that can have the maximum ASI $a^{(7,2)}_{\max} = 7 - 1 = 6$. There are four such bitstrings containing two clear triplets and the starting bit at the end or the ending bit at the start, that is

$$[\ast\ast\ast\star\star\star\ast] \quad \text{and} \quad [\star\ast\ast\ast\star\star\star], \tag{49}$$

and their lengths cannot be increased without a repetition of a doublet, which keeps the ASI at the same level $a^{(8,2)}_{\max} = 8 - 2 = 6$.

This observation and Theorem 2 motivated us to develop a general method to construct the longest possible string having the maximum ASI $a^{(N,b)}_{\max}(C_{(N-1)}) = N - 1$, as a function of the radix b. We denote the length of this string by $N_{(N-1)}$ or $N_{(N-1)}(b)$, and we call this string a $C_{(N-1)}$ string.

After a few groping try-outs, we eventually reached two stable methods (cf. Appendices, Methods A and B). In both methods, we start with an initial balanced string of length $3b$ containing b clear triplets ordered as

$$[0001112\ldots(b-2)(b-1)(b-1)(b-1)]. \tag{50}$$

The doublets that can be inserted into the initial string (50) can be arranged in a $b \times b$ matrix

$$\begin{bmatrix} \cancel{00} & \cancel{01} & 02 & \ldots & 0(b-1) \\ 10 & \cancel{11} & \cancel{12} & \ldots & 1(b-1) \\ 20 & 21 & \cancel{22} & \ldots & 2(b-1) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ (b-2)0 & (b-2)1 & (b-2)2 & \ldots & \cancel{(b-2)(b-1)} \\ (b-1)0 & (b-1)1 & (b-1)2 & \ldots & \cancel{(b-1)(b-1)} \end{bmatrix}, \tag{51}$$

where the crossed out entries on a diagonal cannot be reused, as they would form repetitions in this string. Due to the order of triplets in the string (50) we can also cross out the entries in the first

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

superdiagonal of the matrix (51). The strings of odd lengths generated by these general methods are not only the longest but also the most balanced. This can be stated in the following theorem.

Theorem 25. *The longest length of a string that has the ASI of $N-1$ is given by*

$$N_{(N-1)} = 3b + (b-1)^2 = b^2 + b + 1 \tag{52}$$

(Squarefree numbers, OEIS A353887) and this string is nearly balanced, that is

$$N_{(N-1)} = bN_c + 1, \tag{53}$$

where $N_c = b+1$ is the number of occurrences of all but one symbol within the string, and its Shannon entropy is

$$\begin{aligned} H(C_{(N-1)}) = -\sum_{c=0}^{b-1} p_c \log_2(p_c) = -(b-1)\frac{N_{(N-1)}-1}{bN_{(N-1)}} \log_2\left(\frac{N_{(N-1)}-1}{bN_{(N-1)}}\right) - \frac{N_{(N-1)}-1+b}{bN_{(N-1)}} \log_2\left(\frac{N_{(N-1)}-1+b}{bN_{(N-1)}}\right) = \\ = \frac{1-b^2}{b^2+b+1} \log_2\left(\frac{b+1}{b^2+b+1}\right) - \frac{b+2}{b^2+b+1} \log_2\left(\frac{b+2}{b^2+b+1}\right) \lesssim \log_2(b). \end{aligned} \tag{54}$$

The proof of Theorem 25 is given in Appendix D. A $C_{(N-1)}$ string must contain all clear triplets and all doublets and if it is generated by Method A or B it is terminated with 0 and has a form

$$C_{(N-1)} = [000111222\ldots0]. \tag{55}$$

Although the case for $b=1$ is degenerate, as no information can be conveyed using only one symbol ($H(C_{(N-1)}) = 0$ in this case), nothing precludes the assembly of such defunct strings and the formula (52) yields the correct result; the string [000] is the longest string with $a_{\max}^{(N,1)} = N-1$ by Theorem 1, as for $b=1$ the upper and the lower bound on the ASI are the same, $a_{\max}^{(N,1)} = a_{\min}^{(N)}$ (OEIS A003313). This is the only case where the maximum ASI is not a monotonically nondecreasing function of N.

For $b=3$, only two doublets can be introduced without repetitions into the initial string (50), leading to twelve unique strings of length $N_{(N-1)} = 13$

$$\begin{aligned} &[000111222|0210], [000111222|1020], [20|21|000111222], [21|02|000111222], [0001112|02|22|10], [0001112|10|22|20], \\ &[21|000|20|111222], [000|20|111222|10], [02|000111222|10], [20|00|21|0111222], [21|0001112|02|22], [21|000111222|02]. \end{aligned} \tag{56}$$

Finally, we have to multiply the cardinality of this set by $3! = 6$ to account for permutations. For example, the first string [0001112220210], is equivalent to five strings [0002221110120], [1110002221201], [1112220001021], [2220001112102], and [2221110002012]. Hence, there are seventy-two different strings of length $N_{(N-1)}(3) = 13$.

Subsequently, we considered other $C_{(N-k)}$ strings of length $N_{(N-k)}$ with the maximum ASI $a_{\max}(C_{(N-k)}) = N-k$ for $k > 1$.

Theorem 26. *For all $b > 1$ and $2 \le k \le 9$ the longest length of a string that has the ASI of $N-k$ is given by*

$$N_{(N-k)} = b^2 + b + 2k. \tag{57}$$

The proof of Theorem 26 is given in Appendix E. This result disproves our upper bound Conjecture 1 for $b=2$ stated in our previous study [20]. If the strings of Theorem 26 are based on strings generated by Method A or B, for $b > 2$ they owe their properties to the following distributions of symbols

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

$$
\begin{aligned}
C_{(N-2)} &= [010000111222\ldots10\ldots0],\\
C_{(N-3)} &= [01010000111222\ldots10\ldots0],\\
C_{(N-4)} &= [0101010000111222\ldots10\ldots0],\\
C_{(N-5)} &= [010101000000111222\ldots10\ldots0],\\
C_{(N-6)} &= [010101000000111111222\ldots10\ldots0],\\
C_{(N-7)} &= [0101010000000111111222\ldots10\ldots0],\\
C_{(N-8)} &= [01010100000000110111111222\ldots10\ldots0],\\
C_{(N-9)} &= [0101010010000000110111111222\ldots10\ldots0].
\end{aligned}
\tag{58}
$$

For the strings of the form (58) the fractions in the Shannon entropy are

$$
p_0 = \frac{b+k+f_0}{b^2+b+2k}, \quad p_1 = \frac{b+k+f_1}{b^2+b+2k}, \quad p_{2,\ldots,b-1} = \frac{b+1}{b^2+b+2k}, \tag{59}
$$

where $f_0 = 3$, $f_1 = -1$ if $k = 5$ and $f_0 = 2$, $f_1 = 0$ otherwise, as $[00]$ is inserted into $C_{(N-5)}$, $[11]$ into $C_{(N-6)}$ and $[01]$ or $[10]$ otherwise. This leads to Shannon entropy

$$
H(C_{(N-k)}) = -\frac{b^2-b-2}{b^2+b+2k}\log_2\left(\frac{b+1}{b^2+b+2k}\right) - \frac{b+k+f_1}{b^2+b+2k}\log_2\left(\frac{b+k+f_1}{b^2+b+2k}\right) - \frac{b+k+f_0}{b^2+b+2k}\log_2\left(\frac{b+k+f_0}{b^2+b+2k}\right), \tag{60}
$$

of any $C_{\max}$ string having length $N-k$, for $2 \le k \le 9$. The entropies (54) and (60) are shown in Figure 6. Radix $b = 4$ is the smallest one at which the entropy (60) is a monotonically decreasing function of k. For $b \in \{2,3\}$ there is a local entropy minimum for $k = 5$ and for $b = 2$ an additional local entropy minimum for $k = 2$. Perhaps, the entropy (60) has other local entropy minima for $b < 4$ and for $k > 9$.

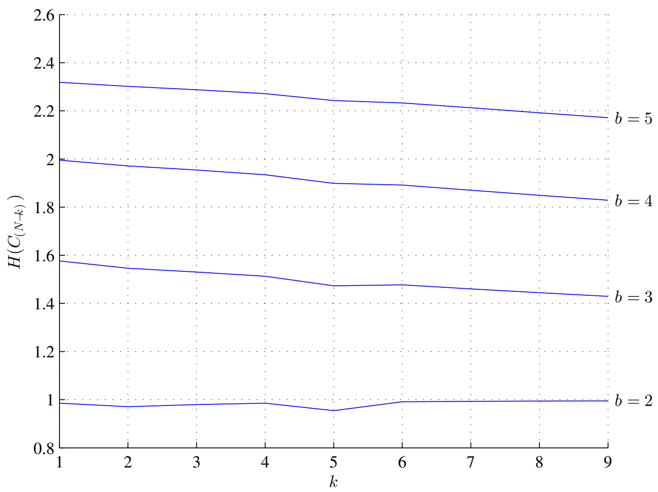

Figure 6. Shannon entropies $H(C_{(N-k)})$ for $1 \le k \le 9$ and $2 \le b \le 5$.

