mathematics

MDPI

Article

An Upper Bound for Locating Strings with High Probability Within Consecutive Bits of Pi

Víctor Manuel Silva-García [1], Manuel Alejandro Cardona-López [2,3,*] and Rolando Flores-Carapia [1,*]

[1] Centro de Innovación y Desarrollo Tecnológico en Cómputo, Instituto Politécnico Nacional, Mexico City 07738, Mexico; vsilvag@ipn.mx
[2] Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico City 07738, Mexico
[3] Escuela Superior de Ingeniería Mecánica y Eléctrica Unidad Zacatenco, Instituto Politécnico Nacional, Mexico City 07738, Mexico
* Correspondence: mcardonal2022@cic.ipn.mx (M.A.C.-L.); rfloresca@ipn.mx (R.F.-C.)

Abstract: Numerous studies on the number pi (π) explore its properties, including normality and applicability. This research, grounded in two hypotheses, proposes and proves a theorem that employs a Bernoulli experiment to demonstrate the high probability of encountering any finite bit string within a sequence of consecutive bits in the decimal part of π. This aligns with findings related to its normality. To support the hypotheses, we present experimental evidence about the equiprobable and independent properties of bits of π, analyzing their distribution, and measuring correlations between bit strings. Additionally, from a cryptographic perspective, we evaluate the chaotic properties of two images generated using bits of π. These properties are evaluated similarly to those of encrypted images, using measures of correlation and entropy, along with two hypothesis tests to confirm the uniform distribution of bits and the absence of periodic patterns. Unlike previous works that solely examine the presence of sequences, this study provides, as a corollary, a formula to calculate an upper bound N. This bound represents the length of the sequence from π required to ensure the location of any n-bit string at least once, with an adjustable probability p that can be set arbitrarily close to one. To validate the formula, we identify sequences of up to $n = 40$ consecutive zeros and ones within the first N bits of π. This work has potential applications in Cryptography that use the number π for random sequence generation, offering insights into the number of bits of π required to ensure good randomness properties.

Keywords: Bernoulli experiment; chaos; entropy; normal numbers; Pi number; upper bound

MSC: 68Q87

Academic Editor: Alexander Dudin

Received: 10 December 2024
Revised: 10 January 2025
Accepted: 17 January 2025
Published: 19 January 2025

Citation: Silva-García, V.M.; Cardona-López, M.A.; Flores-Carapia, R. An Upper Bound for Locating Strings with High Probability Within Consecutive Bits of Pi. *Mathematics* **2025**, *13*, 313. https://doi.org/10.3390/math13020313

1. Introduction

Throughout history, different discoveries have been made concerning the mathematical constant π (pi) [1,2]. For instance, efforts have been made to find rational approximations of π [3], propose methods for calculating the nth digit of π [4], and compute the most accurate approximations of π [5]. In addition, the significance of π extends beyond mere mathematical formulas [6]. It finds applications across diverse fields, including Cryptography. Image encryption algorithms, for instance, leverage π to simulate random numbers by utilizing successive decimals of π [7]. This approach is valuable due to the high level of randomness inherent in the generated sequence, making it well-suited for cryptographic purposes [8]. Additionally, in cryptographic algorithms π has been employed in combination with a secret key to generate permutations [9]. Furthermore, in the realm of chaos

theory, a 2D chaotic map has been devised based on Euler's number and π, demonstrating hyperchaotic behavior in sequence generation [10].

In addition, investigations into the digits of π have revealed patterns, such as the Feynman point, a sequence of six 9s starting from the 762nd decimal place [11]. In fact, in a normal number x in base b (where $b \geq 2$), each of the b^m different strings of length m occurs with equal frequency [12]. This concept is expressed mathematically as $\lim_{n\to\infty}(N(s,n)/n) = b^{-m}$, where $N(s,n)$ denotes the number of occurrences of the string s of length m in the first n digits of x. This idea has led to analyses of the normality of π [13], suggesting that every possible string s eventually appears within the digits of π [14].

The present study demonstrates that any bit string A_n of length n appears with consecutive bits somewhere within A_N, it is the first N bits to the right of the decimal point of π with a probability approaching 1. In essence, it offers a formula to determine the length of a discrete interval of bit positions where a specified string A_n occurs at least once, with an error α as close to zero as desired.

The proposed upper bound contributes to the study of the normality of π, a topic previously explored in other works [13,15]. By establishing such a bound, the assertion that any sequence of bits can be found within π is further reinforced. This is because the bound provides an interval within which any given sequence can be located with a probability arbitrarily close to one.

Regarding the significance of this work in cryptographic applications, it can be particularly relevant for efforts that utilize the digits of π to generate random sequences [9,10]. In such applications, the number of bits of π used is often selected based on the algorithm's requirements, without necessarily considering the randomness properties. The proposed upper bound can provide guidance on the number of bits needed to ensure that all possible sequences of a given length are highly likely to appear within the chosen range.

