

Novel Lossless Compression Method Based on the Fourier Transform to Approximate the Kolmogorov Complexity of Elementary Cellular Automata

Mohammed Terry-Jack 

Department of Computer Science, University of York, York, UK

Email: mohammedterryjack@gmail.com

How to cite this paper: Terry-Jack, M. (2022) Novel Lossless Compression Method Based on the Fourier Transform to Approximate the Kolmogorov Complexity of Elementary Cellular Automata. *Journal of Software Engineering and Applications*, 15, 359-383.

<https://doi.org/10.4236/jsea.2022.1510021>

Received: October 3, 2022

Accepted: October 23, 2022

Published: October 26, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

We propose a novel, lossless compression algorithm, based on the 2D Discrete Fast Fourier Transform, to approximate the Algorithmic (Kolmogorov) Complexity of Elementary Cellular Automata. Fast Fourier transforms are widely used in image compression but their lossy nature exclude them as viable candidates for Kolmogorov Complexity approximations. For the first time, we present a way to adapt fourier transforms for lossless image compression. The proposed method has a very strong Pearsons correlation to existing complexity metrics and we further establish its consistency as a complexity metric by confirming its measurements never exceed the complexity of nothingness and randomness (representing the lower and upper limits of complexity). Surprisingly, many of the other methods tested fail this simple sanity check. A final symmetry-based test also demonstrates our method's superiority over existing lossless compression metrics. All complexity metrics tested, as well as the code used to generate and augment the original dataset, can be found in our github repository: ECA complexity metrics¹.

Keywords

Fast Fourier Transform, Lossless Compression, Elementary Cellular Automata, Algorithmic Information Theory, Kolmogorov Complexity

1. Introduction

Kolmogorov complexity is a unique formalisation of complexity which quantifies “information” from an algorithmic point of view [1] and has since become

¹https://github.com/mohammedterryjack/ECA_complexity_metrics.

very well-established [1], resolving many open problems in mathematics and computational complexity theory. It is an ideal metric for analysing the dynamical properties of systems known to exhibit chaotic and complex behaviours, like cellular automata, by providing insights into the information content of their evolution [1] [2].

However, kolmogorov complexity is an uncomputable function [1] [3] (or lower semi-computable [4]) and only approximable [1]. Traditionally, the way to approach algorithmic complexity is to use a general lossless compression algorithm [3] [4]. While there exist many generic lossless compression methods which successfully compress large regions of uniformity and repetitive behaviour (e.g. the Lempel Ziv Welch algorithm [3], Huffman coding [5] [6], Run-length encoding [6], Arithmetic encoding [6], etc.), most of these methods fail to sufficiently compress irregular patterns or find redundancies in more exotic symmetries (e.g. self-similarities) as well as other common features of fractals [7] often found in complex classes of CA dynamics. Furthermore, most of these methods are designed to compress one-dimensional data (*i.e.* strings and sequences) [5] [8], which make them less suited to two-dimensional data (*i.e.* images, spacetime evolutions, etc.) which naturally contain nested structures [5] and patterns that are only distinguishable in their natural dimension [8]. All of these compression inaccuracies inevitably lead to poorer kolmogorov complexity approximations.

Fortunately, there also exist compression methods better suited to two-dimensional images and more complex patterns (e.g. Chroma sampling, Transform coding, Fractal compression [6], Fourier transforms, etc.) [9]. The Fourier Transform, in particular, is a very popular method for image compression, analysis, filtering and reconstruction. Unfortunately, Fourier Transform based compression methods and the like are lossy [10], which is not a problem for most image compression applications (e.g. compressing large images to save disk space and optimise databases, remote sensing or facilitating data transmission [6]), but does exclude them as viable candidates for Kolmogorov Complexity approximations.

In this paper, we propose a novel adaptation of the 2D-FFT which produces lossless compressions of black and white images. This, in turn, is used to more accurately measure the Kolmogorov Complexity of ECA spacetime diagrams. To the best of our knowledge, we are the first to propose a lossless compression method based on the Fourier Transform.

2. Fundamental Principles

2.1. 1D Elementary Cellular Automata

An elementary cellular automaton (ECA) is the simplest CA which exists; a discrete dynamical system [2] [11] [12] consisting of a configuration-transition function, g_ρ , and a one-dimensional, W -wide lattice of identical, locally-interconnected, binary-state cells, $\{0,1\}^W$.

$$\text{ECA} = \langle g_\rho, \{0,1\}^W \rangle \quad (1)$$

1D-ECA Rules. The configuration-transition function, g_ρ , dictates how to update the ECA's lattice configuration, $\chi \in \{0,1\}^W$, (*i.e.* a particular combination of cell states at any given time), as a result of simultaneously applying a local transition rule, ρ , to all cells in the lattice.

$$\begin{aligned} g_\rho : \{0,1\}^W &\rightarrow \{0,1\}^W \\ [g_\rho(\chi)]_w &= \rho(\chi_{w-n}, \dots, \chi_w, \dots, \chi_{w+n}) \\ w, n &\in \mathbb{N} : 0 \leq w \leq W-1, n \geq 1 \end{aligned} \quad (2)$$

The local transition rule, ρ , returns a cell's new state given its current state and the states of n neighbouring cells to either side of it. These $2n+1$ consecutive cells are collectively referred to as a neighbourhood (therefore, the lattice width must at least be the size of one neighbourhood, $W \geq 2n+1$).

$$\begin{aligned} \rho : \{0,1\}^{2n+1} &\rightarrow \{0,1\} \\ \chi_w &= \rho(\chi_{w-n}, \dots, \chi_w, \dots, \chi_{w+n}) \end{aligned} \quad (3)$$

The local transition rule, ρ , for the simplest ECA considers only nearest-neighbour interactions (*i.e.* a neighbourhood with a single-cell radius, $n=1$), and so its neighbourhood has three cells; those of the nearest neighbour to either side, χ_{w-1} and χ_{w+1} , and that of the central cell between them, χ_w . It should be noted that an ECA lattice wraps around, such that $\chi_0 = \chi_W$, and so the cells at either "end" of the lattice also have complete neighbourhoods (e.g.

$\chi_{W-1}, \chi_0, \chi_1$ and $\chi_{W-2}, \chi_{W-1}, \chi_0$ respectively).

There are $2^{2^{2n+1}} = 256$ different local transition rules that can be effectively enumerated [13] and since they are commonly formulated as lookup tables (making them very simple to describe [1] [2] [12] [14] [15]) they can end up having all sorts of functional forms (including discontinuous ones [14]). This is more apparent when they are reformulated as boolean functions (which are intrinsically nonlinear anyway), since the resulting expressions can be utterly irregular [14].

1D-ECA Evolutions. A spacetime evolution is a binary matrix $s \in \{0,1\}^{W \times T}$ representing the evolution of 1D lattice configurations over time, produced by iteratively applying the configuration-transition function, g_ρ , from some given initial configuration, χ_0 , up to $T \in \mathbb{Z}^+$ time steps. Such that $t \in \mathbb{N} : 0 \leq t \leq T-1$, and $g_\rho^0(\chi') = \chi'$.

$$s_{:,t} = g_\rho^{ot}(\chi') \quad (4)$$

This evolution can be behaviourally rich and permit a wide range of intriguing spatiotemporal dynamics [1] [2] [3] [11] [16] which has been the focus of many studies. In fact, a spacetime evolution, s , is often compared to other spatially extended dynamical systems, such as classical, discretised, partial differential equations [1] [14].

2.2. Measuring the Complexity of 1D-ECAs

Qualitative Complexity. Since the evolution of a one-dimensional CA can be

visualised as a two-dimensional image [2] [16], Stephen Wolfram looked at thousands of space-time diagrams [4] [5] [16] [17] and assigned each CA rule to a class of typical behaviour based purely on the general visual appearance of the patterns he saw [4] [5] [16]. After extensive experimentations with computer simulations, Wolfram proposed what is now the most famous classification of CAs, consisting of four classes that are primarily qualitative [3] [18] and ultimately subjective [17] (with a considerable degree of freedom in deciding what each class description means [17]) as opposed to being determined according to objective definitions. Wolfram supported his heuristic classification approach which relied on “how things look to our eyes” [5] by appealing to the superiority of our innate pattern recognition ability (“despite all the various methods of mathematical and other analysis that have been developed, our visual system still represents one of the most powerful and reliable tools we have” [5]).

However, his classification had many problems which primarily stem from its empirical underpinning [3] [12] [16] [17] [18] [19]. Without a formal mathematical definition for each class [1] [17] [20], or even some statistical quantities which sharply elaborate the characteristics of each class [16] [20], they lacked computable qualities which prevent them from being effective (*i.e.* determined algorithmically) [1] [17] or usable for precise answers and predictions in the CA universe [19] throwing the practicality and usefulness of his classification into question [1]. It also meant that there are no quick tests to distinguish classes from one another nor any mathematical evidence to determine whether a given cellular automata belonged to one class or another [12]. Certain distinctions motivated by the visual detection of complex, localised structures may have also been superficial or artificial when viewed from a deeper, dynamical systems viewpoint [2]. Equally, far more subtle class distinctions (which are not easily detected by the naked eye) may have been overlooked (e.g. some rules in Wolfram’s chaotic class III merely preserve the randomness already present in the initial configuration, whereas rules 30 and 45 are uniquely capable of generating randomness from an ordered initial configuration [19]). While his classification was a breakthrough and inspiration for future research in the field, it was insufficient [19] [21] and incomplete [1] [16]. Subsequent attempts relied far less on qualitative criteria [18], leveraging instead more numerical approaches to quantify distinguishing characteristics of CAs, such as their complexity.