Theorem 27. *If $b > 1$ and $N > N_{(N-9)}$ then*

$$
a_{max}^{(N,b)} \le \begin{cases} a_{max}^{(N-1,b)} + 1 & \text{if} \quad N = 2l, \\ a_{max}^{(N-1,b)} & \text{if} \quad N = 2l+1, \end{cases} \tag{61}
$$

or equivalently

$$
a_{max}^{(N,b)} \le \left\lfloor \frac{N}{2} \right\rfloor + \frac{b(b+1)}{2}, \tag{62}
$$

 doi:10.20944/preprints202409.1581.v9

Proof. Formulas (61) and (62) capture the stepwise linear relation of Theorem 26, shown in Figure 7 for all N. In other words, if $N \geq N_{(N-2)}$, then ASI increases by one, where N increases by two ($b(b+1)/2$ are triangular numbers, OEIS A000217). Once $a_{\max}^{(N,b)}$ is defined by this relation it can only decrease its slope for $N > N_{(N-9)}$. □

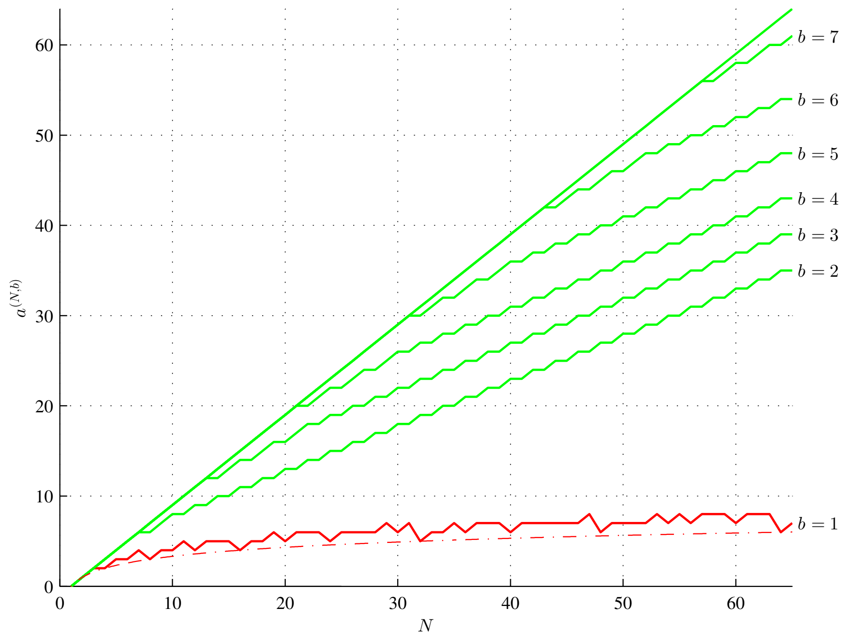

Figure 7. The minimum assembly index (OEIS A003313, red; $\log_2(N)$, red, dash-dot), and the maximum assembly index (green) for $1 \leq b \leq 7$ and $0 < N \leq 65$.

Conjecture 28. *If*

$$N > N_{max} = \begin{cases} 4b^4 & \text{if} \quad b = 2l, \\ 4(b^4+1) & \text{if} \quad b = 2l+1, \end{cases} . \tag{63}$$

then $a_{max}^{(N,b)} < \lfloor N/2 \rfloor + b(b+1)/2$.

W.l.o.g. Conjecture 28 can be proven (or falsified) for $b = 2$. We note that inserting any doublet into a $C_{(N-3)}^{(12,2)}$ string (E9) at any position forms a triplet. Using the equation (26) of Theorem 14 we have

$$\begin{aligned}
a_s &= a_{s-2} + 1, \quad N_s = N_{s-2} + 2, \\
a_s &= N_s - 1 - \sum_{r=1}^{R_r} \left[k_r(n_r - 1) - a\left(C_r^{(n_r,b)}\right) \right], \\
a_{s-2} &= N_{s-2} - 1 - \sum_{p=1}^{R_{s-2}} \left[k_p(n_p - 1) - a\left(C_p^{(n_p,b)}\right) \right], \\
a_s - a_{s-2} &= (N_{s-2} + 2) - 1 - \sum_{r=1}^{R_r} \left[k_r(n_r - 1) - a\left(C_r^{(n_r,b)}\right) \right] - \left(N_{s-2} - 1 - \sum_{p=1}^{R_p} \left[k_p(n_p - 1) - a\left(C_p^{(n_p,b)}\right) \right] \right) = \\
&= 2 - \sum_{r=1}^{R_r} \left[k_r(n_r - 1) - a\left(C_r^{(n_r,b)}\right) \right] + \sum_{p=1}^{R_p} \left[k_p(n_p - 1) - a\left(C_p^{(n_p,b)}\right) \right] = 1, \\
&\sum_{r=1}^{R_r} \left[k_r(n_r - 1) - a\left(C_r^{(n_r,b)}\right) \right] = \sum_{p=1}^{R_p} \left[k_p(n_p - 1) - a\left(C_p^{(n_p,b)}\right) \right] + 1,
\end{aligned} \tag{64}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

for any step s if only $N_{(N-2)} \leq N_s \leq N_{\max}$. Now, assume that $\forall r$, $a\left(C_r^{(n_r,b)}\right) = n_r - 1$ and $\forall p$, $a\left(C_p^{(n_p,b)}\right) = n_p - 1$. Then

$$\begin{aligned}\sum_{r=1}^{R_r}[(k_r-1)(n_r-1)] &= \sum_{p=1}^{R_p}[(k_p-1)(n_p-1)]+1,\\ \sum_{r=1}^{R_r} n_r k_r - \sum_{r=1}^{R_r} n_r - \sum_{r=1}^{R_r} k_r + R_r &= \sum_{p=1}^{R_p} n_p k_p - \sum_{p=1}^{R_p} n_p - \sum_{p=1}^{R_p} k_p + R_p + 1.\end{aligned} \tag{65}$$

The proof of the Conjecture 28 must show the conditions for the equations (64) and (65) to hold. We note that the assumption used in the equation (65) is valid only for $n_r \leq N_{(N-1)}$ and $n_p \leq N_{(N-1)}$. We note that maximum ASI must rise. If it were constant for $N > \hat{N}_{max}$, then at some even larger N it would inevitably become lower than the minimum ASI bound 2 which also rises, and this would be a contradiction. The bounds of Theorems 25 and 26 and Conjecture 27 are illustrated in Figure 7.

Table 2. Distributions of n-plets in strings of maximum ASI.

N	$\times 2_{(b=1)}$	$\times 2_{(b=2)}$	$\times 2_{(b=3)}$	$\times 2_{(b=4)}$	$\times 4_{(b)}$	$\times 8_{(b)}$	$\times 16_{(b)}$	$\times 32_{(b)}$	last $\times 8$	last $\times 4$	last $\times 2$	last $\times 1$	$a_{\max}^{(N,1)}$	$a_{\max}^{(N,2)}$	$a_{\max}^{(N,3)}$	$a_{\max}^{(N,4)}$
1	0	0	0	0	0	0	0	0	N	N	N		0	0	0	0
2	1	1	1	1	0	0	0	0	N	N		N	1	1	1	1
3	1	1	1	1	0	0	0	0	N	N		Y	2	2	2	2
4	1	2	2	2	1	0	0	0	N		N	N	2	3	3	3
5	1	2	2	2	1	0	0	0	N		N	Y	3	4	4	4
6	1	3	3	3	1	0	0	0	N		Y	N	3	5	5	5
7	1	3	3	3	1	0	0	0	N		Y	Y	4	6	6	6
8	1	3	4	4	2	1	0	0		N	N	N	3	6	7	7
9	1	3	4	4	2	1	0	0		N	N	Y	4	7	8	8
10	1	4	5	5	2	1	0	0		N	Y	N	4	8	9	9
11	1	3	5	5	2	1	0	0		N	Y	Y	5	8	10	10
12	1	4	6	6	3	1	0	0		Y	N	N	4	9	11	11
13	1	3	6	6	3	1	0	0		Y	N	Y	5	9	12	12
14	1	4	6	7	3	1	0	0		Y	Y	N	5	10	12	13
15	1	3	6	7	3	1	0	0		Y	Y	Y	6	10	13	14
16	1	4	7	8	4	2	1	0	N	N	N	N	4	11	14	15
17	1	3	6	8	4	2	1	0	N	N	N	Y	5	11	14	16
18	1	4	7	9	4	2	1	0	N	N	Y	N	5	12	15	17
19	1	3	6	9	4	2	1	0	N	N	Y	Y	6	12	15	18
20	1	4	7	10	5	2	1	0	N	Y	N	N	5	13	16	19
21	1	3	6	10	5	2	1	0	N	Y	N	Y	6	13	16	20
22	1	4	7	10	5	2	1	0	N	Y	Y	N	6	14	17	20
23	1	3	6	10	5	2	1	0	N	Y	Y	Y	7	14	17	21
24	1	4	7	11	6	3	1	0	Y	N	N	N	5	15	18	22
25	1	3	6	10	6	3	1	0	Y	N	N	Y	6	15	18	22
26	1	4	7	11	6	3	1	0	Y	N	Y	N	6	16	19	23
27	1	3	6	10	6	3	1	0	Y	N	Y	Y	7	16	19	23
28	1	4	7	11	7	3	1	0	Y	Y	N	N	6	17	20	24
29	1	3	6	10	7	3	1	0	Y	Y	N	Y	7	17	20	24
30	1	4	7	11	7	3	1	0	Y	Y	Y	N	7	18	21	25
31	1	3	6	11	7	3	1	0	Y	Y	Y	Y	8	18	21	25
32	1	4	7	11	8	4	2	1	N	N	N	N	5	19	22	26
33	1	3	6	11	8	4	2	1	N	N	N	Y	6	19	22	26