This approach minimizes the risk of omitting any specific sequence, thereby enhancing the randomness properties of the generated sequences. For instance, if π is used as a source of randomness for a cryptographic application requiring the random selection of n-bit strings, the upper bound proposed in this study specifies the number N of bits that should be considered from π. By using these N bits, it is highly probable that all possible n-bit sequences will be present, making π a reliable source for random string generation.

Also, the present work includes a validation of the proposed formula by computing the upper bound N for various sizes of n, where a string A_n is contained within A_N with an error of $\alpha = 0.01$. Beginning with a string length of $n = 5$, the increments progress in intervals of 5 up to $n = 40$. Additionally, it denotes the position within π where the first bit of A_n lies. While this investigation works on strings comprised solely of consecutive zeros or ones, it is worth noting that other combinations are also feasible.

Furthermore, this study relies on two key hypotheses in the number π: the equiprobable distribution of bits and the independent events of selecting two different strings. The former asserts that when selecting a bit randomly from the bits of the decimal part of π, each bit can be either zero or one, with an equal probability of 0.5. The latter hypothesis deals with the probabilistic experiment of selecting two distinct strings comprised of n and m consecutive bits, respectively, from the decimal part of π. In this case, the probability of selecting the strings A_n and B_m is determined by the product of the probabilities of selecting each individual string: $P(A_n \cap B_m) = P(A_n) \times P(B_m)$.

To validate these hypotheses, two numerical experiments will be presented as evidence in Section 3. Additionally, other studies corroborate these hypotheses. An analysis of the distribution for four trillion hexadecimal digits of π was conducted [15], which later was complemented by several tests [16]. Furthermore, a hypothesis test was performed using

the first 100 million decimal digits of π, indicating that it is 1.86×10^{30} times more likely that π is normal than the contrary [17].

Moreover, to substantiate the two hypotheses, we offer evidence about the chaotic distribution of information within the bits of the number π. For this purpose, we employ methodologies commonly utilized in image encryption studies. Within this domain, key metrics such as entropy and the correlation coefficient serve to assess the chaotic nature of sequences. This research undertakes a comparative analysis between the outcomes derived from applying these metrics to encrypted images against those obtained from images generated by extracting blocks of random bits from the decimal part of π.

Regarding the distribution of this research, this section contains the introduction and a comprehensive overview of the current state of the field. Subsequently, the theoretical framework utilized in this study is expounded upon in Section 2. Section 3 elucidates the experimental results, validating the two hypotheses posited in this investigation. In Section 4, the theorem regarding finite sequences of consecutive bits of π is proven, along with establishing the upper bound for locating the strings within π. Section 5 offers the analysis and discussion of the obtained results. Lastly, Section 6 encapsulates the conclusions drawn from this research.

2. Materials and Methods

This section begins by describing the Bernoulli model utilized for analyzing the bits of the number π. Subsequently, encryption tools commonly employed to measure chaotic properties of information are applied to evaluate two images constructed from the bits of π. These tools include correlation analysis, entropy measurement, and two hypothesis tests: the Discrete Fourier Transform and the Goodness-of-Fit test.

2.1. Bernoulli Model

In a Bernoulli test, there are only two possible outcomes: 1 and 0, where one is defined as success and the other as failure [18]. In this context, the probability of obtaining the value 1 is denoted by p, and q represents the probability of obtaining the value 0, satisfying $p + q = 1$. When there are reproduced n different trials, which are also independent of each other, the Binomial distribution is constructed [19]. In this context, the mathematical model that describes this process is in Equation (1), where x represents the number of successes after n trials for $0 \leq x \leq n$.

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \tag{1}$$

On the other hand, an unbiased estimator $\bar{X}$ for the parameter p is expressed as $\bar{X} = \sum_{i=1}^{n} x_i / n$.

2.2. Information Entropy

Information entropy H is a measure frequently utilized in Cryptography [20–22], this parameter offers insights into the level of chaos within a block of bits. It is defined in Equation (2). In this research, it serves as a key metric to apply it to consecutive bits of π. However, it is essential to note that while information entropy provides essential insights, it alone does not suffice to conclusively determine the chaotic nature of the data. Hence, the assessment of chaotic properties involves the consideration of various parameters to reinforce the results.

$$H(x) = -\sum_{x \in X} P(x) \log_2 P(x) \tag{2}$$

The parameter will be utilized to assess the chaotic level of pixels within an image generated using a block of bits from the number π. These images, depicted in color

in Figures 1 and 2, undergo segmentation into blocks of 8 bits per color for entropy computation. It is noted that information is deemed to be chaotically distributed when the entropy value approaches 8 [23].

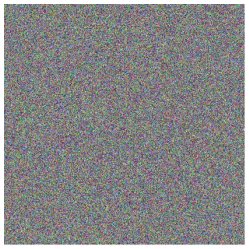

Figure 1. Image of dimensions 512 × 512 pixels, generated using 786,432 blocks of 8 bits randomly selected from the decimal part of π. Each pixel comprises 24 bits (three blocks), representing the red, green, and blue color channels.