Quantifying Complexity. Complexity, however, is easier to identify than it is to quantify [18] (even the meaning in everyday parlance has a lot of ambiguity; with meanings ranging from “interwoven”, “hard to describe” [22], “randomness” [5] [23], etc). Complexity is, more precisely, a measure of “information” [24] (*i.e.* the amount of “novelty”, “surprise” or “unpredictability” something has) and so many information-theoretic tools have been used to quantify the complexity of CAs [1] [18], such as algorithmic (kolmogorov) complexity, index complexity, communication complexity, integrated information (IIT) [18] [25], causal complexity (a CA is treated as a directed causal graph and perturbed into

all possible states to note every possible state-to-state transitions over one time-step. The intrinsic cause-effect structure of the CA can be evaluated and taken as the foundation for the rule's potential for dynamical complexity, which may or may not manifest itself under the observed circumstances [25]), morphological complexity [25] (*i.e.* block shannon entropy [4] [22] which generalises the concept of shannon entropy to the number of distinct $N \times N$ block patterns in an ECA evolution for a particular initial configuration [4] [25]), metric entropy (the asymptotic mean entropy of local patterns [26]), transfer entropy [18] (a measure of complexity via the averaged, directed exchange of information processed between two parts of an ECA; a source cell and target cell, to determine the reduction in uncertainty of the future state of the source cell given knowledge of the target cell. *i.e.* rules that duplicate the input in a trivial way or annihilate to produce all 0s or 1s yield a transfer entropy of 0 [18]), etc.

A few approaches measure the complexity of CAs through the underlying, local rule [3]; by computing the rule's lookup frequency [22] (*i.e.* the number of times each input neighbourhood in the table is looked up [22]) or its compression ratio [1], or minimised boolean function [22] (*i.e.* representing the rule table as a boolean function and applying minimisation techniques to it from boolean algebra to remove irrelevant input variables from the truth table [22]), or the complexity of the geometrical structure of its boolean cube representation (*i.e.* a rule table can be represented as a boolean cube by identifying each of its eight unique input neighbourhoods to a vertex, and assigning colours to the eight vertices to represent their corresponding outputs [27]). The complexity is thus the minimum number of parallel planes necessary to separate the boolean cube's coloured vertices [22] [27]).

Unfortunately, purely rule-based complexity measures cannot possibly provide sufficient conditions to predict the complexity of CAs because their dynamical evolutions depend on their initial configuration² and the CAs size, in addition to its local rule [25] (small lattices with only a few cells are unable to unfold the full dynamical potential of their rule, even though the local interactions are the same as in larger systems [25]). Since the CAs behaviour is decided locally (by the transition rule) it is tempting to assume the resulting global dynamics are indifferent to non-local CA parameters that do not inform the decision of the local transition rule (such as the CA's lattice structure or its dimensionality). However, it would be premature to conclude that the expressed CA dynamics are solely determined by the CA transition rule (even though the global CA classification task is often equated to the local "rule characterisation task" [14] [28]) because more and more examples are being found that demonstrate

²We can better understand why CA dynamics are sensitive to initial configurations if we take a dynamical systems perspective [13]. A CAs spacetime evolution represents a particular trajectory through state-space [15] (unique to each rule). A particular seed value (an initial configuration) might start within an attractor basin and result in a stable, periodic orbit (a periodic spacetime pattern), while another seed falls within the basin of a strange attractor and follows a chaotic orbit. State-spaces containing many attractors would produce rules that are more sensitively dependent upon initial conditions (initial configurations).

CA dynamics are sensitively dependent on the initial configuration [28] (a non-local property). For example, some rules exhibit a large increase in complexity when starting from more complex initial configurations [13] (e.g. ECA rule 40 is ordered when starting from a configuration of all 0 s but chaotic when starting from random initial configurations [3]. Likewise, ECA rule 106 generates simple dynamics when starting from an initial configuration of all 0 s containing a single 1, but becomes dynamically complex when the initial configuration contains a second, adjacent 1 [28]). In fact, a major shortcoming of choosing the single-cell input is that many of Wolfram's class IV rules lead to simpler dynamics than what they are truly capable of expressing [18]. Other rules seem less sensitive to initial configurations [11] and changes to their initial configurations have no noticeable impact on their resulting dynamics [13] (e.g. rules 30, 45 and 110). However, even rule 30, which is well known for its chaotic dynamics, can suddenly exhibit simple, periodic behaviour for certain initial configurations (see **Figure 1**) [5].

Since the output of a CA rule can be highly dependent on its initial configuration [18] (such that the same rule may behave differently and show patterns of very different complexity [25] when applied to a different initial configuration [4] [11], it can also belong to multiple classes of behaviour depending on the initial configuration [3]. However, it is important to note that despite a CA rule being able to take on multiple class labels, rule-based complexity measures attempt to label the rule as a single class. Only by enumerating or perturbing a rule for arbitrary initial conditions can its full potential complexity be revealed [25] and any measure which ignores the delicate interplay of the initial configuration on the expressed dynamics will inevitably fail to capture the nuances of a rule's full dynamical range. Incomplete classification methods have hastily labelled rules as uninteresting [29], despite their potential to express dynamics of interest, and only later are researchers stumbling upon important discoveries in previously overlooked rules (e.g. rules 20 and 52) [29] [30], even finding complex emergent phenomena in rules previously dismissed as ordered (Wolfram's class I) or periodic (Wolfram's class II).

Most approaches, therefore, attempt to quantify the complexity expressed in the transient patterns of a CA's spacetime evolution [3] [8] [25] (*i.e.* the global

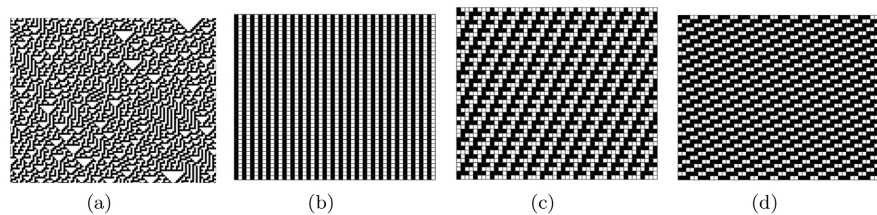


Figure 1. A chaotic spacetime pattern typically produced by ECA Rule 30 is shown (a) next to three particular initial configurations where that same rule produces non-chaotic, perfectly periodic patterns (b, c, d) [5]. (a) ECA Rule 30 (chaotic); (b) ECA Rule 30 (periodic); (c) ECA Rule 30 (periodic); (d) ECA Rule 30 (periodic).

result of the local rule being unfolded from some initial configuration), ignoring any specifics about the underlying rule structure itself [22].

2.3. Kolmogorov Complexity

Kolmogorov complexity [1] [24] [31] (also known as kolmogorov-chaitin complexity [4], algorithmic complexity [1] [4], program-size complexity [3] [4], etc.) is a unique formalisation of complexity derived from algorithmic information theory (AIT) [1] [4] which quantifies “information” from an algorithmic point of view [1] and has since become very well-established [1], resolving many open problems in mathematics and computational complexity theory. The Kolmogorov Complexity, K_U , of a sequence, x , (*i.e.* a string of characters, a sentence, etc.) is the length, l , in bits of the shortest computer program, p , running on a Universal Turing machine, U , written in any programming language, to output the sequence, x , exactly (*i.e.* losslessly).

$$K_U(x) = \min(\{l(p) \mid U(p) = x\}) \quad (5)$$

The program, p , is a description of the sequence, x , [3] [24] [31]. Simpler sequences can be fully described with shorter programs [1] [31], resulting in a smaller kolmogorov complexity [31]. Complex sequences, however, require longer programs to be described fully, and so their kolmogorov complexity is higher [31]. For example, a regular sequence (e.g. $\dots 01010101\dots$) with a lot of redundant information [5] can be described more compactly by a program (e.g. $(01)^n$) [3]. On the other hand, a sequence with no recognisable regularities could require the sequence to be spelt out entirely [1] [5].

Complexity vs Entropy. Shannon’s Information Entropy (also known as Shannon entropy [31]) is an alternative way to quantify information [23]. It is a measure of complexity [32] which connects the amount of information about an event to the probability of the occurrence of the event (*i.e.* more likely events provide less novelty or new information, while highly uncertain and unpredictable events are far more informative and thus provide a lot of new information). High entropy implies high uncertainty or unpredictability and thus high complexity [32].

While it is true that entropy and complexity are both alternative measures for information, entropy actually measures the “expected” (average rate of) information inherent in the possible outcomes of a stochastic source of data (*i.e.* a random variable) [24] [31] and therefore rests upon probability distributions [24] [31]. Entropy, therefore, measures the expectation of complexity [31], while kolmogorov complexity measures complexity itself. This subtle distinction can lead to differing results between the two metrics, such that high complexity is not always equivalent to high entropy [33]. Several authors speculate that the relationship between complexity and entropy is unimodal (*i.e.* high complexity values correspond with intermediate entropy values [23] and highly ordered states (low complexity) which are very predictable (low entropy) [23] are, oddly

similar to highly disordered states (high complexity) since they tend to be just as simple on average [12] (low entropy). Others highlight there is no one-to-one relationship between complexity and entropy, rather they have a one-to-many or many-to-one relationship [23] (*i.e.* given an entropy value, there are many possible complexity values, and vice-versa). It is, thus, possible that no universal relationship between complexity and entropy exists and any perceived relationship depends purely on the measured sequence [23].