5. A Method of Generating a Maximum Assembly Index String

The results thus far led us to a simple method of determining the ASI of a maximum ASI string and strengthened our Conjectures 3 and 4 stated in the previous study [20]. The method is based on unique 2^s-plets and powers of two, as shown in Table 2. First, a maximum ASI string is sequenced, every two symbols to find the number n_{UAD} of unique adjoining doublets $\times 2_{(b)}$. In particular, a $C_{(N-1)}$ string (A3) or (B1) contain the maximum of $\lfloor N_{(N-1)}/2 \rfloor$ unique adjoining doublets, a $C_{(N-2)}$ string (E3) contains the maximum of $N_{(N-2)}/2-1$ unique adjoining doublets, and so on. In general, a $C_{(N-k)}$ string contains the maximum of

$$n_{UAD} = \left\lfloor \frac{N_{(N-k)}}{2} \right\rfloor - k + 1 = \begin{cases} b(b+1)/2 = \sum_{l=1}^{b} l & \text{if } k = 1, \\ b(b+1)/2 + 1 = \sum_{l=1}^{b} l + 1 & \text{if } k \neq 1, \end{cases} . \tag{66}$$

unique adjoining doublets, where $N_{(N-k)}$ is given by the relations (52) or (57), which is independent of k.

Subsequently, these doublets form $\times 4_{(b)}$ unique adjoining quadruplets, quadruplets form $\times 8_{(b)}$ unique adjoining octuples, and so on depending on the length of the string N and the radix b, as there can be at most b^{2^s} unique 2^s-plets. The columns "last 2^s" indicate if the assembled string should be terminated with a single substring of length 2^s in descending order. The empty fields in the respective columns for $N > 1$ indicate that a given $\times 2^s$ substring can be interpreted as either a "regular" single $\times 2^s$ substring or a last $\times 2^s$ substring if $\times 2^s = 1$. Furthermore, each $\times 2^s$ step and all last 2^s steps are tantamount to one ASD level.

For example, the $N_{(N-3)}$ string (E10) of length $N_{(N-3)} = 18$ for $b = 3$ can be assembled as

$$\begin{aligned} 0 \circ 1 = [01], 0 \circ 0 = [00], 1 \circ 1 = [11], 1 \circ 2 = [12], & \\ 2 \circ 2 = [22], 1 \circ 0 = [10], 2 \circ 0 = [20] & \quad (\times 2_{(b=3)} = 7), \\ [01] \circ [01] = [0101], [00] \circ [00] = [0000], [11] \circ [12] = [1112], [22] \circ [10] = [2210] & \quad (\times 4 = 4), \\ [0101] \circ [0000] = [01010000], [1112] \circ [2210] = [11122210] & \quad (\times 8 = 2), \\ [01010000] \circ [11122210] = [0101000011122210] & \quad (\times 16 = 1), \\ [0101000011122210] \circ [20] = [010100001112221020] & \quad (\text{last} \times 2), \\ & \quad \overline{} \\ 7+4+2+1+1 = & 15 \text{ steps}, d_{15} = 5. \end{aligned} \tag{67}$$

Similarly, the $N_{(N-1)}$ string (A3) of length $N_{(N-1)} = 21$ for $b = 4$ can be assembled, as shown in Table 2 as

$$\begin{aligned} 0 \circ 0 = [00], 0 \circ 1 = [01], 1 \circ 1 = [11], 2 \circ 2 = [22], 2 \circ 3 = [23], & \\ 3 \circ 3 = [33], 1 \circ 0 = [10], 2 \circ 1 = [21], 3 \circ 2 = [32], 0 \circ 3 = [03] & \quad (\times 2_{(b=4)} = 10), \\ [00] \circ [01] = [0001], [11] \circ [22] = [1122], [23] \circ [33] = [2333], & \\ [10] \circ [21] = [1021], [32] \circ [03] = [3203] & \quad (\times 4 = 5), \\ [0001] \circ [1122] = [00011122], [2333] \circ [1021], [23331021] & \quad (\times 8 = 2), \\ [00011122] \circ [23331021] = [0001112223331021] & \quad (\times 16 = 1), \\ [0001112223331021] \circ [3203] = [00011122233310213203] & \quad (\text{last} \times 4), \\ [00011122233310213203] \circ 0 = [000111222333102132030] & \quad (\text{last} \times 1), \\ & \quad \overline{} \\ 10+5+2+1+1+1 = & 20 \text{ steps}, d_{20} = 5. \end{aligned} \tag{68}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

Furthermore, for $b = 1$ the method produces the DPI (OEIS A014701). For example, the string of length $N = 15$ can be assembled in six steps as

$$\begin{aligned} 0 \circ 0 &= [00], & (\times 2_{(b=1)} &= 1), \\ [00] \circ [00] &= [0000] & (\times 4_{(b=1)} &= 1), \\ [0000] \circ [0000] &= [00000000] & (\times 8_{(b=1)} &= 1), \\ [00000000] \circ [0000] &= [000000000000] & (\text{last} \times 4), \\ [000000000000] \circ [00] &= [00000000000000] & (\text{last} \times 2), \\ [00000000000000] \circ [0] &= [000000000000000] & (\text{last} \times 1), \\ \hline 1+1+1+1+1+1 &= 6 \text{ steps}, d_6 = 4, \end{aligned} \tag{69}$$

where obviously $\max(\times 2^s) = 1$. However, this is the 1st exception for $b = 1$ as the ASI of this string is five if it is assembled using doublet [00] and triplet [000].

We further note that the method illustrated in Table 2 cannot be used to construct the maximum ASI string. For example, both the following two distributions of doublets for $N = 6$ satisfy the distributions of Table 2. However, only the left one correctly reflects the maximum ASI of the assembled string.

$$\begin{aligned} 0 \circ 0 = [00], 0 \circ 1 = [01], 1 \circ 1 = [11] && (\times 2_{(b=2)} = 3), && 0 \circ 0 = [00], 1 \circ 0 = [10], 1 \circ 1 = [11] && (\times 2_{(b=2)} = 3), \\ [00] \circ [01] = [0001] && (\times 4 = 1), && [00] \circ [10] = [0010] && (\times 4 = 1), \\ [0001] \circ [11] = [000111] && (\text{last} \times 2), && [0010] \circ [11] = [001011] && (\text{last} \times 2), \\ 3+1+1 = && 5 \text{ steps}, d_5 = 3, && 3+1+1 = && 5 \neq 4 \text{ steps}, d_5 = 3, \end{aligned} \tag{70}$$

as the right one can be assembled in four steps with $P_4^{(2)} = \{0, 1, 01, \dots\}$. Similarly, only the top distribution of doublets below correctly reflects the maximum ASI of the assembled string for $N = 10$

$$\begin{aligned} 0 \circ 1 = [01], 0 \circ 0 = [00], 1 \circ 1 = [11], 1 \circ 0 = [10] && (\times 2_{(b=2)} = 4), \\ [01] \circ [00] = [0100], [00] \circ [11] = [0011] && (\times 4 = 2), \\ [0100] \circ [0011] = [01000011] && (\times 8 = 1), \\ [01000011] \circ [10] = [0100001110] && (\text{last} \times 2), \\ 4+2+1+1 = && 8 \text{ steps}, d_8 = 4 \\ \\ 0 \circ 0 = [00], 0 \circ 1 = [01], 1 \circ 0 = [10], 1 \circ 1 = [11] && (\times 2_{(b=2)} = 4), \\ [00] \circ [01] = [0001], [10] \circ [11] = [1011] && (\times 4 = 2), \\ [0001] \circ [1011] = [00011011] && (\times 8 = 1), \\ [0001011] \circ [11] = [0001101111] && (\text{last} \times 2), \\ 4+2+1+1 && 8 \neq 6 \text{ steps}, d_8 = 4, \end{aligned} \tag{71}$$