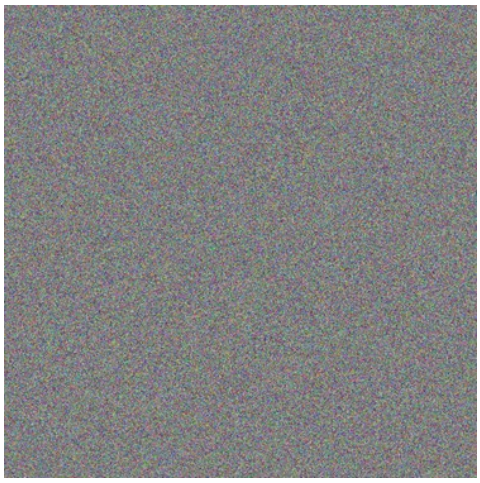

Figure 2. Image of dimensions 1024 × 1024 pixels, generated using 3,145,728 blocks of 8 bits randomly selected from the decimal part of π. Each pixel comprises 24 bits (three blocks), representing the red, green, and blue color channels.

2.3. Correlation Coefficient

Another tool utilized to assess the absence of a linear relationship between the bits within an image is the correlation coefficient r_d [24]. For the computation of this parameter, w pixels are haphazardly chosen from an image. Each pixel has three possible neighborhoods [25], having adjacent pixels in horizontal, vertical, and diagonal directions d. In addition, a pixel is represented with the colors c: red, green, and blue, and a byte with 256 different intensity levels.

Subsequently, the correlation for each direction d and color c is evaluated. Let the value of the ith randomly chosen pixel be denoted as $x_{i,d,c}$, where the subscripts d and c represent the basic color direction and type, respectively. Similarly, $y_{i,d,c}$ denotes the value of the adjacent pixel in direction d and for color c. The formula to calculate it is provided in Equation (3).

$$r_d = \frac{\frac{1}{w}\left(\sum_{i=1}^{w}(x_{i,c,d} - \overline{x}_{c,d})(y_{i,c,d} - \overline{y}_{c,d})\right)}{\sqrt{\frac{1}{w^2}\left(\sum_{i=1}^{w}(x_{i,c,d} - \overline{x}_{c,d})^2\right)\left(\sum_{i=1}^{w}(y_{i,c,d} - \overline{y}_{c,d})^2\right)}} \tag{3}$$

where the means of each variable are as follows $\overline{x}_{c,d} = \sum_{i=1}^{w} x_{i,c}/w$ and $\overline{y}_{c,d} = \sum_{i=1}^{w} y_{i,c}/w$.

2.4. Discrete Fourier Transform

The Discrete Fourier Transform (DFT) is used as a statistical hypothesis test that serves to assess the degree of disorder in which the bits of a string appear [26], particularly in our case, a sequence of consecutive bits from π forming an image. Specifically, this parameter evaluates the presence of periodic features within a given sequence of m bits $\{\delta_1\delta_2 \ldots \delta_m\}$. In addition, the null hypothesis posits that the information in a randomly selected block of bits, sourced from the decimal part of π, follows a random distribution. Within the framework of hypothesis testing the test statistic is d, presented in Equation (4) using the length m of the sequence and two constants M_0, M_1.

$$d = \frac{M_1 - M_0}{\sqrt{\frac{m(0.95)(0.05)}{4}}} \tag{4}$$

For computing the constant M_0, it is defined below in Equation (5).

$$M_0 = \frac{(0.95) \times m}{0.05} \tag{5}$$

Regarding M_1, it requires the bound l, it is defined in Equation (6) and a sequence $Y = y_1y_2, \ldots, y_m$. The computation of Y is using the bits δ_k in the following manner: $y_k = 2\delta_k - 1$, for $1 \leq k \leq m$.

$$l = \sqrt{m\mathrm{Ln}\frac{1}{0.05}} \tag{6}$$

Additionally, it is necessary to compute $m/2 - 1$ complex functions f_j, for $1 \leq j \leq m/2 - 1$ according to Equation (7), where $i = \sqrt{-1}$ is the complex unit. In this way, m is the number of pixels expressed in bits and even. Subsequently, the module $\| f_j \|$ is computed for each function. If $\| f_j \| < l$, 1 is added to M_1; otherwise, the value of M_1 does not change. It is important to note that the initial value of M_1 is zero. After comparing all the moduli, the final value of M_1 is obtained, and therefore, d can be computed.

$$f_j = \sum_{k=1}^{m} y_k e^{\frac{2\pi(i)(k-1)j}{n}} \tag{7}$$

In this context, the decision parameter of the hypothesis test is the p-value and is expressed in Equation (8). The rejection of the null hypothesis of randomness occurs when p-value < 0.01; otherwise, it is accepted.