A major advantage of information entropy is that it is computable [4]. This may be the reason entropy has been more widely adopted as the dominant measure of complexity [32]. However, a major disadvantage of entropy is that, unlike Kolmogorov complexity, it lacks the power of a universal measure of complexity [4]. Entropy is unable to take into account spatial or structural characteristics of 2D patterns present in CA spacetime configurations [32] which directly contrasts with our intuitive perception of complexity in patterns and can be problematic as a complexity measure for CAs [32] (see Figure 2).

Computing Complexity. While algorithmic information theory provides a natural choice for an objective measure of complexity [8] [22] [32], Kolmogorov complexity, K_U , is an incomputable function [1] [3] (or lower semi-computable [4]) and only approximable [1]. Kolmogorov Complexity is thus reliant on methods of approximation [22] (*i.e.* using a lossless compression ratio, CR , to indirectly show the complexity [4] [22], since regular patterns are more compressible [8] [31] than random ones, which tend toward incompressibility [3] [4] [8], and thus the difference in length between the compressed and uncompressed sequence is a good approximation of its computer program size [3] [31]).

$$K_U(x) \approx \frac{1}{CR(x)} \quad (6)$$

Depending on the compression method used, approximations can differ slightly [22]. Examples of general lossless compression algorithms used to approximate Kolmogorov Complexity [3] [4] include Huffman encoding (more frequently occurring data, *i.e.* pixels, are encoded with fewer bits) [6], Lempel-Ziv (LZ77 used with Huffman encoding in GZIP [8] [34] and LZ78 [3] variants), Lempel-Ziv-Welch (LZW) algorithm (a combination of Lempel-Ziv and Huffman

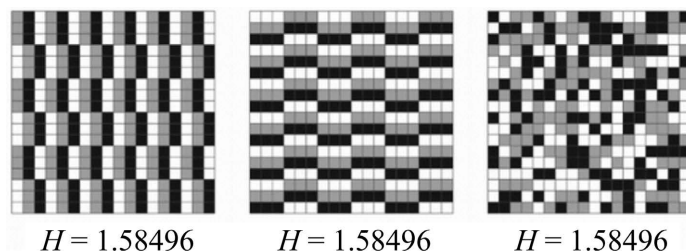


Figure 2. Since entropy ignores structural differences, it can contradict our intuitive concept of complexity. Consider the following three patterns with identical entropies (H) of 1.58496 [32], in spite of the periodic structure of the first two patterns which make them appear far less complex than the third.

encoding [3] [8]), BZIP2 (Burrows-Wheeler block sorting compression and Huffman encoding) [8], Run-length encoding [5] [6] (a simple and fast algorithm for compressing any type of data irregardless of its content [6]), Entropy encoding [6], Arithmetic encoding [6], etc. Block Decomposition Method (BDM) [4] [8] [22] is one of the few alternatives that approximates kolmogorov complexity without compression. BDM uses the frequency of the occurrence of a string [4] [8] instead (where frequency is connected to complexity based on a relation established by the coding theorem [4] [8] which states that the more frequent something is, the less unpredictable it is—*i.e.* the less information/complexity it has).

Despite the many lossless compression techniques available, almost all of them were designed for one-dimensional data (*i.e.* strings and sequences) [5] [8], which make them fundamentally unsuitable for measuring two-dimensional spacetime evolutions (which naturally contain patterns and nested structures [5] that are only distinguishable in their natural dimension and not in lower dimensions [8]). Even when these methods are scaled to two-dimensional data (*i.e.* by reading in one dimension of data at a time), geometrical regularities get destroyed in the process (e.g. the triangular structures in rule 30—see **Figure 3**). Defining a 2-dimensional complexity measure has been identified as an open problem in complexity science [8].

2.4. Lossy vs Lossless Compression

Compression techniques are broadly categorised into either lossy or lossless compression [6] [35]. The reconstruction loss, ε , is defined as the difference between the original image (*i.e.* the spacetime evolution, s) and the image reconstructed from the compression process, $\hat{s}$.

$$\varepsilon : \mathbb{R}^{W \times T} \times \mathbb{R}^{W \times T} \rightarrow \mathbb{R}^+$$

$$\varepsilon(s, \hat{s}) = \sum_{w=0}^{W-1} \sum_{t=0}^{T-1} |s_{w,t} - \hat{s}_{w,t}| \quad (7)$$

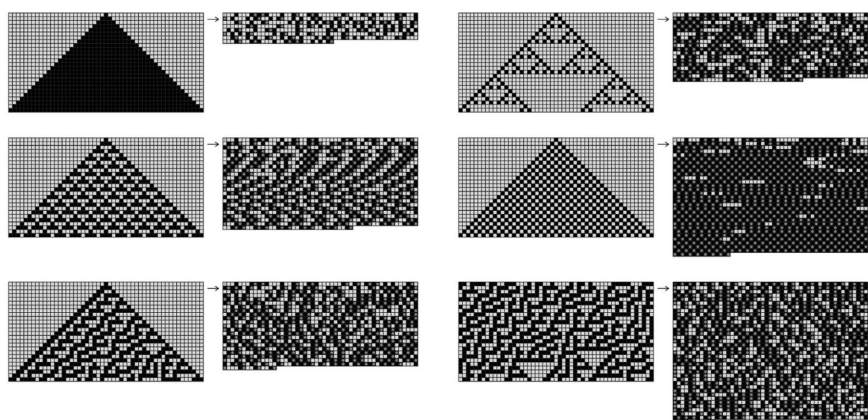


Figure 3. Examples of run-length encoding compressions of various elementary cellular automata spacetimes [5]. 2D triangular patterns or nested patterns fail to be sufficiently compressed.

Lossless compression reconstructs an image identical to the original, but lossy compression only reconstructs similar images since some information is lost during the compression [6] (*i.e.* the compression is “lossy” if $\varepsilon(s, \hat{s}) \neq 0$ and “lossless” if $\varepsilon(s, \hat{s}) = 0$).

2.5. Lossy Compression Using the Fast Fourier Transform

The Fast Fourier Transform (FFT) is a computationally efficient algorithm for computing the Discrete Fourier Transform (a mathematical representation of a signal in its frequency domain [6]) which even has a natural extension for processing 2D data [10]. The 2D-FFT, $\mathcal{F}$, transforms a real-valued image matrix, $r \in \mathbb{R}^{W \times T}$, into complex-valued coefficients, $z \in \mathbb{C}^{U \times V}$, that encode the information of the original image [6] from the spatial domain into the frequency domain [6]. The Discrete 2D-FFT is used for simplicity, since it returns the sum of a finite set of coefficients, z , large enough to fully describe the original image, r . A spacetime evolution, $s \in \{0, 1\}^{W \times T} \subset \mathbb{R}^{W \times T}$, is also a real-valued image, r (since binary-values are a subset of the reals).

$$\begin{aligned}\mathcal{F} : \mathbb{R}^{W \times T} &\rightarrow \mathbb{C}^{U \times V} \\ z &= \mathcal{F}(s)\end{aligned}$$

The 2D-FFT, $\mathcal{F}$, of each coefficient is calculated as:

$$[\mathcal{F}(s)]_{u,v} = \frac{1}{(W-1)(T-1)} \sum_{w=0}^{W-1} \sum_{t=0}^{T-1} s_{w,t} \cdot e^{-i2\pi\left(\frac{u \cdot w}{W} + \frac{v \cdot t}{T}\right)} \quad (8)$$

where w, t and u, v represent spatial and frequency coordinates respectively and the exponential term represents the linearly-independent, orthogonal basis pattern (sine and cosine waves with increasing frequencies).

$e^{-i2\pi\left(\frac{u \cdot w}{W} + \frac{v \cdot t}{T}\right)} = \cos\left(2\pi\left(\frac{u \cdot w}{W} + \frac{v \cdot t}{T}\right)\right) - i \sin\left(2\pi\left(\frac{u \cdot w}{W} + \frac{v \cdot t}{T}\right)\right)$. When every coefficient's basis patterns are combined together, they reconstruct the original image (as in **Figure 4**).

Frequency Filtering. Traditional (*i.e.* lossy [6]) FFT compression, which underlies much of signal processing, image processing and image compression [6], uses a process called fourier filtering; an image, s , is fourier transformed, $\mathcal{F}$, and thus encoded, z into the frequency domain where a frequency filter processes it. Let $m \in \{0, 1\}^{U \times V}$ be a binary-mask used for filtering specific coefficients, then the filtered coefficients, $\hat{z} \in \mathbb{C}^{U \times V}$, are the product of the fourier coefficients and the mask: $\hat{z} = m \cdot z$.

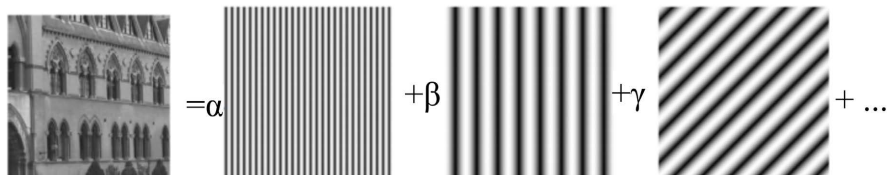


Figure 4. The 2D-FFT is able to transform any image into a linear combination of basis patterns (e.g. sine and cosine waves).