as the bottom one can be assembled in six steps with $P_6^{(2)} = \{0, 1, 11, 011, \dots\}$. Furthermore, this method tends to exaggerate the estimated maximum ASI value, that is,

$$a_{\max}^{(N,b)} \leq a_{\text{method}}^{(N,b)}(C_k), \tag{72}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

where $a^{(N,b)}_{\text{method}}$ is the ASI of a string C_k determined by the method illustrated in Table 2. For example, the first six strings below contain four unique doublets instead of the required three. Therefore

$$\begin{aligned}
C_1 &= [00|10|01|11], \quad a^{(8,2)}(C_1) = 5, \quad a^{(8,2)}_{\text{method}}(C_1) = 7,\\
C_2 &= [00|10|11|01], \quad a^{(8,2)}(C_2) = 5, \quad a^{(8,2)}_{\text{method}}(C_2) = 7,\\
C_3 &= [00|01|10|11], \quad a^{(8,2)}(C_3) = 5, \quad a^{(8,2)}_{\text{method}}(C_3) = 7,\\
C_4 &= [00|01|11|10], \quad a^{(8,2)}_{\max}(C_4) = 6, \quad a^{(8,2)}_{\text{method}}(C_4) = 7,\\
C_5 &= [00|11|10|01], \quad a^{(8,2)}(C_5) = 5, \quad a^{(8,2)}_{\text{method}}(C_5) = 7,\\
C_6 &= [00|11|01|10], \quad a^{(8,2)}(C_6) = 5, \quad a^{(8,2)}_{\text{method}}(C_6) = 7,\\
\\
C_7 &= [00|01|11|00], \quad a^{(8,2)}_{\max}(C_7) = 6 = \quad a^{(8,2)}_{\text{method}}(C_7) = 6.
\end{aligned} \tag{73}$$

Further research should consider researching the formula equivalent to (52) that captures a quadruplet repetition, similarly as $b^2 + b^1 + b^0$ captures a doublet repetition.

6. Discussion

The mathematical findings of this study, especially the theorems concerning the ASI, DPI, and ASD, provide a framework for understanding the principles underlying the assembly of biological macromolecules such as DNA and proteins. These theorems offer insights into how the complexity and functionality of biological sequences are governed by underlying mathematical principles. For instance, here we demonstrate that a DNA strand of length N containing four nucleobases cannot represent a minimum ASI string without violating Chargaff's rules and Theorem 3, which establishes that a minimum ASI string can contain at most two distinct symbols if $N = 2^s$ and at most three, otherwise.

The fundamental interplay of entropy, energy, and temperature is inherent in thermodynamics. Although increasing entropy is a natural tendency of thermodynamic systems following the second law of thermodynamics, dissipative structures, including biological ones, which are open, can decrease their internal entropy by increasing the entropy of their surroundings. By evolving sequences with lower entropies, organisms may achieve more stable and energetically favorable configurations. Despite the mathematical ideal of maximum entropy in balanced strings, biological systems often deviate from this balance. This is evident in natural sequences, where certain nucleotides or amino acids are more prevalent, resulting in lower entropy. For example, the Shannon entropy of the SARS-CoV genome containing N = 29903 nucleobases decreased from $H = 1.3565$ to 1.3562 within two years after the Wuhan outbreak [20,31], ($a^{(29903)}_{\min} = 19$). If the length of a DNA strand is constant, it will tend to evolve to decrease the Shannon entropy [7,31] and, hence, to become less balanced. Here we show that any maximum length string without substring repetitions, that is a maximum ASI string with $a^{(N,b)}_{\max} = N - 1$, is inherently the most balanced: all but one symbol occur $b + 1$ times and one symbol occurs $b + 2$ times within such string $C_{(N-1)}$. However, longer maximum ASI strings $C_{(N-k)}$ become less balanced and, hence, their entropies (60) decrease. Notably, radix $b = 4$ is the smallest one at which the entropy (60) is a monotonically decreasing function of k. Together with Theorem 3 this could be the reason why nature has chosen four nucleobases to encode genetic information. The tendency of biological sequences to become less balanced and thus exhibit lower entropy may reflect an underlying drive toward minimizing the energy required for their assembly and maintenance. This evolution toward energetically favorable states supports the principle that natural systems evolve in ways that reduce free energy, aligning with fundamental thermodynamic and biological principles. More complex sequences require more assembly steps and, consequently, more energy. This energy consideration can influence evolutionary processes, as organisms that synthesize essential proteins and nucleic acids more efficiently may have a selective advantage. Metabolic efficiency is critical for

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

survival, especially in environments where resources are limited, so there is evolutionary pressure to minimize the energy costs associated with macromolecule assembly. Understanding these relationships enhances our comprehension of molecular evolution and the factors that influence the complexity of biological macromolecules.

Analogously, in theoretical physics, black holes - objects of maximal entropy - also consolidate all available energy, suggesting that systems may reach energy minima through configurations that balance entropy and energy. The energy of a black hole conceptualized as a balanced bitstring [28], can be two times the energy of the entropy variation sphere that it generates [29], indicating that a tendency toward imbalance seems to be associated with the minimum energy condition.

In summary, our theorems provide a mathematical underpinning biological phenomena such as the preference for radix $b = 4$ in genetic encoding and the evolutionary trend toward lower entropy. Integrating AT into biological contexts opens avenues for a fundamental mathematical understanding of evolutionary processes, responding to the call for a precise and abstract mathematical theory of evolution [32].

Author Contributions: W.B.: first concept of a general method for constructing the string of length $N_{(N-1)}$ leading to Theorem 25; the concept of the doublet matrix (51); outline of the general Method A; proposition of Theorem 13; a string with exactly two copies of all doublets (C1) idea and the formula (C2) for its length; finding more balanced minimum ASI bitstrings of lengths $N \in \{27, 45, 59, 63\}$; numerous clarity corrections and improvements; P.M.: outline of the general Method B; the hint for ASI combinatorics; creation of a string $C_{\max}^{(24,2)}$; observation of the relation between Theorems 6 and 7; crucial observations leading to the proofs of Theorems 18 and 24; novel Strings (40); the concept of a Table 1; Conjecture 22; numerous clarity corrections and improvements; A.T.: formal proof of Theorem 3; proof that the Shannon entropy (54) can be approximated by $\log_2(b)$ for large b; proof of the Theorem 15; conceptualization of the proof of the Theorem 7 and equation (8); outline and the crucial input to the discussion section 6; numerous clarity corrections and improvements; S.Ł.: The remaining part of the study.

Funding: This research received no external funding.

Data Availability Statement: The public repository for the code written in the MATLAB computational environment and C++ is given under the link https://github.com/szluk/Evolution_of_Information (accessed on 19 September 2024).

Acknowledgments: The authors thank Mariola Bala for her motivation and Rafał Winiarski for noting that the relation (26) is inequality. SŁ thanks his wife, Magdalena Bartocha, for her everlasting support, and his partner and friend, Renata Sobajda, for her prayers.

Conflicts of Interest: Authors Wawrzyniec Bieniawski and Piotr Masierak were employed by the company Łukaszyk Patent Attorneys. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