$$p\text{-value} = \mathrm{erfc}\frac{|\, d \,|}{\sqrt{2}} \tag{8}$$

Finally, the erfc function uses the cumulative function Φ of the standard normal distribution as Equation (9) presents it.

$$\mathrm{erfc}\frac{|\, d \,|}{\sqrt{2}} = 2(1 - \Phi(|\, d \,|)) \tag{9}$$

2.5. Goodness-of-Fit Test

This hypothesis test has a null hypothesis positing that the information distribution is uniform [27], contrasting with the alternative hypothesis suggesting otherwise. The aim of this tool is to assess the extent to which the information in the histograms conforms to a uniform distribution. This assessment is conducted for each color channel red, green, and blue, given that every pixel in the images we employ is depicted by these colors. Each of

them is defined by a byte, allowing for 256 possible decimal values. Within this framework, the test statistic is expressed in Equation (10).

$$\chi^2 = \sum_{i=1}^{256} \frac{(o_i - \exp)^2}{\exp} \tag{10}$$

Equation (10) must be applied to each color channel individually. For a specific color channel, o_i represents the number of occurrences of intensity level i across the entire image, while exp denotes the expected frequency of each intensity level to create a uniform distribution. The expected frequency can be calculated as the total number of pixels divided by 256, which corresponds to the total number of possible intensity values in an 8-bit representation.

According to the Central Limit Theorem, the distribution of the χ^2 statistic tends towards a normal distribution [28] of mean $\mu = 255$ and standard deviation $\sigma = 22.5$. Moreover, a significance level $\alpha = 0.01$ rejects the null hypothesis if $\chi^2 \geq 308$; otherwise, it is accepted.

In this work, the test is used to analyze the distribution of 8-bit blocks within the pixels. Specifically, it assesses whether the bits of the number π, grouped into 8-bit blocks, exhibit a uniform distribution. This means verifying whether all 256 possible values that can be represented with 8 bits occur with approximately equal frequency.

3. Two Hypothesis Derived from Pi

In this section, the analysis is focused on the bits of the decimal expansion of π. First, the distribution of bits equal to zero and one is examined. Next, the correlation between two strings of bits is analyzed, followed by an evaluation of the chaotic properties of two images constructed from the bits of π. This comprehensive analysis provides further evidence supporting two hypotheses.

3.1. Equiprobable Property of Bits

This subsection will provide evidence that the bits of the number π present an equiprobable distribution. In other words, if a bit x_i is randomly selected from these bits, the probability of it being either 0 or 1 is $1/2$. To verify this, eight strings of distinct length were extracted, all starting from the first bit of the decimal part of π and consisting of n consecutive bits. Furthermore, the probability was assessed using the estimator mean $\bar{X} = (\sum_{i=0}^{n-1} x_i)/n$. Table 1 presents the percentages of zeros and ones for each string. Pertinent comments will be provided in Section 5.

Table 1. Percentage distribution of zeros and ones in strings of varying lengths extracted from the bits of π. The results support the hypothesis that each bit of π is equiprobable, with a 50% chance of being either zero or one.

Bit String Length n	Percentage of Zero Bits	Percentage of One Bits
2^5	49.999583	50.000417
2^{10}	51.074219	48.925781
2^{15}	49.935913	50.064087
2^{20}	50.023270	49.976730
2^{25}	49.990329	50.009671
2^{30}	49.999331	50.000669
2^{35}	50.000190	49.999810
2^{40}	50.000034	49.999966
2^{45}	49.999905	50.000095

As a result of the observations presented in Table 1, the evidence suggests a hypothesis regarding the random selection of a bit from π, proposing that it can be either zero or one with an equal probability of 50%. This is stated in Hypothesis 1. However, the evidence in Table 1 is based on our analysis of up to 2^{45} bits of π. In this way, this result aligns with other studies examining the properties of π. For instance, Bailey et al. analyzed the first trillion hexadecimal digits of π, reporting that each hexadecimal digit from 0 to f appears approximately the expected number of 62,500,000,000 times in a sample of 1,000,000,000,000 digits [15]. Although their study focuses on the normality of π across its first four trillion hexadecimal digits, it also incorporates complementary tests on the randomness of π's digits [16], yielding results consistent with our results.

Hypothesis 1. *The probability of x_i being 0 or 1 is $P(x_i = 0) = P(x_i = 1) = 1/2$. In other words, when selecting a bit randomly from the bits to the right of π, it can have a value of either zero or one, each with a probability of 0.5.*

3.2. Independence Property of Bit Strings

To provide evidence for this property, we first define the strings that will be used in a subsequent experiment. Let $A_{n,0} = \{x_0, \ldots, x_{n-1}\}$ represent a string of n bits, where x_0 denotes the first bit, located immediately to the right of the decimal point in the number π. More generally, a string of length n starting at the i-th bit to the right of the decimal point in π is represented as $A_{n,i} = \{x_i, \ldots, x_{i+n-1}\}$. For instance, following this notation, we have $A_{n,1} = \{x_1, \ldots, x_{n-1}, x_n\}$, and the sequence continues to generate $A_{n,2}$, $A_{n,3}$, up to $A_{n,10}$.