Compression Ratio. The compression ratio, CR , (the ratio of the original image to its compressed form [6]) can be calculated by comparing the image's representation length, l , before and after filtering.

$$CR: \mathbb{R}^{W \times T} \rightarrow \mathbb{Z}^+$$

$$CR(s) = \frac{l(\mathcal{F}(s))}{l(m \cdot \mathcal{F}(s))} \quad (9)$$

where the representation length, l , is the number of non-zero fourier frequencies.

$$l: \mathbb{C}^{U \times V} \rightarrow \mathbb{Z}^+$$

$$l(z) = \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} 1 - \mathbb{I}_{\{0\}}(z_{u,v}) \quad (10)$$

where $\mathbb{I}$ is the indicator function, such that:

$$\mathbb{I}_{\{0\}}(\alpha) = \begin{cases} 1 & \text{if } \alpha \in \{0\} \\ 0 & \text{otherwise} \end{cases}$$

A higher compression ratio, CR , signifies the image has been compressed more due to more frequencies being filtered (zeroed) and thus a filtered representation with a smaller length, l .

Inverse Transformation. By applying the inverse 2D-FFT, $\mathcal{F}^{-1}$, to the filtered coefficients, $\hat{z}$, (*i.e.* the compressed representation), one obtains a reconstruction of the original image without any noticeable loss in image quality [10] (as the filtering process discards coefficients of little significance representing negligible image data).

$$\mathcal{F}^{-1}: \mathbb{C}^{U \times V} \rightarrow \mathbb{R}^{W \times T}$$

$$\left[\mathcal{F}^{-1}(\hat{z}) \right]_{w,t} = \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} \hat{z}_{u,v} \cdot e^{i2\pi \left(\frac{w \cdot u}{U} + \frac{t \cdot v}{V} \right)} \quad (11)$$

3. Method

Fourier compressions are normally lossy, excluding them from Kolmogorov Complexity approximations. However, we propose a novel adaptation which allows for lossless fourier compression of black and white images. Our method is based on traditional frequency filtering, but adapted such that the filtered coefficients, $\hat{z}$, result in absolutely no reconstruction loss, $\varepsilon(s, \hat{s}) = 0$.

3.1. Decompression

The final, binarised reconstruction of the spacetime evolution, $\hat{s}$, is obtained by applying the decompression function (the inverse of the compression function, C^{-1}) to the filtered coefficients, $\hat{z}$. The decompression, C^{-1} , is simply the inverse 2D-FFT, $\mathcal{F}^{-1}$, with an additional quantisation step, Q_θ .

$$C^{-1}: \mathbb{C}^{U \times V} \rightarrow \{0,1\}^{W \times T}$$

$$C^{-1}(\hat{z}) = Q_{\theta}(\mathcal{F}^{-1}(\hat{z})) \quad (12)$$

3.2. Quantisation

The quantisation step, Q_{θ} , is used to convert the real-valued reconstruction of the spacetime evolution, $\hat{r} \in \mathbb{R}^{W \times T}$, into a binary-valued reconstruction, $\hat{s} \in \{0, 1\}^{W \times T}$, by setting each value to 1 if it is equal to or above the quantisation threshold, $\theta \in \mathbb{R}$, and 0 if below.

$$Q_{\theta} : \mathbb{R}^{W \times T} \rightarrow \{0, 1\}^{W \times T}$$

$$[Q_{\theta}(\hat{r})]_{w,t} = \begin{cases} 1 & \text{if } \hat{r}_{w,t} \geq \theta \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

This quantisation step, Q_{θ} , limits our lossless compression to black and white images (such as spacetime evolutions, s). However, this limitation is also a source of strength because the quantised reconstruction, $\hat{s}$, ends up being far more robust to small errors that would have otherwise resulted from the frequency filtering step. During the quantisation step, Q_{θ} , a certain level of noise gets removed from the reconstructed image, $\hat{r}$, and some loss (that would otherwise result from frequency filtering) ends up being corrected. Small differences in the noisy (*i.e.* lossy) reconstructed image, $\hat{r}$, get removed so that the final, binary-valued, reconstruction, $\hat{s}$, is error-free (*i.e.* lossless).

Example. The following example demonstrates how a lossy reconstruction may become lossless as a side effect of the quantisation step. Let s' be a simple spacetime evolution such that $s'_{w,t} = 1, \forall w \in [0, W), \forall t \in [0, T)$. Let $\hat{r}'$ be a noisy reconstruction such that $\hat{r}'_{w,t} = 1 \pm \Delta, \forall w \in [0, W), \forall t \in [0, T)$, where $\Delta \in \mathbb{R}^+$ represents some noise. Therefore, this reconstruction is lossy,

$$\mathcal{E}(s', \hat{r}') = \sum_{w=0}^{W-1} \sum_{t=0}^{T-1} |s'_{w,t} - \hat{r}'_{w,t}| = WT\Delta \neq 0.$$

Now let the quantisation threshold $\theta \leq (1 - \Delta)$, such that the binarised reconstruction becomes

$$\hat{s}' = Q_{\theta}(\hat{r}') = [\hat{s}'_{w,t} = 1 \mid \forall w \in [0, W-1), \forall t \in [0, T-1)],$$

since $\hat{r}'_{w,t} \geq \theta$ holds for all values. Thus, the binarised reconstruction is lossless

$$\Rightarrow \mathcal{E}(s', \hat{s}') = \sum_{w=0}^{W-1} \sum_{t=0}^{T-1} |s'_{w,t} - \hat{s}'_{w,t}| = \sum_{w=0}^{W-1} \sum_{t=0}^{T-1} |1 - 1| = 0.$$

3.3. Lossless Frequency Filtering

We define the space of lossless frequency filter masks, $\mathcal{M}^{\phi}$, as all frequency filter masks that do not result in any reconstruction loss, $\mathcal{E}(s, \hat{s}) = 0$.

$$\mathcal{M} = \left\{ m \in \{0, 1\}^{U \times V} \mid \mathcal{E}\left(s, C^{-1}\left(m \cdot \mathcal{F}(s)\right)\right) = 0 \right\} \quad (14)$$

For a given image, s , there will always exist at least one, lossless, frequency filter mask, $\exists m \in \mathcal{M}$, and that is the filter which leaves all frequencies unfiltered, $m' = \{1\}^{U \times V}$, such that $m' \cdot \mathcal{F}(s) = \mathcal{F}(s)$ (since no information is removed during filtering, its corresponding reconstruction will have no loss). However, this trivial filter has a compression ratio of 1, meaning it produces a very poor compression (the higher the compression ratio, the better the compression),

³The space of lossless frequency filter masks, $\mathcal{M}$, is actually increased due to the quantisation step, Q_{θ}

even though it is lossless, $CR(s) = \frac{l(\mathcal{F}(s))}{l(m' \cdot \mathcal{F}(s))} \Rightarrow \frac{l(\mathcal{F}(s))}{l(\mathcal{F}(s))} = 1$.

3.4. Optimal Lossless Filter

A good compression method aims to maximise the compression ratio, CR , so an optimal lossless filter mask, m^* , will filter as many frequencies as possible without becoming lossy (for a maximal compression ratio, CR). In other words, the optimal lossless filter mask, m^* , is the mask producing a filtered representation, $\hat{z}$, with the smallest representation length, l (since the compression ratio is inversely proportional to the representation length, $CR \propto \frac{1}{l(\hat{z})}$, therefore the lower the filtered representation length, $l(\hat{z})$, the higher the compression ratio, CR).

$$m^* = \min \{l(m \cdot \mathcal{F}(s)) \mid m \in M\} \quad (15)$$

(In **Appendix C** we include an alternative, non-exhaustive search method for more quickly approaching the optimal lossless filter mask, m^*).

3.5. Compression

The compression method, C , is defined as the optimal lossless filter, m^* , applied to the fourier transform, $\mathcal{F}$, of a given spacetime evolution, s .

$$\begin{aligned} C : \{0,1\}^{W \times T} &\rightarrow \mathbb{C}^{U \times V} \\ C(s) &= m^* \cdot \mathcal{F}(s) \end{aligned} \quad (16)$$

This compression method, C , is set apart from traditional fourier image compression by virtue of its lossless nature (it will always produce an identical reconstruction of the original spacetime evolution given)

Lemma 1. $C^{-1}(C(s)) = s$

Proof. Since $m^* \in M$ by definition 15, then $\varepsilon(s, C^{-1}(m^* \cdot \mathcal{F}(s))) = 0$ by definition 14, then we get $\sum_{w=0}^{W-1} \sum_{t=0}^{T-1} |s_{w,t} - [C^{-1}(m^* \cdot \mathcal{F}(s))]_{w,t}| = 0$ by substituting 7 $\Rightarrow s = C^{-1}(m^* \cdot \mathcal{F}(s))$, then by substituting 16 we get: $s = C^{-1}(C(s))$

□

3.6. Complexity Metric

Using our novel lossless compression method, C , we can approximate the Kolmogorov Complexity, K , of a given spacetime evolution, s , as:

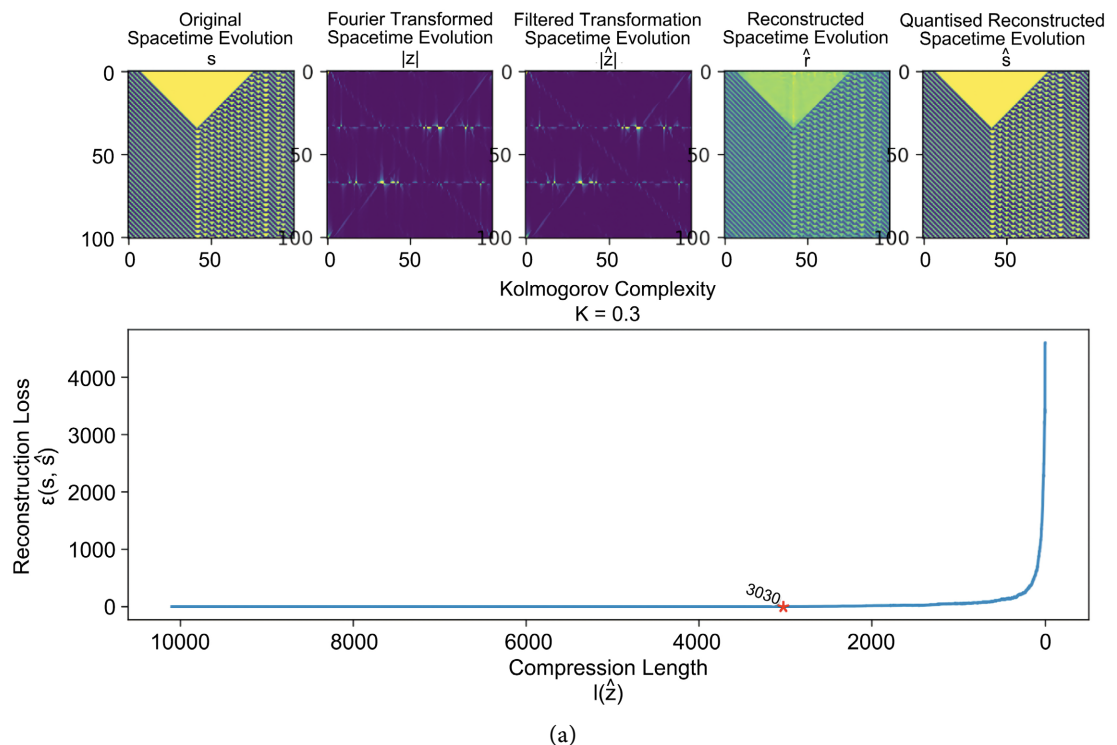
$$\begin{aligned} K_U : \{0,1\}^{W \times T} &\rightarrow \mathbb{R}^+ \\ K_U(s) &\approx \frac{l(C(s))}{l(\mathcal{F}(s))} \end{aligned} \quad (17)$$

Since, by definition 6, we can approximate Kolmogorov complexity, K_U , us-

ing any lossless compression ratio, $K_U(x) \approx \frac{1}{CR(s)}$, then $\Rightarrow \frac{l(m^* \cdot \mathcal{F}(s))}{l(\mathcal{F}(s))}$ by substituting 9 and $\Rightarrow \frac{l(C(s))}{l(\mathcal{F}(s))}$ by substituting 16.

Examples (Figure 5). Below are some examples of the Kolmogorov Complexity measured for a couple of 1D-ECA spacetime evolutions exhibiting different classes of dynamics. The examples show the steps involved to losslessly compress a given image by fourier transforming it into the frequency domain, filtering coefficients and then reconstructing the spacetime via the inverse fourier transform and binary quantisation step. Each example illustrates the following information (laid out as in Figure 6).

- 1) An image of the given 1D-ECA spacetime evolution, s .
- 2) An image of the magnitude of its fourier coefficients, z , returned by the 2D-FFT, $\mathcal{F}$, (*i.e.* the real components. The phase image from the imaginary components is not shown for simplicity).
- 3) An image of the magnitude of its filtered coefficients, $\hat{z}$, produced by the lossless filter mask, m^* , (*i.e.* the real components. The phase image from the imaginary components is not shown for simplicity).
- 4) An image of the reconstructed spacetime evolution, $\hat{r}$, returned by the inverse 2D-FFT, $\mathcal{F}^{-1}$.
- 5) An image of the binarised reconstructed spacetime evolution, $\hat{s}$, returned by the quantisation step, Q_θ (using a default threshold, θ , set to 0.5).
- 6) The resulting Kolmogorov Complexity, K , of the spacetime evolution, s .



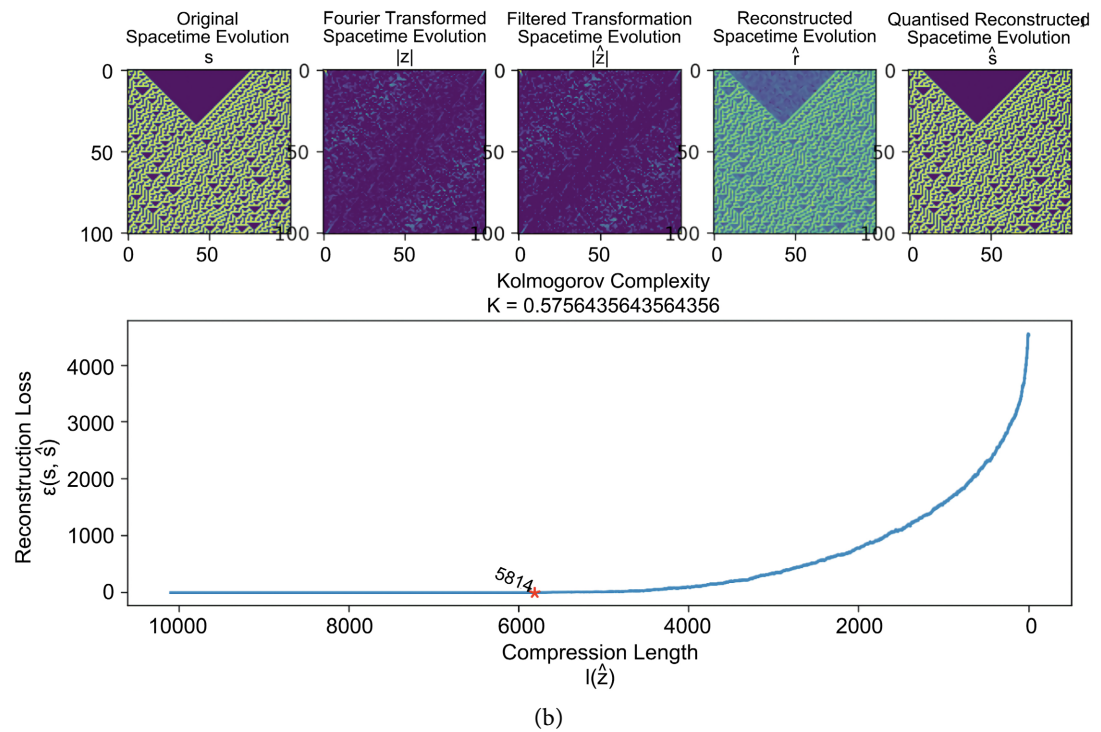


Figure 5. An example of a relatively simple (a) and complex (b) spacetime evolution. (a) ECA rule 145 (IC 8932) has a frequency representation which relies on relatively few basis patterns, which mean more coefficients can be filtered without affecting the reconstruction loss, allowing for a higher compression ratio and a lower kolmogorov complexity; (b) ECA rule 30 (IC 8932) requires relatively more basis patterns in its frequency representation and so fewer coefficients can be filtered before the reconstruction loss becomes positive, resulting in a lower compression ratio and a higher kolmogorov complexity.

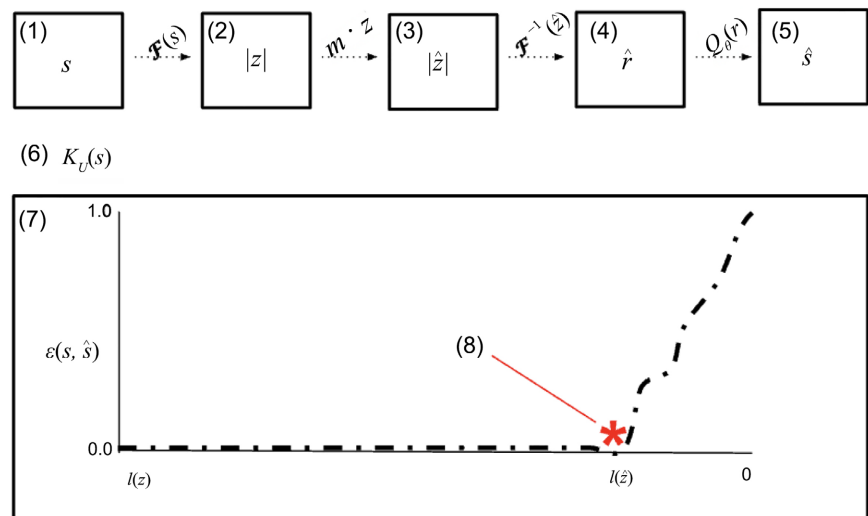


Figure 6. Layout of information for each example.

7) The loss from iteratively filtering all N ordered coefficients, one by one (to search for a near-optimal, lossless filter). For more details, see **Appendix C**.

8) The best lossless filter found (*i.e.* the smallest compression length before the filters become lossy).