Appendix A. Method A for Generating $C_{(N-1)}$ String

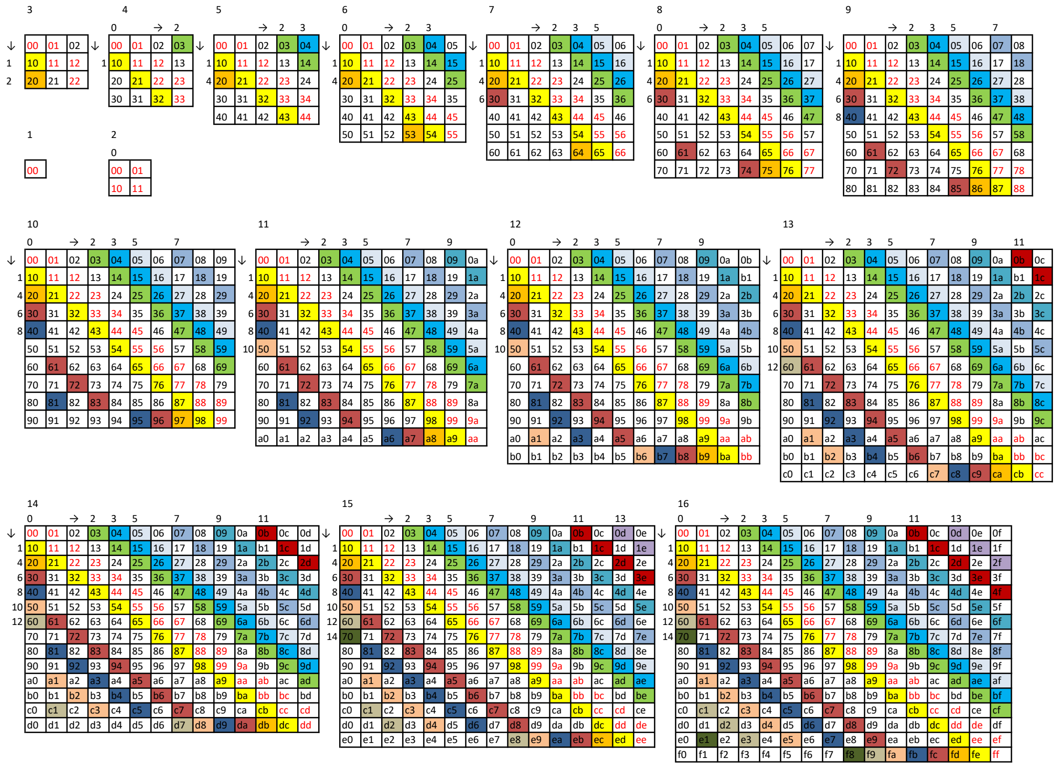

Figure A1. Doublet matrices for $1 \leq b \leq 16$ that illustrate the generation of $N_{(N-1)}$ strings according to Method A. Colored doublets are appended to the initial string of clear triplets in the order indicated by arrows starting from the 1st column or row. Finally, 0 is appended at the end, if b is even.

We start with a string of clear triplets (50). In the 1st step, we form a string containing doublets on the first subdiagonal of the matrix (51) starting with 10

$$[102132\ldots(b-2)(b-3)(b-1)(b-2)], \tag{A1}$$

and we append it to the string (50). With this step, we also eliminate the doublets on the second superdiagonal starting with the doublet 02, as well as the doublet $(b-1)1$. In the 2nd step, we form a string containing doublets on the third superdiagonal beginning with the doublet 03

$$[0314\ldots(b-5)(b-2)(b-4)(b-1)], \tag{A2}$$

and append it to the string formed so far. With this step, we also remove the doublet $(b-2)0$ and the middle part of the second subdiagonal containing $\{31, 42, \ldots, (b-2)(b-4)\}$. And so on. Finally, we append 0 if b is even. This process is illustrated in Figure A1 and for $3 \leq b \leq 13$ generates the following $C_{(N-1)}$ strings

$$\begin{aligned}
&[000111222|10|20],\\
&[000111222333|102132|03|0],\\
&[000111222333444|10213243|0314|20|40],\\
&[000111222333444555|1021324354|031425|0415|2053|0],\\
&[000111222333444555666|102132435465|03142536|041526|2064|0516|30],\\
&[000111222333444555666777|10213243546576|0314253647|04152637|2075|051627|306174|0],\\
&[\ldots|1021324354657687|031425364758|0415263748|2086|05162738|30617285|0718|40],\\
&[\ldots|102132435465768798|03142536475869|041526374859|2097|0516273849|\\
&3061728396|071829|408195|0],\\
&[\ldots|102132435465768798a9|031425364758697a|0415263748596a|20a8|\\
&05162738495a|3061728394a7|0718293a|408192a6|091a|50],\\
&[\ldots|102132435465768798a9ba|031425364758697a8b|0415263748596a7b|20b9|\\
&05162738495a6b|3061728394a5b8|0718293a4b|408192a3b7|091a2b|50a1b6|0],\\
&[\ldots|102132435465768798a9bacb|031425364758697a8b9c|0415263748596a7b8c|20ca|\\
&05162738495a6b7c|3061728394a5b6c9|0718293a4b5c|408192a3b4c8|091a2b3c|50a1b2c7|0b1c|60].
\end{aligned} \tag{A3}$$

Appendix B. Method B for Generating $C_{(N-1)}$ String

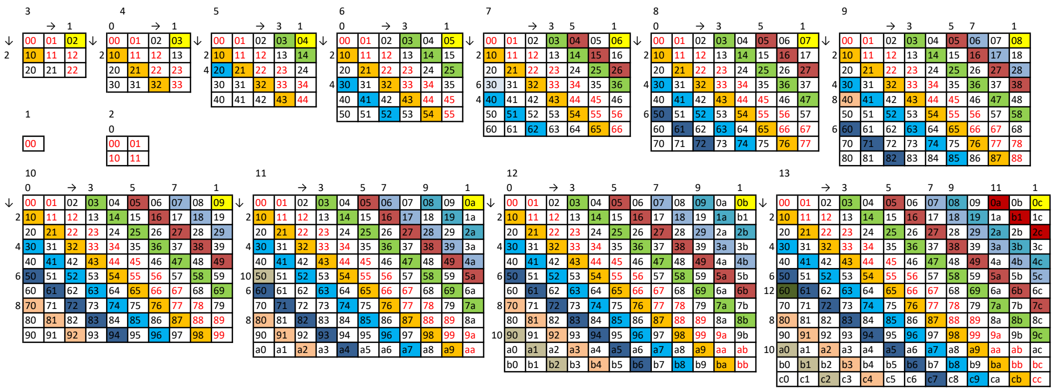

Figure A2. Doublet matrices for $1 \leq b \leq 13$ that illustrate the generation of $N_{(N-1)}$ strings according to Method B. Colored doublets are appended to the initial string of clear triplets in the order indicated by arrows starting from the 1^{st} column or row. Finally, 0 is appended at the end, if b is even.

This method is similar to the Method A. We also start with a string of clear triplets (50) and the matrix of doublets (51) with a crossed diagonal and the first superdiagonal. In the first step, we append the doublet $0(b-1)$ (top right doublet of the matrix of doublets (51)) at the end of the string (50). Next, we generally perform the following pairs of iterations:

1. we check subsequent subdiagonals until we find one that does not contain a doublet present in the string formed so far, we append it at the end of this string and proceed to step 2;
2. we check subsequent superdiagonals until we find one that does not contain a doublet present in the string formed so far, we append it at the end of this string and proceed to step 1.

Finally, we append 0 if b is even. The method is illustrated in Figure A2 and for $3 \leq b \leq 13$ generates the $C_{(N-1)}$ strings in the form

$$
\begin{aligned}
&[000111222|0210],\\
&[000111222333|03|102132|0],\\
&[000111222333444|04|10213243|0314|20],\\
&[000111222333444555|05|1021324354|031425|304152|0],\\
&[000111222333444555666|06|102132435465|03142536|405162|041526|30],\\
&[000111222333444555666777|07|10213243546576|0314253647|3041526374|051627|506172|0],\\
&[\ldots|08|1021324354657687|031425364758|304152637485|05162738|607182|061728|40],\\
&[\ldots|09|102132435465768798|03142536475869|30415263748596|0516273849|5061728394|071829|708192|0],\\
&[\ldots|0a|102132435465768798a9|031425364758697a|30415263748596a7|05162738495a|\\
&60718293a4|061728394a|8091a2|08192a|50],\\
&[\ldots|0b|102132435465768798a9ba|031425364758697a8b|30415263748596a7b8|05162738495a6b|\\
&5061728394a5b6|0718293a4b|708192a3b4|091a2b|90a1b2|0],\\
&[\ldots|0c|102132435465768798a9bacb|031425364758697a8b9c|30415263748596a7b8c9|05162738495a6b7c|\\
&5061728394a5b6c7|0718293a4b5c|8091a2b3c4|08192a3b4c|a0b1c2|0a1b2c|60].
\end{aligned}
\tag{B1}
$$

Appendix C. A String with Exactly Two Copies of All Doublets and No Repeated Triplets

A string that has exactly two copies of all doublets and no repeated triplets can have a form (for $b \in \{1,2,3,4,5\}$)

$$
\begin{aligned}
&[0000]\\
&[00001111|010]\\
&[000011112222|1021|202010]\\
&[0000111122223333|102132|101202303203130]\\
&[00001111222233334444|10213243|1012023034041304242143203140]
\end{aligned}
\tag{C1}
$$

and has a length of

$$
N_{2D} = 2b^2 + b + 1. \tag{C2}
$$

A suboptimal method for its generating (with repeated triplets) is illustrated in Figure A3.