Using these strings, an experiment is conducted to measure the correlation coefficient between pairs of strings. Specifically, the correlations are computed between $A_{n,0}$ and $A_{n,1}$, $A_{n,0}$ and $A_{n,2}$, $A_{n,0}$ and $A_{n,3}$, and so on. This test is performed for four different string lengths (n), and the results are summarized in Table 2.

Table 2. Correlation coefficients between pairs of string $A_{n,0}$ and shifted strings $A_{n,i}$, for four lengths n, derived from π's digits. Values near 0 suggest independence, with correlations decreasing as string length n increases.

Starting Bit (i) of $A_{n,i}$	**Bit String Length (n) of $A_{n,0}$, and $A_{n,i}$**			
	100	**1000**	**1,000,000**	**100,000,000**
1	−0.04166	−0.00200	0.00080	0.000025
2	−0.08333	0.00459	−0.00121	0.00012
3	−0.08333	0.00799	−0.00052	0.000072
4	0.10790	0.00439	0.00081	−0.000055
5	0.02490	−0.00800	−0.00050	−0.000075
6	−0.09960	−0.00320	0.00048	0.000076
7	−0.11580	0.00119	0.00100	0.000100
8	0.03298	−0.01040	0.00021	−0.000100
9	−0.04947	0.00219	0.00062	0.000059
10	0.11544	−0.00200	0.00184	0.000088

A correlation value close to 0, as the values reported in Table 2, between two strings ($A_{n,0}$ and $A_{n,i}$) derived from the digits of π, indicates the absence of a relationship between them and, therefore, suggests independence. Additionally, as the string length n increases, the linear correlation approaches zero, further reinforcing their independence.

To provide additional support for the hypothesis of independence between different strings from π, in addition to the correlation coefficient analysis, other independence tests commonly applied to encrypted images were reproduced. For this purpose, two color images were created by selecting blocks of 8 consecutive bits randomly from the decimal

expansion of π. The first image, shown in Figure 1, has dimensions of 512×512 pixels, while the second, shown in Figure 2, has dimensions of 1024×1024 pixels.

Regarding the construction of these images, as they are color images, each pixel consists of 24 bits (8 bits per color channel: red, green, and blue). Thus, the creation of a single pixel requires three blocks of 8 bits. For this work, the 8-bit blocks were selected randomly from the digits of π. To generate the $512 \times 512 = 262{,}144$ pixels in Figure 1, a total of 786,432 bytes were extracted from π (calculated as $3 \times 262{,}144$). Similarly, for Figure 2, which contains $1024 \times 1024 = 1{,}048{,}576$ pixels, 3,145,728 bytes were utilized.

After creating the images, we evaluated them using four parameters similar to those employed in analyzing encrypted images [29,30]. These methods assess the degree of chaos in the information and the relationships between pixels. In this case, they were applied to Figures 1 and 2 to evaluate the chaotic nature of the corresponding bits (represented by pixels and derived from the digits of π) in both figures.

In this case, the tests were applied to Figures 1 and 2 to evaluate the chaotic nature of the corresponding bits (represented by pixels and derived from the digits of π) in both images. Additionally, conducting this analysis in such a manner implicitly involves analyzing the bits of π in 8-bit blocks rather than solely in binary form, as it was shown in Tables 1 and 2. This is feasible because the evaluation of the images is performed per color channel for each pixel, which, as previously explained, comprises 8 bits.

The parameters analyzed include entropy and the correlation coefficient between adjacent pixels in three directions (horizontal, vertical, and diagonal) for each primary color. The results of these evaluations are summarized in Table 3.

From the data in Table 3, it is shown that the entropy values are close to the ideal value of 8.0, indicating that the 256 possible values representing a pixel's color channel occur with nearly uniform frequency. This observation supports the equiprobable distribution of the bits of π. Additionally, the correlation coefficients are near the ideal value of 0.0, suggesting an absence of relationships between adjacent pixels. In this context, these results imply a lack of correlation between the bits of π from which the pixels are generated, across all three analyzed directions.

Table 3. Entropy and correlation coefficients for adjacent pixels in three directions (horizontal, vertical, and diagonal) for each primary color, derived from π's bits in Figures 1 and 2. Entropy values near 8.0 indicate uniform distribution, while near-zero correlations suggest independence between adjacent pixels.