4. Experimental Setup

4.1. Complexity Metrics and Compression Techniques

We evaluate the proposed, lossless fourier compression for approximating kolmogorov complexity alongside several existing compression methods: Huffman encoding, Run-length encoding (RLE), Lempel-Ziv-Markov-Chain Algorithm (LZMA), ZLIB, GZIP, and BZIP2, as well as one alternative information-theoretic complexity metric: Block Decomposition Method (BDM). We restrict our evaluation to complexity metrics applied to ECA spacetime evolutions (as opposed to those applied to the ECA transition rules).

4.2. Dataset

Each metric is run over a dataset of spacetime evolutions (generated from all 256 ECA rules, always from the same initial configuration corresponding to the binary value of 8932 on 100-cell wide lattices, iterated for 110 time steps, ignoring the initial 10 steps to allow for sufficient transient time for simpler evolutions to settle into their characteristic patterns) and then augmented with several symmetry transformations (*i.e.* identity, reflection, inversion and rotation—see **Figure 7**) comprising a total of 768 spacetime patterns ($256 \text{ rules} \times 1 \text{ initial configuration} \times 4 \text{ symmetry transformations}$).

4.3. Code

We conduct all experiments in python 3, simulating elementary cellular automata using `eca`⁴ (our open-source python package published on pypi). All complexity metrics tested, as well as the code used to generate and augment the original dataset, can be found on our github repository: `ECA_complexity_metrics`.

5. Results and Evaluation

Three evaluation phases were carried out:

- 1) Pearsons Correlation
- 2) Limits of Complexity

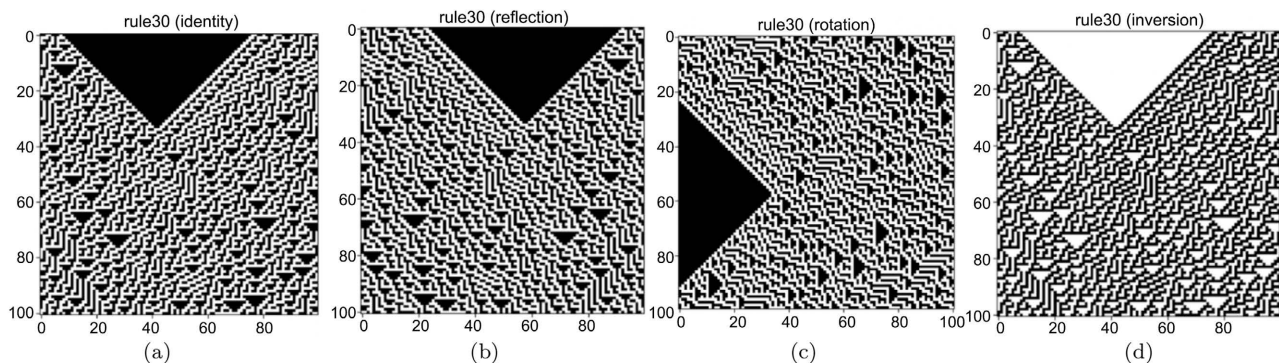


Figure 7. The spacetime evolution of rule 30 and its augmented symmetry transformations. (a) ECA Rule 30 (Identity); (b) ECA Rule 30 (Reflection); (c) ECA Rule 30 (Rotation); (d) ECA Rule 30 (Inversion).

⁴<https://pypi.org/project/eca/>.

3) Symmetry Test

5.1. Pearsons Correlation

First, we observe how similar the proposed metric is to the existing complexity metrics (**Table 1**). Taking the Pearson correlation coefficient reveals the proposed method to be strongly correlated to all existing complexity metrics tested (>0.8), with the exception of RLE (which has a weaker positive correlation ≈ 0.4).

5.2. Limits of Complexity

The second test assumes that the least and most complex patterns possible are that of absolute nothingness and absolute randomness, respectively. Therefore, by measuring a blank image and an image of white-noise, we can obtain the lower and upper limits of complexity (see **Figure 8**). The minimum and maximum complexity measurements slightly vary from metric to metric (see **Table 2**), however, their measurements for all images in the dataset should not exceed these complexity limits.

Table 1. The proposed complexity metric shows a positive correlation to existing complexity metrics.

Pearsons Correlation	
Fourier Transform (ours)	1.000000
ZLIB	0.857042
GZIP	0.857042
BZIP2	0.854448
BDM	0.849258
Huffman Encoding	0.830718
LZMA	0.809293
RLE	0.427328

Table 2. Limits of complexity result. The minimum and maximum complexities for each complexity metric obtained by measuring a blank image of nothingness and a random image of white-noise. The metrics which exceed their own limits are highlighted in red.

	min	max	count (<min)	count (>max)
Fourier Transform (ours)	0.000000	0.648200	0	0
BZIP2	0.004400	0.170200	0	0
BDM	0.000000	9.284512	0	0
GZIP	0.004600	0.171100	80	0
Huffman Encoding	0.333400	1.038100	80	0
LZMA	0.010800	0.166400	80	0
RLE	0.004000	0.494000	82	163
ZLIB	0.003400	0.169900	80	0

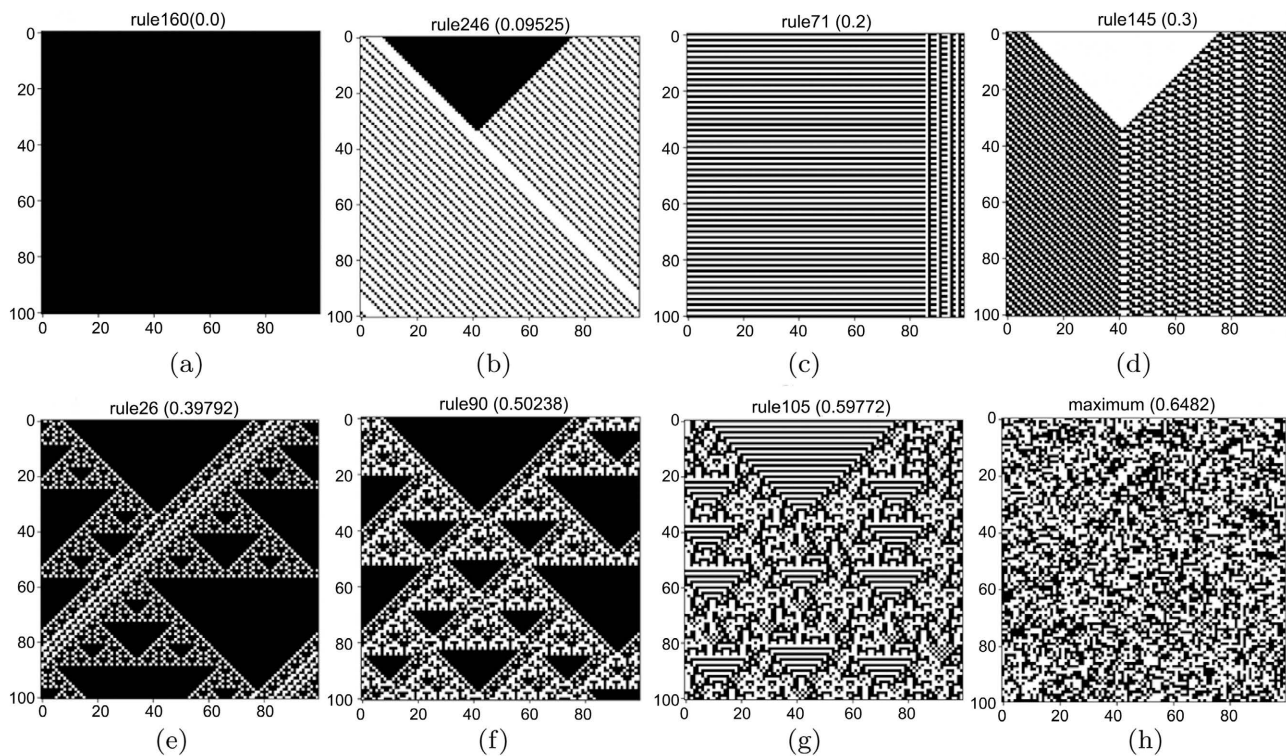


Figure 8. A sample of increasingly complex spacetime evolutions, as measured by our novel method.

BDM, BZIP2 and Fourier Transform (our proposed method) pass this simple sanity check. However, the remaining lossless compression methods (ZLIB, GZIP, LZMA, RLE and Huffman Encoding) exceed their own lower complexity limits, while RLE was the only method which also exceeded its own upper complexity limit, indicating these methods produce incoherent approximations of complexity.

5.3. Symmetry Test

Lastly, we perform a symmetrical equivalence test, which assumes a group of symmetrical images to have equivalent complexity. In other words, spacetime evolutions which only differ by a symmetrical transformation should measure the same complexity (this assumption is fairly prevalent in ECA literature whereby experiments, classifications and complexity measurements are performed on a reduced set of 88 symmetrically equivalent sets of rules, as opposed to using all 256 ECA rules [5] [36]). Therefore, an ideal complexity metric should produce identical measurements across symmetrical images and any differences in measurement can be attributed to imprecisions in the metric. To ensure we have perfectly symmetrical patterns within the dataset, we use image augmentation techniques to transform each rule's spacetime according to the following symmetry transformations:

- 1) identity transformation (to include the original spacetime pattern in the dataset).

Table 3. Symmetry test results. The average normalised complexity difference across symmetrically equivalent images (Symmetrical) and a breakdown by symmetry type (Inversion, Rotation, Reflection).