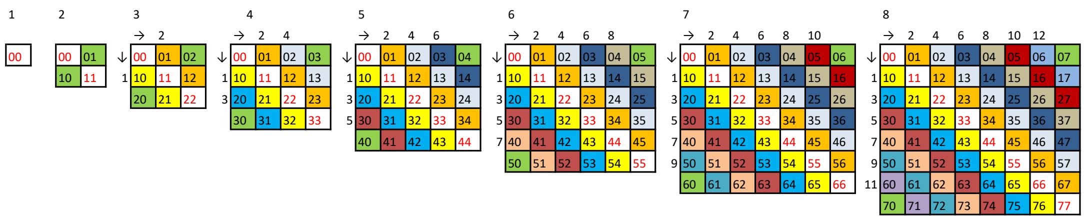

Figure A3. Doublet matrices for $1 \leq b \leq 8$ that illustrate the generation of N_{2D} strings containing exactly two copies of all doublets. Colored doublets are appended to the initial string of clear quadruplets in the order indicated by arrows starting from the 1st column or row. Finally, $0(b-1)0$ is appended at the end. The 1st superdiagonal is appended as 01234 . . . .

Appendix D. Proof of $C_{(N-1)}$ String Theorem

The $N_{(N-1)}$ given by the formula (52) is an odd number for all b. The first element $3b$ is the length of the initial string (50) containing b clear triplets and $b^2 - b - (b-1)$ is the number of doublets available in the matrix (51) after crossing out b doublets on its diagonal and $b-1$ doublets on its

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

superdiagonal that are present in the starting string (50). By definition, a $C_{(N-1)}$ string cannot have any repetitions. To be the longest, it must contain all doublets in the matrix (51) and all clear triplets. Furthermore, to be the most patternless, this string must maximize Shannon entropy; must be the most balanced. For the string of the form (53) the fractions in the Shannon entropy are

$$p_0 = \frac{N_c+1}{N_{(N-1)}}, \quad p_{1,2,\ldots,b-1} = \frac{N_c}{N_{(N-1)}}, \tag{D1}$$

where w.l.o.g. we assume that the symbol occurring $N_c(b)+1$ times within the string is $c=0$. To see that the Shannon entropy (54) of a $C_{(N-1)}$ string can be approximated by $\log_2(b)$ for large b, first notice that $1-b^2<0$ and $b^2+b+1>0, \forall b>1$. Furthermore, $\forall b>0$, $b+1 \ll b^2+b+1$, which implies that the first term

$$\log_2\left(\frac{b+1}{b^2+b+1}\right) < 0. \tag{D2}$$

Similarly the second term,

$$\log_2\left(\frac{b+2}{b^2+b+1}\right) < 0. \tag{D3}$$

Hence, the entropy (54) can be approximated by the dominant contribution from the first term, which is $\log_2(b)$.

The strings given by the relation (52) are not the shortest possible ones. Strings satisfying the equation (53) and satisfying $\min(bN_c(b)+1) > N_{(N-1)}(b-1)$ are given by b^2+1 (OEIS A002522). They can be constructed to contain all possible doublets but without any triplets, starting with an initial balanced string of length $2b$ containing b clear doublets ordered from the main diagonal of the doublet matrix (51). Furthermore, their entropies are smaller than the entropies of the strings given by the equation (52). Namely $\forall b>1$

$$\frac{1-b^2}{b^2+b+1}\log_2\left(\frac{b+1}{b^2+b+1}\right) - \frac{b+2}{b^2+b+1}\log_2\left(\frac{b+2}{b^2+b+1}\right) > \frac{b(1-b)}{b^2+1}\log_2\left(\frac{b}{b^2+1}\right) - \frac{b+1}{b^2+1}\log_2\left(\frac{b+1}{b^2+1}\right). \tag{D4}$$

Now, assume *a contrario* that a string $C'_{(N-1)}$ longer than $N_{(N-1)}$ can be constructed, say of length $N'_{(N-1)} = N_{(N-1)}+1$. But in this case, the corresponding $H(C'_{(N-1)}) < H(C_{(N-1)})$. The string of the length given by the formula (52) maximizes the Shannon entropy if it must additionally satisfy the relation (53). Thus, Theorem 25 is proven.

Appendix E. Proof of $C_{(N-k)}$ String Theorem

We start by noting that for $b=1$, $N_{(N-2)}(1)=5$, as the ASI of [00000] is the same as the ASI of [000000], $N_{(N-3)}(1)=7$, as the ASI of strings of seven and eight same symbols is three, there is no $N_{(N-4)}(1)$, and so on. Hence, Theorem 26 does not hold for $b=1$.

A $C_{(N-1)}$ string contains all doublets. Hence, inserting any basic symbol into any position inevitably leads to a repetition of a doublet. W.l.o.g. we append it at the start of the $C_{(N-1)}$ string, obtaining a string

$$C_k = [*000111222\ldots], \quad a_{\max}^{(N_{(N-1)}+1,b)}(C_k) = N-2. \tag{E1}$$

Another symbol can be introduced to this string without an additional doublet repetition provided that it adjoins the previously introduced symbol, which gives a string

$$C_l = [\star * 000111222\ldots], \quad a_{\max}^{(N_{(N-1)}+2,b)}(C_l) = N-2, \tag{E2}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 January 2025 doi:10.20944/preprints202409.1581.v9

leading to the repetition of the doublet $\star\ast$ or $\ast 0$ but not both of them (here we allow $\star = \ast$). Hence, both the length and the ASI of this string increase by one. Finally, 0 can be appended at the start of this string without an additional doublet repetition provided that $\star \neq 0$ and $\ast = 0$ and the string becomes

$$C_{(N-2)} = [0 \star 0000111222\ldots], \quad a_{\max}^{(N_{(N-1)}+3,b)}(C_{(N-2)}) = N-2, \tag{E3}$$

leading to the mutually exclusive repetition of the doublet $0\star$, $\star 0$ or 00, so that also both length and the ASI of this string increase by one. An insertion of another symbol into the string (E3) at any position will maintain or even decrease the ASI of this newly formed string. For example, appending 0 at the start of the $C_{(N-2)}$ string (E3), where $\star = 1$

$$[0010000111222\ldots]. \tag{E4}$$

forms a 001 triplet based on 00 doublet leading to a decrease of the ASI of this longer string to $a = N-4$ as compared to $a = N-2$ of the string (E3).

$C_{(N-2)}$ string (E3) must contain only two copies of a doublet. Hence, a clear quadruplet ($bbbb$) and a pattern binding different symbols adjoining this quadruplet, such as $[\ldots abbbbc \ldots abc \ldots]$, $[\ldots abbbbaba \ldots]$, etc. must be present, so that any $C_{(N-2)}$ string contains only one pair of repeated doublets ab, bb, or $\{bc, ba\}$ (See also Appendix C). For example, for $N = 10$, sixteen bitstrings

$$\begin{matrix} [0100011110], & [0111100010], & [0111101000], & [\underline{0100001110}], \\ [0001011110], & [0001111010], & [0101111000], & [0111000010] \end{matrix} \tag{E5}$$

(an additional eight are given by swapping 0 with 1) have the ASI $a = N-2 = 8$, where the underlined string (E5) is the one that we formed for $b = 2$. Each string $C_{(N-2)}$ (E5) contains three pairs of doublets [01], [10], and $[\ast\ast]$ overlapped in such a way that only one pair can be reused from the Ω to decrease the maximum $N-1$ ASI by one.

Searching for a $C_{(N-3)}$ string, w.l.o.g. we append $\ast \neq 0$ at the start of the $C_{(N-2)}$ string (E3)

$$C_k = [\ast 010000111222\ldots], \quad a_{\max}^{(N_{(N-1)}+4,b)}(C_k) = N-3. \tag{E6}$$

If $\ast = 1$, we have the same three doublets 10. Otherwise, we have two pairs of the same doublets $\ast 0$ and 10. Both cases are equivalent by Theorem 8. An insertion of another symbol to this string may maintain or even decrease the ASI of this newly formed string. To maximize its ASI, another symbol must adjoin $\ast$. Hence, we append $\star$ at the start, where $\forall\star$ and $\forall\ast \neq 0$, a string

$$C_l = [\star \ast 010000111222\ldots], \quad a_{\max}^{(N_{(N-1)}+5,b)}(C_l) = N-3, \tag{E7}$$

has an increased length and ASI. W.l.o.g. for $b = 2$ we have four bitstrings (E7), wherein three of them

$$\begin{aligned} C_1^{(12,2)} &= [000100001110], & a\left(C_1^{(12,2)}\right) &= 12-4 = 8, \\ C_2^{(12,2)} &= [110100001110], & a\left(C_2^{(12,2)}\right) &= 8, \\ C_3^{(12,2)} &= [100100001110], & a\left(C_3^{(12,2)}\right) &= 8, \end{aligned} \tag{E8}$$

have the same non-maximum ASI and only one have the maximum ASI

$$C_{(N-3)}^{(12,2)} = [010100001110], \quad a_{\max}^{(N_{(N-1)}+5,2)}(C_{(N-3)}^{(12,2)}) = 12-3 = 9, \tag{E9}$$