Measure	Figure	Red	Green	Blue
Horizontal correlation	Figure 1	0.00462	−0.00067	−0.00979
	Figure 2	0.00102	−0.00151	−0.00396
Vertical correlation	Figure 1	0.00290	−0.00189	0.00287
	Figure 2	−0.00152	0.00049	−0.00103
Diagonal correlation	Figure 1	−0.00687	0.00195	−0.00292
	Figure 2	0.00137	0.00548	−0.00336
Entropy	Figure 1	7.99920	7.99939	7.99931
	Figure 2	7.99981	7.99982	7.99982

Additionally, two hypothesis tests were performed on Figures 1 and 2: the Discrete Fourier Transform (DFT) test and the goodness-of-fit test [31]. The results, presented in Table 4, use the symbol ✓ to denote acceptance of the null hypothesis. For the DFT test, the acceptance of the null hypothesis indicates the absence of periodic features within the sequence, which in this context pertains to the pixels but can also be interpreted as the bits of π from which the pixels are derived.

Table 4. Results of the Discrete Fourier Transform (DFT) test and the goodness-of-fit test conducted on Figures 1 and 2. The symbol ✓ indicates acceptance of the null hypothesis, signifying that the distribution of information (pixels, derived from the bits of π) is random.

Test	Figure	Red	Green	Blue
χ^2	Figure 1	288.7 < 308 ✓	218.3 < 308 ✓	250.2 < 308 ✓
	Figure 2	262.6 < 308 ✓	247.9 < 308 ✓	262.0 < 308 ✓
p-value of DFT	Figure 1	0.789 > 0.01 ✓	0.501 > 0.01 ✓	0.292 > 0.01 ✓
	Figure 2	0.782 > 0.01 ✓	0.502 > 0.01 ✓	0.294 > 0.01 ✓

The acceptance of the null hypothesis in the goodness-of-fit test confirms that the pixel intensity distribution closely resembles a uniform distribution. This result is significant, as it suggests that the digits of π, when expressed as 8-bit numbers (within a 256-numerical system), each of them occurs approximately with the same frequency. The conclusions derived from entropy and correlation are further validated by these two hypothesis tests, as the acceptance of the null hypothesis, in summary, signifies a random distribution of the information. In this case, the information pertains to the pixels but is fundamentally tied to the bits of π.

With these observations in mind, and based on the correlation between the pairs of strings analyzed in this section, the second hypothesis is stated in Hypothesis 2. To ensure clarity, the relevant elements referenced within the hypothesis are defined below.

- Let x_i, y_i represent a bit to the right of the decimal point of the number π.
- Let $A_n = \{x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+n-1}\}$ denote a string of n bits.
- Let $B_m = \{y_i, y_{i+1}, y_{i+2}, \ldots, y_{i+m-1}\}$ a different bit string with m bits. It is assumed that there is at least a bit x_j at position j in A_n that does not exist at the same position in B_m, and vice-versa.

Hypothesis 2. *In the probabilistic experiment of selecting two strings, each comprising n and m consecutive bits, respectively, from the decimal part of the number π, the probability of selecting the strings A_n and B_m is calculated as the product of the probability of selecting the string A_n and the probability of selecting the string B_m. In other words, $P(A_n \cap B_m) = P(A_n) \times P(B_m)$.*

4. The Upper Bound

This section outlines the methodology to establish an upper bound on the number of bits of π's decimal expansion required to locate any given string at least once with a probability p.

4.1. The Probability of Selecting a String A_n

From the presented information in Hypothesis 1 and 2, the following Lemma is presented.

Lemma 1. *In the probabilistic experiment of selecting a string of n consecutive bits from the decimal part of the number π, the probability of selecting the string A_n is equal to $(1/2)^n$, expressed as $P(A_n) = (1/2)^n$.*

Proof of Lemma 1. The proof of this lemma shall proceed through the method of mathematical induction.

1. Base case: The base case is considered when $n = 1$, representing the probability of selecting the 1-bit string A_1, which consists solely of the bit x_i. In other words, $P(A_1) = P(x_i)$. According to the first hypothesis, it follows that

$$P(x_i) = (1/2)^1 = P(A_1).$$

This result aligns with the statement of the Lemma, confirming consistency.

2. Inductive hypothesis: Assume that the Lemma holds true for $n = k$. Under this assumption, the inductive hypothesis states that the probability of selecting the k-bit string A_k is given by $P(A_k) = (1/2)^k$.
3. Inductive step: Prove that the formula also holds for $k + 1$. We then analyze the scenario wherein a string of $k + 1$ bits is selected, it is $n = k + 1$. Consequently, the probability $P(A_{k+1})$ of selecting the string A_{k+1} can be expressed as the probability $P(A_k \cap x_{i+k})$ of selecting the k-bit string A_k and the 1-bit string $\{x_{i+k}\}$.
   Leveraging the Hypothesis 2, we deduce

$$P(A_k \cap x_{i+k}) = P(A_k) \times P(x_{i+k}).$$

   Subsequently, utilizing the inductive hypothesis $P(A_k) = (1/2)^k$ and the Hypothesis 1 indicating $P(x_i) = 1/2$, it follows that

$$P(A_k) \times P(x_{i+k}) = (1/2)^k \times 1/2.$$

   Consequently,

$$P(A_{k+1}) = P(A_k) \times P(x_{i+k}) = (1/2)^{k+1}.$$

□

4.2. The Probability of Finding A_n

Before presenting the theorem, another string will be presented. Consider A_N as a string of $N > n$ bits, representing the first N consecutive bits to the right of the decimal point within the number π. Here, the initial bit of A_N corresponds to the first bit within the decimal part of π.