	Inversion	Rotation	Reflection	Symmetrical
BDM	0.000	0.000	0.000	0.000
Fourier Transform (ours)	0.001	0.011	0.010	0.022
GZIP	0.000	0.028	0.004	0.032
ZLIB	0.000	0.028	0.004	0.032
LZMA	0.000	0.027	0.009	0.036
Huffman Encoding	0.000	0.030	0.008	0.038
BZIP2	0.003	0.034	0.004	0.041
RLE	0.000	0.233	0.001	0.234

2) rotation (90 degrees anticlockwise).

3) reflection (about the vertical axis—*i.e.* left-right reflection).

4) black-white inversion (also referred to as negation or complementation [37]).

Our proposed method performed the best of all lossless compression methods tested and was beaten only by BDM, the alternative information-theoretic complexity metric which does not rely on lossless compression. BDM was the only metric measuring absolutely no difference across any symmetrically equivalent image. RLE was worst as it had the largest normalised complexity difference across symmetrical images on average (Table 3).

6. Conclusion

In this paper, we propose a novel adaptation of the 2D-FFT which produces lossless compressions of black and white images. This, in turn, is used to more accurately measure the Kolmogorov Complexity of ECA spacetime diagrams. To the best of our knowledge, we are the first to propose a lossless compression method based on the Fourier Transform. The proposed method is strongly correlated to existing complexity metrics (with a Pearson correlation >0.8) and, unlike many existing methods, is internally consistent as its measurements never exceed the upper and lower complexity limits set using the least and most complex patterns possible (*i.e.* that of nothingness and absolute randomness). A final symmetry-based test quantifiably demonstrates the proposed method's superiority over existing lossless compression methods. Our novel lossless compression method is thus a superior approximation for the kolmogorov complexity of ECA spacetime evolutions.

Acknowledgements

To my colleague and good friend, Nikolai Rozanov, thank you for the many stimulating discussions and invaluable mathematical insights. To my supervisor at

the University of York, Dr. Simon O’Keefe, thank you for the guidance throughout. To my dear mother, Elsa Weizemann, thank you for your massive enthusiasm and support. A big thank you to all my family and loved ones, especially my darling wife and daughter, Rouhy and Fahtima Zahra.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Dubacq, J.-C., Durand, B. and Formenti, E. (2001) Kolmogorov Complexity and Cellular Automata Classification. *Theoretical Computer Science*, **259**, 271-285. [https://doi.org/10.1016/S0304-3975\(00\)00012-8](https://doi.org/10.1016/S0304-3975(00)00012-8)
- [2] Baetens, J. and De Baets, B. (2011) Towards the Full Lyapunov Spectrum of Elementary Cellular Automata. *AIP Conference Proceedings*, **1389**, 981-986. <https://doi.org/10.1063/1.3637774>
- [3] Zenil, H. (2009) Compression-Based Investigation of the Dynamical Properties of Cellular Automata and Other Systems. arXiv preprint arXiv:0910.4042. <https://doi.org/10.1142/S0218127413501599>
- [4] Zenil, H. and Villarreal-Zapata, E. (2013) Asymptotic Behavior and Ratios of Complexity in Cellular Automata. *International Journal of Bifurcation and Chaos*, **23**, Article ID: 1350159. <https://doi.org/10.1115/1.1553433>
- [5] Wolfram, S. and Gad-el Hak, M. (2003) A New Kind of Science. *Applied Mechanics Reviews*, **56**, B18-B19.
- [6] Gomathi and Santhanam (2018) Performance Analysis of Lossless and Lossy Image Compression Techniques for Human Object. *International Journal of Applied Engineering Research*, **13**, 11715-11723.
- [7] Allis, N., Dumont, J.P., Heiss, F.J. and Reiter, C.A. (2000) Fast Fourier Transforms, Diffraction Patterns, and J. *The Journal of the British APL Association*, **16**, 111-116.
- [8] Zenil, H., Soler-Toscano, F., Delahaye, J.-P. and Gauvrit, N. (2015) Two-Dimensional Kolmogorov Complexity and an Empirical Validation of the Coding Theorem Method by Compressibility. *PeerJ Computer Science*, **1**, e23. <https://doi.org/10.7717/peerj-cs.23>
- [9] Rasheed, M.H., Salih, O.M., Siddeq, M.M. and Rodrigues, M.A. (2020) Image Compression Based on 2D Discrete Fourier Transform and Matrix Minimization Algorithm. *Array*, **6**, Article ID: 100024. <https://doi.org/10.1016/j.array.2020.100024>
- [10] Brunton, S. (2000) Image Compression and the FFT. <https://www.youtube.com/watch?v=gGEBUdM0PVc>
- [11] Adamatzky, A. and Martinez, G.J. (2010) On Generative Morphological Diversity of Elementary Cellular Automata. *Kybernetes*, **39**, 72-82. <https://doi.org/10.1108/03684921011021282>
- [12] Dürr, C., Rapaport, I. and Theyssier, G. (2004) Cellular Automata and Communication Complexity. *Theoretical Computer Science*, **322**, 355-368. <https://doi.org/10.1016/j.tcs.2004.03.017>
- [13] Culik II, K., Hurd, L.P. and Yu, S. (1990) Computation Theoretic Aspects of Cellular Automata. *Physica D: Nonlinear Phenomena*, **45**, 357-378. [https://doi.org/10.1016/0167-2789\(90\)90194-T](https://doi.org/10.1016/0167-2789(90)90194-T)

- [14] Li, W. and Packard, N. (1990) The Structure of the Elementary Cellular Automata Rule Space. *Complex Systems*, **4**, 281-297.
- [15] Wuensche, A. (1994) Complexity in One-D Cellular Automata: Gliders, Basins of Attraction and the Z Parameter. University of Sussex, School of Cognitive and Computing Sciences, Brighton.
- [16] Sutner, K. (2009) Classification of Cellular Automata. In: Meyers, R., Ed., *Encyclopedia of Complexity and Systems Science*, Springer, New York, 755-768.
https://doi.org/10.1007/978-0-387-30440-3_50
- [17] Cattaneo, G. and Vogliotti, C.Q. (1997) The “Magic” Rule Spaces of Neural-Like Elementary Cellular Automata. *Theoretical Computer Science*, **178**, 77-102.
[https://doi.org/10.1016/S0304-3975\(96\)00053-9](https://doi.org/10.1016/S0304-3975(96)00053-9)
- [18] Borriello, E. and Imari Walker S. (2017) An Information-Based Classification of Elementary Cellular Automata. *Complexity*, **2017**, Article ID: 1280351.
<https://doi.org/10.1155/2017/1280351>
- [19] Powley, E.J. and Stepney, S. (2009) Automorphisms of Transition Graphs for Elementary Cellular Automata. *Journal of Cellular Automata*, **4**, 125-136.
- [20] De Sales, J., Martins, M. and Moreira, J. (1997) One-Dimensional Cellular Automata Characterization by the Roughness Exponent. *Physica A: Statistical Mechanics and Its Applications*, **245**, 461-471. [https://doi.org/10.1016/S0378-4371\(97\)00320-8](https://doi.org/10.1016/S0378-4371(97)00320-8)
- [21] Ruivo E.L. and de Oliveira P.P. (2012) Spectral Similarity among Elementary Cellular Automata. In: Formenti, E., Ed., *Proceeding of the 18th International Workshop on Cellular Automata and Discrete Complex Systems: Exploratory Papers Proceedings*, Rapport de Recherche I3S-ISR: I3S/RR-2012-04-FR, 89-98.
- [22] Ewert, T. (2019) A Measure for the Complexity of Elementary Cellular Automata. *Complex Systems*, **28**, Article No. 219.
<https://doi.org/10.25088/ComplexSystems.28.2.219>
- [23] Li, W. (1991) On the Relationship between Complexity and Entropy for Markov Chains and Regular Languages. *Complex Systems*, **5**, 381-399.
- [24] Teixeira, A., Matos, A., Souto, A. and Antunes, L. (2011) Entropy Measures vs. Kolmogorov Complexity. *Entropy*, **13**, 595-611. <https://doi.org/10.3390/e13030595>
- [25] Albantakis, L. and Tononi, G. (2015) The Intrinsic Cause-Effect Power of Discrete Dynamical Systems—From Elementary Cellular Automata to Adapting Animats. *Entropy*, **17**, 5472-5502. <https://doi.org/10.3390/e17085472>
- [26] Lei, Q., Lee, J., Huang, X. and Kawasaki, S. (2021) Entropy-Based Classification of Elementary Cellular Automata under Asynchronous Updating: An Experimental Study. *Entropy*, **23**, 209. <https://doi.org/10.3390/e23020209>
- [27] Chua, L.O., Yoon, S. and Dogaru, R. (2002) A Nonlinear Dynamics Perspective of Wolfram’s New Kind of Science Part I: Threshold of Complexity. *International Journal of Bifurcation and Chaos*, **12**, 2655-2766.
<https://doi.org/10.1142/S0218127402006333>
- [28] Machicao, J., Ribas, L.C., Scabini, L.F. and Bruno, O.M. (2018) Cellular Automata Rule Characterization and Classification Using Texture Descriptors. *Physica A: Statistical Mechanics and Its Applications*, **497**, 109-117.
<https://doi.org/10.1016/j.physa.2017.12.072>
- [29] Freire, J.G., Brison, O.J. and Gallas, J.A. (2010) Complete Sets of Initial Vectors for Pattern Growth with Elementary Cellular Automata. *Computer Physics Communications*, **181**, 750-755. <https://doi.org/10.1016/j.cpc.2009.12.007>
- [30] J.G. Freire, Brison, O.J. and Gallas, J.A. (2009) Exact Quantification of the Com-