and cannot be further extended along with the increment of the ASI. Therefore

$$C^{(N,b)}_{(N-3)} = [01010000111222\ldots10\ldots], \quad a_{\max}^{(N_{(N-1)}+5,b)}(C^{(N,b)}_{(N-3)}) = N-3, \tag{E10}$$

and the ASI of this newly formed string increases again. However, the insertion of another symbol into this string will maintain or even decrease the ASI of this newly formed string. Any $C_{(N-3)}$ string must contain only three copies of a doublet, two copies of a triplet, or two pairs of different doublets. W.l.o.g. we have found the following $C_{(N-k)}$ strings for $b=2$ and $4 \le k \le 8$

$$\begin{aligned}
C^{(10,2)}_{(N-2)} &= [0100001110], & a_{\max}^{(10,2)} &= 8, \\
C^{(12,2)}_{(N-3)} &= [010100001110], & a_{\max}^{(12,2)} &= 9 \left([01] \text{ to } C_{\max}^{(10,2)}\right), \\
C^{(14,2)}_{(N-4)} &= [01010100001110], & a_{\max}^{(14,2)} &= 10 \left([01] \text{ to } C_{\max}^{(12,2)}\right), \\
C^{(16,2)}_{(N-5)} &= [0101010000001110], & a_{\max}^{(16,2)} &= 11 \left([00] \text{ to } C_{\max}^{(14,2)}\right), \\
C^{(18,2)}_{(N-6)} &= [010101000000111110], & a_{\max}^{(18,2)} &= 12 \left([11] \text{ to } C_{\max}^{(16,2)}\right), \\
C^{(20,2)}_{(N-7)} &= [01010100000001111110], & a_{\max}^{(20,2)} &= 13 \left([01] \text{ to } C_{\max}^{(18,2)}\right), \\
C^{(22,2)}_{(N-8)} &= [0101010000000110111110], & a_{\max}^{(22,2)} &= 14 \left([10] \text{ to } C_{\max}^{(20,2)}\right), \\
C^{(24,2)}_{(N-9)} &= [010101001000000110111110], & a_{\max}^{(24,2)} &= 15 \left([01] \text{ to } C_{\max}^{(22,2)}\right),
\end{aligned} \tag{E11}$$

which led us to the strings (58) for all $b > 1$. Thus, Theorem 26 is proven.

Appendix F. Assembly Spaces of Minimum Assembly Index Strings

Table A1. Pathways leading to strings having the minimum assembly index (maximizing the number of independent assembly steps - MIA, maximizing the bitstring Shannon entropy - MBL). for $2 \le N \le 65$ (see Section 3 for details).

N	$d^{(N)}_{\min} = \lceil \log_2(N) \rceil$	$d^{(N)}_{a_{\min}}$	$a^{(N)}_{\min}$	$\hat{a}^{(N)}_{\min}$	MIA pathway	MBL pathway (Hamming weight N_1)	String
2	1	1	1	1	$\{2\}$ (1)		$\hat{N}_1$
3	2	2	2	2	$\{2,3\}$ (1)		$\hat{N}_1$
4	2	2	2	2	$\{2,4\}$ (2)		$\hat{N}_1$
5	3	3	3	3	$\{2,4,5\}$ (2)		$\hat{N}_1$
6	3	3	3	3	$\{2,4,6\}$ (3)		$\hat{N}_1$
7	3	3	4	4	$\{2,(3,4),7\}$	$\{2,4,6,7\}$ (3)	$\tilde{N}_3$
8	3	3	3	3	$\{2,4,8\}$ (4)		$\hat{N}_1$
9	4	4	4	4	$\{2,4,8,9\}$ (4)		$\hat{N}_1$
10	4	4	4	4	$\{2,4,8,10\}$ (5)		$\hat{N}_1$
11	4	4	5	5	$\{2,(3,4),7,11\}$	$\{2,4,8,10,11\}$ (5)	$\tilde{N}_3$
12	4	4	4	4	$\{2,3,6,12\}$	$\{2,4,8,12\}$ (6)	$\hat{N}_1$
13	4	4	5	5	$\{2,4,(5,8),13\}$	$\{2,4,8,12,13\}$ (6)	$\tilde{N}_5$
14	4	4	5	5	$\{2,(3,4),7,14\}$	$\{2,4,8,12,14\}$ (7)	$\tilde{N}_3$
15	4	5	5	6	$\{2,3,5,10,15\}$ (**6**)		$\tilde{N}_7$
16	4	4	4	4	$\{2,4,8,16\}$ (8)		$\hat{N}_1$
17	5	5	5	5	$\{2,4,8,16,17\}$ (8)		$\hat{N}_1$
18	5	5	5	5	$\{2,4,8,16,18\}$ (9)		$\hat{N}_1$
19	5	5	6	6	$\{2,(3,4),8,11,19\}$	$\{2,4,8,10,18,19\}$ (9)	$\tilde{N}_3$
20	5	5	5	5	$\{2,3,5,10,20\}$	$\{2,4,8,16,20\}$ (10)	$\hat{N}_1$
21	5	5	6	6	$\{2,4,(5,8),16,21\}$	$\{2,4,8,16,20,21\}$ (10)	$\tilde{N}_5$
22	5	5	6	6	$\{2,(3,4),7,11,22\}$	$\{2,4,8,16,20,22\}$ (11)	$\tilde{N}_3$
23	5	6	6	7	$\{2,3,5,10,20,23\}$ (**9**)		$\tilde{N}_{2^n+1,b}$
24	5	5	5	5	$\{2,4,8,12,24\}$ (12)		$\hat{N}_1$
25	5	5	6	6	$\{2,4,8,(9,16),25\}$	$\{2,4,8,16,24,25\}$ (12)	$\tilde{N}_9$
26	5	5	6	6	$\{2,4,(5,8),13,26\}$	$\{2,4,8,16,24,26\}$ (13)	$\tilde{N}_5$
27	5	6	6	7	$\{2,3,6,12,24,27\}$	$\{2,4,5,9,18,27\}$ (**12**)	$\tilde{N}_{2^n+1,a}$
28	5	5	6	6	$\{2,(3,4),7,14,28\}$	$\{2,4,8,16,24,28\}$ (14)	$\tilde{N}_3$
29	5	6	7	7	$\{2,4,8,(9,10),20,29\}$	$\{2,4,8,16,24,28,29\}$ (14)	
30	5	6	6	7	$\{2,3,5,10,15,30\}$	$\{2,4,6,10,20,30\}$ (15)	$\tilde{N}_7$
31	5	6	7	8	$\{2,4,(5,8),13,26,31\}$	$\{2,4,8,10,20,30,31\}$ (15)	$\tilde{N}_{15}$
32	5	5	5	5	$\{2,4,8,16,32\}$ (16)		$\hat{N}_1$