Theorem 1. *As the length N of the string A_N approaches infinity, the probability of the n-bit string A_n appearing at least once as a contiguous bit string within A_N tends to 1.*

Proof of Theorem 1. Next, we demonstrate the probability that A_n appears at least once as a sequence of consecutive bits within the string A_N of length N. This involves analyzing the probability that A_n appears $y \geq 0$ times within A_N.

To achieve this, we present the following probabilistic experiment: each group of n consecutive bits from A_N is compared to the string A_n. If they match, it is considered a success. This setup constitutes a Bernoulli experiment, as each trial has exactly two possible outcomes: success or failure. The details of the experiment are outlined below.

1. First trial: The initial n consecutive bits of the string A_N are selected and compared to determine if they match the string A_n. A successful outcome occurs when the selected sequence matches the A_n string, while failure arises if there is no match. Referring to Lemma 1, the probability of success is $P(A_n) = (1/2)^n$. In addition, it is possible to form $N - n + 1$ strings of n consecutive bits in one of length N, which implies that the number of trials is $N - n + 1$.
2. Second trial: Another string of n bits is obtained from A_N starting from bit position 1 and extending to bit n. Following the same reasoning, the probability of this string being equal to A_n is $P(A_n)$. Additionally, since the string of n bits from trial 1 is different from that of trial 2, it is independent as per the second hypothesis.
3. i-th trial: In general, for the i-th trial, the string to be compared with A_n starts from bit number $i - 1$ of A_N, encompassing the consecutive n bits.

Under this experiment, the probability of the string A_n appearing y times as consecutive bits within the string A_M is expressed in Equation (11).

$$P(Y = y) = \binom{N+1-n}{y}((1/2)^n)^y(1-(1/2)^n)^{(N+1-n)-y} \tag{11}$$

Then, the probability $P(Y \geq 1)$ of A_n appearing at least once as a consecutive bit string in A_N can be expressed through its complement, where it does not appear at all:

$$P(Y \geq 1) = 1 - P(Y = 0).$$

By substituting $y = 0$ into Equation (11), we obtain Equation (12), which represents the probability that the string A_n appears at least once within A_N.

$$P(Y \geq 1) = 1 - (1-(1/2)^n)^{N+1-n}. \tag{12}$$

Subsequently, by applying the limit as $N \to \infty$ to both sides of Equation (12), we have

$$\lim_{N\to\infty} P(Y \geq 1) = \lim_{N\to\infty}(1-(1-(1/2)^n)^{N+1-n}),$$

which can be simplified to

$$\lim_{N\to\infty} P(Y \geq 1) = \lim_{N\to\infty}(1) - \lim_{N\to\infty}(1-(1/2)^n)^{N+1-n}.$$

It is important to note that $(1/2)^n > 0$ for any arbitrary but fixed $n > 0$. Therefore, $-(1/2)^n < 0$ and $0 < 1-(1/2)^n < 1$. Consequently, as $N \to \infty$, $(1-(1/2)^n)^{N+1-n} \to 0$. In other words, $\lim_{N\to\infty}(1-(1/2)^n)^{N+1-n} = 0$. In conclusion, given that the first term is a constant ($\lim_{N\to\infty}(1) = 1$) and the limit obtained earlier corresponds to the second term, it follows that

$$\lim_{N\to\infty} P(Y \geq 1) = 1 - 0 = 1.$$

□

From this, we conclude that, on the long path to infinity, it is highly probable that the string A_n appears at least once, within A_N.

4.3. An Upper Bound N for Finding A_n

Continuing this discussion, another problem to address is determining the length N of a bit string, starting from the zero bit position of the decimal part of π, such that a given string of length n is contained at least once within A_N with consecutive bits and a probability of p. As a result of this consideration, the following corollary is formulated:

Corollary 1. *The length N at which A_n appears at least once within A_N as a string of consecutive bits can be determined using Equation (13) in terms of the probability p of its occurrence and the length n.*

$$N = n - 1 + \left\lceil \frac{\ln(1-p)}{\ln(1-(1/2)^n)} \right\rceil \tag{13}$$