- plexity of Spacewise Pattern Growth in Cellular Automata. *Journal of Physics A: Mathematical and Theoretical*, **42**, Article ID: 395003. <https://doi.org/10.1088/1751-8113/42/39/395003>
- [31] Grunwald, P. and Vitányi, P. (2004) Shannon Information and Kolmogorov Complexity. arXiv preprint: cs/0410002.
- [32] Ali Javaheri Javid, M., Blackwell, T., Zimmer, R. and Majid al Rifaie, M. (2016) Analysis of Information Gain and Kolmogorov Complexity for Structural Evaluation of Cellular Automata Configurations. *Connection Science*, **28**, 155-170. <https://doi.org/10.1080/09540091.2016.1151861>
- [33] Shalizi, C.R., Haslinger, R., Rouquier, J.-B., Klinkner, K.L. and Moore, C. (2006) Automatic Filters for the Detection of Coherent Structure in Spatiotemporal Systems. *Physical Review E*, **73**, Article ID: 036104. <https://doi.org/10.1103/PhysRevE.73.036104>
- [34] Rajeswaran, K. and Winberg, S. (2013) Lossless Compression of Ska Data Sets. *Communications and Network*, **5**, 369-378. <https://doi.org/10.4236/cn.2013.54046>
- [35] Alarabeyyat, A., Al-Hashemi, S., Khmour, T., Btoush, M.H., Bani-Ahmad S., Al-Hashemi R., Bani-Ahmad, S., *et al.* (2012) Lossless Image Compression Technique Using Combination Methods. *Journal of Software Engineering and Applications*, **5**, 752-763. <https://doi.org/10.4236/jsea.2012.510088>
- [36] Krawczyk, M.J. (2015) New Aspects of Symmetry of Elementary Cellular Automata. *Chaos, Solitons & Fractals*, **78**, 86-94. <https://doi.org/10.1016/j.chaos.2015.07.012>
- [37] Martinez, G.J. (2013) A Note on Elementary Cellular Automata Classification. arXiv preprint arXiv: 1306.5577.

Appendices

A. Contributions

- A novel adaptation of the 2D-FFT for lossless compressions of black and white images.
- More accurate and consistent approximations of the kolmogorov complexity for ECA spacetime evolutions.
- First to propose a lossless compression method based on the Fourier Transform.
- A symmetry-based test demonstrating the superiority of the proposed method over existing lossless compression methods.

B. Summary of Novel Method

Kolmogorov Complexity Approximation. We propose the following fourier transform compression based approximation of Kolmogorov Complexity:

$$K_U : \{0,1\}^{W \times T} \rightarrow \mathbb{R}^+$$

$$K_U(s) \approx \frac{l(C(s))}{l(\mathcal{F}(s))}$$

where l is the number of non-zero coefficients in the 2D fourier representation with dimensions $U \times V$ where $U, V \in \mathbb{Z}^+$ and $\mathbb{I}$ represents the indicator function

$$l : \mathbb{C}^{U \times V} \rightarrow \mathbb{Z}^+$$

$$l(z) = \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} 1 - \mathbb{I}_{\{0\}}(z_{u,v})$$

1D-ECA. $s \in \{0,1\}^{W \times T}$, $W, T \in \mathbb{Z}^+$ is a binary matrix of some image to be compressed. In our case, a spacetime evolution of a 1D Elementary Cellular Automata (1D-ECA).

$$s_{:,t} = g_{\rho}^{ot}(\chi')$$

where $\chi' \in \{0,1\}^W$ is the initial lattice configuration of the ECA, $t \in \mathbb{Z} : 0 \leq t \leq T-1$ and g_{ρ} is the global transition function of the ECA responsible for deterministically updating the lattice configuration, defined as:

$$g_{\rho} : \{0,1\}^W \rightarrow \{0,1\}^W$$

$$[g_{\rho}(\chi)]_w = \rho(\chi_{w-n}, \dots, \chi_w, \dots, \chi_{w+n})$$

where $w \in \mathbb{Z} : 0 \leq w \leq W-1$ and the neighbourhood radius $n \in \mathbb{Z} : n \geq 1$ (thus an ECA lattice width must be equal to or greater than one neighbourhood, $W \geq 2n+1$). Where $\rho : \{0,1\}^{2n+1} \rightarrow \{0,1\}$ is the local transition rule of the ECA (specified as a lookup table). For the simplest ECA with a neighbourhood radius $n=1$, there are $2^{2^{2n+1}} = 256$ possible local rules (each yielding unique global spacetime evolutions).

2D-FFT. $\mathcal{F}$, is the 2D Fast Fourier Transform (2D-FFT), defined as:

$$[\mathcal{F}(s)]_{u,v} = \frac{1}{(W-1)(T-1)} \sum_{w=0}^{W-1} \sum_{t=0}^{T-1} s_{w,t} \cdot e^{-i2\pi\left(\frac{u \cdot w}{W} + \frac{v \cdot t}{T}\right)}$$

where $u, v \in \mathbb{Z} : 0 \leq u \leq U-1, 0 \leq v \leq V-1$ are the frequency coordinates.

Lossless Compression. C is the compression method based on the 2D-FFT, defined as:

$$C : \{0,1\}^{W \times T} \rightarrow \mathbb{C}^{U \times V}$$

$$C(s) = m^* \cdot \mathcal{F}(s)$$

where $m^* \in M$ and $M = \{m \in \{0,1\}^{U \times V} \mid \mathcal{E}(s, C^{-1}(m \cdot \mathcal{F}(s))) = 0\}$ is the space of lossless filter masks and C^{-1} is defined as:

$$C^{-1} : \mathbb{C}^{U \times V} \rightarrow \{0,1\}^{W \times T}$$

$$C^{-1}(\hat{z}) = Q_{\theta}(\mathcal{F}^{-1}(\hat{z}))$$

The optimal lossless filter mask $m^* = \min(\{l(m \cdot z) \mid m \in M\})$ is the mask that leaves the fewest number of fourier coefficients z without the resulting image reconstruction $\hat{s}$ becoming lossy, where loss is defined as:

$$\mathcal{E} : \mathbb{R}^{W \times T} \times \mathbb{R}^{W \times T} \rightarrow \mathbb{R}^+$$

$$\mathcal{E}(s, \hat{s}) = \sum_{w=0}^{W-1} \sum_{t=0}^{T-1} |s_{w,t} - \hat{s}_{w,t}|$$

where $\hat{z} \in \mathbb{C}^{U \times V}$ is the filtered fourier transformed representation of the image and $\mathcal{F}^{-1} : \mathbb{C}^{U \times V} \rightarrow \mathbb{R}^{W \times T}$ is the inverse 2D-FFT and Q_{θ} is a quantisation step introduced to retrieve a binary spacetime evolution (also limiting our lossless compression method to black and white images), defined as:

$$Q_{\theta} : \mathbb{R}^{W \times T} \rightarrow \{0,1\}^{W \times T}$$

$$[Q_{\theta}(\hat{r})]_{w,t} = \begin{cases} 1 & \text{if } \hat{r}_{w,t} \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

where $\theta \in \mathbb{R}$ is the quantisation threshold.

Novelty. Note that C is distinct from traditional (*i.e.* lossy) fourier image compression by virtue of its lossless nature. In other words, it will always produce an identical reconstruction of the original image given, such that $C^{-1}(C(s)) = s$, which, for the first time, allows the Fourier transform to be used as a Kolmogorov Complexity approximation.

C. Fast Search for Near Optimal Lossless Filter Mask

Let ζ be the fourier frequencies z , ordered by magnitude:

$$\zeta = \left[|z_{f(0)}| \leq |z_{f(1)}| \leq \dots \leq |z_{f(UV)}| \right] \quad (18)$$

where f specifies an order of indices, $f : \mathbb{N} \rightarrow \mathbb{N}^2$, such that $\forall i \in [0, (U-1)(V-1)]$:

$$f(i) = \begin{pmatrix} u \\ v \end{pmatrix} \quad (19)$$

where $u \in [0, U)$, $v \in [0, V)$ and f^{-1} specifies the inverse mapping such that $\forall i \in [0, (U-1)(V-1)] : f^{-1}(f(i)) = i$.

Let $\mu : \mathbb{N} \rightarrow \{0, 1\}^{\overline{UV}}$ return a fourier filter mask designed to filter the positions of the k smallest fourier frequencies:

$$[\mu(k)]_{u,v} = \begin{cases} 1 & \text{if } f^{-1}(u, v) \geq k \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

Now since larger values of k result in more fourier frequencies being filtered (*i.e.* multiplied to zero) and fewer non-zero frequencies in the filtered fourier representation (*i.e.* smaller $l(\hat{z})$). Therefore, the smallest lossless filter mask $\mu(i^*)$ will be found at position i^* , beyond which the resulting filter masks start to become lossy

$$\varepsilon\left(s, C^{-1}\left(\mu(i^*) \cdot \mathcal{F}(s)\right)\right) = 0$$

and

$$\varepsilon\left(s, C^{-1}\left(\mu(i^* + 1) \cdot \mathcal{F}(s)\right)\right) > 0$$