Table A1. *Cont.*

N	$d_{\text{min}}^{(N)} = \lceil \log_2(N) \rceil$	$d_{a_{\text{min}}}^{(N)}$	$a_{\text{min}}^{(N)}$	$\hat{a}_{\text{min}}^{(N)}$	MIA pathway	MBL pathway (Hamming weight N_1)	String
33	6	6	6	6	$\{2,4,8,16,32,33\}$ (16)		$\hat{N}_1$
34	6	6	6	6	$\{2,4,8,16,32,34\}$ (17)		$\hat{N}_1$
35	6	6	7	7	$\{2,(3,4),7,14,28,35\}$	$\{2,4,8,16,32,34,35\}$ (17)	$\tilde{N}_3$
36	6	6	6	6	$\{2,4,8,16,32,36\}$ (18)		$\hat{N}_1$
37	6	6	7	7	$\{2,4,(5,8),16,32,37\}$	$\{2,4,8,16,32,36,37\}$ (18)	$\tilde{N}_5$
38	6	6	7	7	$\{2,(3,4),8,11,19,38\}$	$\{2,4,8,16,32,36,38\}$ (19)	$\tilde{N}_3$
39	6	6	7	8	$\{2,4,(5,8),13,26,39\}$	$\{2,4,(5,8),13,26,39\}$ (**18**)	
40	6	6	6	6	$\{2,4,8,16,32,40\}$	$\{2,4,8,16,32,40\}$ (20)	$\hat{N}_1$
41	6	6	7	7	$\{2,4,8,(9,16),25,41\}$	$\{2,4,8,16,32,40,41\}$ (20)	$\tilde{N}_9$
42	6	6	7	7	$\{2,(3,4),7,14,28,42\}$	$\{2,4,8,16,32,40,42\}$ (21)	$\tilde{N}_5$
43	6	7	7	8	$\{2,3,5,10,20,40,43\}$ (**17**)		$\tilde{N}_{2^n+1,b}$
44	6	6	7	7	$\{2,(3,4),7,11,22,44\}$	$\{2,4,8,16,32,40,44\}$ (22)	$\tilde{N}_3$
45	6	7	7	8	$\{2,3,5,10,20,40,45\}$	$\{2,4,5,9,18,27,45\}$ (**20**)	$\tilde{N}_{2^n+1,b}$
46	6	7	7	8	$\{2,3,5,10,20,23,46\}$	$\{2,4,6,10,20,40,46\}$ (23)	$\tilde{N}_{2^n+1,b}$
47	6	7	8	9	$\{2,(3,4),7,11,22,44,47\}$	$\{2,4,6,10,20,40,46,47\}$ (23)	$\tilde{N}_{15}$
48	6	6	6	6	$\{2,4,8,12,24,48\}$ (24)		$\hat{N}_1$
49	6	7	7	7	$\{2,4,8,12,24,48,49\}$ (24)		$\tilde{N}_{17}$
50	6	6	7	7	$\{2,4,8,(9,16),25,50\}$	$\{2,4,8,16,32,40,48,50\}$ (25)	$\tilde{N}_9$
51	6	7	7	8	$\{2,4,8,16,17,34,51\}$ (**24**)		$\tilde{N}_{2^n+1,a}$
52	6	6	7	7	$\{2,4,(5,8),13,26,52\}$	$\{2,4,8,16,32,40,48,52\}$ (26)	$\tilde{N}_5$
53	6	7	8	8	$\{2,4,(5,8),16,32,48,53\}$	$\{2,4,8,16,32,40,48,52,53\}$ (26)	
54	6	7	7	8	$\{2,3,6,12,24,27,54\}$	$\{2,4,6,12,24,48,54\}$ (27)	$\tilde{N}_{2^n+1,a}$
55	6	7	8	9	$\{2,(3,4),7,11,22,44,55\}$	$\{2,4,8,16,18,36,54,55\}$ (27)	
56	6	6	7	7	$\{2,(3,4),7,14,28,56\}$	$\{2,4,8,16,32,48,56\}$ (28)	$\tilde{N}_3$
57	6	7	8	8	$\{2,(3,4),7,14,28,56,57\}$	$\{2,4,8,16,32,48,56,57\}$ (28)	
58	6	7	8	8	$\{2,(3,4),7,14,28,29,58\}$	$\{2,4,8,16,32,48,56,58\}$ (29)	
59	6	7	8	9	$\{2,(3,4),7,14,28,56,59\}$	$\{2,4,5,9,18,27,54,59\}$ (**26**)	$\tilde{N}_{27}$
60	6	7	7	8	$\{2,4,8,12,24,48,60\}$	$\{2,4,6,10,20,30,60\}$ (30)	$\tilde{N}_7$
61	6	8	8	9	$\{2,4,8,12,24,48,60,61\}$	$\{2,4,8,16,20,40,60,61\}$ (30)	
62	6	7	8	9	$\{2,(3,4),7,14,28,31,62\}$	$\{2,4,8,16,20,40,60,62\}$ (31)	$\tilde{N}_{15}$
63	6	7	8	10	$\{2,(3,4),7,14,21,42,63\}$	$\{2,4,5,9,18,27,45,63\}$ (**28**)	
64	6	6	6	6	$\{2,4,8,16,32,64\}$ (32)		$\hat{N}_1$
65	7	7	7	7	$\{2,4,8,16,32,64,65\}$ (32)		$\hat{N}_1$

References

1. M. Levin, Self-constructing bodies, collective minds: the intersection of cs, cognitive bio, and philosophy (2024).
2. M. Levin, Self-Improvising Memory: A Perspective on Memories as Agential, Dynamically Reinterpreting Cognitive Glue, Entropy **26**, 481 (2024).
3. D. Hoffman, *The case against reality: Why evolution hid the truth from our eyes* (WW Norton & Company, 2019).
4. D. D. Hoffman and M. Singh, Perception, Evolution, and the Explanatory Scope of Scientific Theories, Journal of Consciousness Studies **31**, 29 (2024).
5. Č. Brukner, A No-Go Theorem for Observer-Independent Facts, Entropy **20**, 10.3390/e20050350 (2018).
6. S. Łukaszyk and A. Tomski, Omnidimensional Convex Polytopes, Symmetry **15**, 10.3390/sym15030755 (2023).
7. S. Łukaszyk, Shannon entropy of chemical elements, European Journal of Applied Sciences **11**, 443–458 (2024).
8. S. Łukaszyk, The Imaginary Universe (on the Three Complementary Sets of Measurement Units Defining Three Dark Electrons) (2024).
9. A. D. Lorenzo, A relation between pythagorean triples and the special theory of relativity (2018).
10. H. Sporn, Pythagorean triples using the relativistic velocity addition formula, The Mathematical Gazette **108**, 219 (2024).
11. S. Łukaszyk, Metallic Ratios and Angles of a Real Argument, IPI Letters , 26 (2024).
12. S. M. Marshall, A. R. G. Murray, and L. Cronin, A probabilistic framework for identifying biosignatures using Pathway Complexity, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **375**, 20160342 (2017).
13. S. Imari Walker, L. Cronin, A. Drew, S. Domagal-Goldman, T. Fisher, and M. Line, Probabilistic biosignature frameworks, in *Planetary Astrobiology*, edited by V. Meadows, G. Arney, B. Schmidt, and D. J. Des Marais (University of Arizona Press, 2019) pp. 1–1.
14. V. S. Meadows, G. N. Arney, B. E. Schmidt, and D. J. Des Marais, eds., *Planetary astrobiology*, University of Arizona space science series (The University of Arizona Press ; Houston : Lunar and Planetary Institute, Tucson, 2020) oCLC: 1151198948.

15. Y. Liu, C. Mathis, M. D. Bajczyk, S. M. Marshall, L. Wilbraham, and L. Cronin, Exploring and mapping chemical space with molecular assembly trees, Science Advances **7**, eabj2465 (2021).
16. S. M. Marshall, C. Mathis, E. Carrick, G. Keenan, G. J. T. Cooper, H. Graham, M. Craven, P. S. Gromski, D. G. Moore, S. I. Walker, and L. Cronin, Identifying molecules as biosignatures with assembly theory and mass spectrometry, Nature Communications **12**, 3033 (2021).
17. S. M. Marshall, D. G. Moore, A. R. G. Murray, S. I. Walker, and L. Cronin, Formalising the Pathways to Life Using Assembly Spaces, Entropy **24**, 884 (2022).
18. A. Sharma, D. Czégel, M. Lachmann, C. P. Kempes, S. I. Walker, and L. Cronin, Assembly theory explains and quantifies selection and evolution, Nature **622**, 321 (2023).
19. M. Jirasek, A. Sharma, J. R. Bame, S. H. M. Mehr, N. Bell, S. M. Marshall, C. Mathis, A. MacLeod, G. J. T. Cooper, M. Swart, R. Mollfulleda, and L. Cronin, Investigating and Quantifying Molecular Complexity Using Assembly Theory and Spectroscopy, ACS Central Science **10**, 1054 (2024).
20. S. Łukaszyk and W. Bieniawski, Assembly Theory of Binary Messages, Mathematics **12**, 1600 (2024).
21. S. Raubitzek, A. Schatten, P. König, E. Marica, S. Eresheim, and K. Mallinger, Autocatalytic Sets and Assembly Theory: A Toy Model Perspective, Entropy **26**, 808 (2024).
22. S. Łukaszyk, On the "Assembly Theory and its Relationship with Computational Complexity" (2024).
23. P. Francis, Dilexit nos: Encyclical letter on the human and divine love of the heart of jesus christ (2024), accessed: 2024-11-01.
24. Book of John [1.3] (c90).
25. C. P. Kempes, M. Lachmann, A. Iannaccone, G. M. Fricke, M. R. Chowdhury, S. I. Walker, and L. Cronin, Assembly Theory and its Relationship with Computational Complexity (2024).
26. L. Cronin, Exploring assembly index of strings is a good way to show why assembly & entropy are intrinsically different., https://x.com/leecronin/status/1850289225935257665 (2024), accessed: 2024-11-01.
27. S. Pagel, A. Sharma, and L. Cronin, Mapping Evolution of Molecules Across Biochemistry with Assembly Theory (2024).
28. S. Łukaszyk, Black Hole Horizons as Patternless Binary Messages and Markers of Dimensionality, in *Future Relativity, Gravitation, Cosmology* (Nova Science Publishers, 2023) Chap. 15, pp. 317–374.
29. S. Łukaszyk, Life as the explanation of the measurement problem, Journal of Physics: Conference Series **2701**, 012124 (2024).
30. G. J. Chaitin, Randomness and Mathematical Proof, Scientific American **232**, 47 (1975).
31. M. M. Vopson, The second law of infodynamics and its implications for the simulated universe hypothesis, AIP Advances **13**, 105308 (2023).
32. G. Chaitin, *Proving Darwin: Making Biology Mathematical* (Knopf Doubleday Publishing Group, 2013)