Proof of Corollary 1. In this proof, we utilize the results of Theorem 1. Accordingly, the probability of A_n appearing at least once as a consecutive bit string in A_N, ($P(Y \geq 1)$, is given by Equation (12). Let $P(Y \geq 1) = p$, then Equation (12) is written as,

$$p = 1 - (1-(1/2)^n)^{N+1-n}.$$

By rearranging terms, we obtain

$$(1-(1/2)^n)^{N+1-n} = 1 - p.$$

Upon applying the natural logarithm on both sides and its properties

$$(N+1-n)\ln(1-(1/2)^n) = \ln(1-p).$$

Subsequently, by solving for N, we arrive at the result shown in Equation (13),

$$N = n - 1 + \frac{\ln(1-p)}{\ln(1-(1/2)^n)}$$

□

On the other hand, the probability p can be made as close to one as desired, minimizing the error to find any desired length. In this study, a probability of $p = 0.99$ is employed, resulting in an error of $\alpha = 0.01$. Furthermore, it is important to clarify that in this investigation, N is rounded using the ceiling function, denoted by $\lceil \cdot \rceil$, to return the smallest integer greater than or equal to the expression in Equation (13), i.e., the integer part plus one. Subsequently, the results of finding strings for different values of N will be presented in the next section.

5. Results Analysis and Discussion

This section commences with the results derived from the experimental properties. Firstly, Table 1 illustrates that as the length n of the bit string increases, the proportion of zeros and ones tends toward 50%. Secondly, there is a fluctuation in the percentages of zeros and ones, with the percentage of zeros occasionally exceeding 50% and at other times falling below this percentage. This observation forms the basis for the first hypothesis proposed in this research. Additionally, this result is consistent with the concept of normal numbers [32]. To further support the second hypothesis, Table 2 is presented. In this analysis, four different bit lengths are considered to generate strings starting from various bit positions i within the decimal part of the number π. It can be observed that as the length of the bit strings increases, the correlation between pairs of strings tends towards zero. This indicates a lack of a linear relationship between them, providing evidence in favor of the second hypothesis.

Additionally, two color images, presented in Figures 1 and 2, are constructed using randomly selected blocks of bits from the decimal part of the number π. The first image comprises 512×512 pixels, while the second one consists of 1024×1024 pixels. Subsequent evaluations are conducted to assess the level of chaos exhibited by the bits in both images from a cryptographic perspective. This evaluation is based on parameters such as entropy, correlation, goodness-of-fit test using the χ^2 distribution, and the discrete Fourier transform. It is observed that there is no discernible relationship between the bits in both images, as evidenced by Tables 1–4.

On the other hand, significant research in image encryption shows lower entropy levels of encrypted images than those observed in the images generated from the bits of π, suggesting a comparatively lower degree of pixel randomness. For instance, the entropy values of Figures 1 and 2 range from 7.99920 to 7.99982. In contrast, several encryption proposals report lower entropy values: the work in [33] achieves entropy values between 7.9931 and 7.9949, while another reports values ranging from 7.9968 to 7.9975 [34]. Similarly, a third study presents a range of 7.9972 to 7.9994 [35]. Other researches achieve entropy levels closer to those of the π-based images, such as entropy values between 7.9969 and 7.9998 [36], and a range from 7.9992 to 7.9998 [37]. Consequently, it can be observed that the distribution of information within the blocks of the number π exhibits chaotic properties, in some cases surpassing those observed in encrypted images.

Concerning the theorem presented, it is observed that as the string length N increases, the probability of locating the string A_n within consecutive bits of A_N approaches 1. However, this increase in N entails the need to search through more bits, which comes at a computational cost. Additionally, due to the probabilistic nature of the process, there remains a risk that certain strings may fall outside the range of consideration.

After establishing the theorem, a corollary was formulated to determine the length N of the sequence A_N required to encompass, with consecutive bits, any string A_n with a probability p. To validate the formula, 16 distinct strings were located in π. These were composed entirely of zeros or ones, across eight different sizes. The formula's predicted upper bound, computed with a probability $p = 0.99$, was used as a reference.

Table 5 summarizes the results of this validation. The first column specifies the length n of the target string, while the second column presents the calculated upper bound N. The third and fourth columns indicate the starting bit positions of the identified strings of zeros and ones, respectively, within the decimal part of π. The results confirm that all identified strings are located within the proposed upper bound, except for the string of $n = 25$ consecutive zeros, which was found at a position exceeding the predicted upper bound N. This discrepancy aligns with the 1% error margin inherent in the computation of Equation (13) with $p = 0.99$.

Table 5. Validation results for locating strings of n consecutive zeros and ones within the decimal part of π, compared against the predicted upper bound N with a 99% probability.

Length n	Length N	Starting Bit-Position of A_n of Consecutive Zeros	Ones
5	150	95	10
10	4723	901	644
15	150,914	11,790	58,275
20	4,828,882	726,843	1,962,900
25	154,523,896	171,498,579	47,536,570
30	4,944,763,863	1,407,238,213	207,861,697
35	158,232,442,734	21,774,349,073	61,906,790,708
40	5,063,438,167,262	1,584,920,456,449	1,748,147,295,589

In this study, string searching was performed using a finite automaton search algorithm with n states, where n represents the length of the string being searched within a sequence of N bits of π. A finite automaton of these characteristics has a computational complexity of $O(N)$ [38] for locating an n-bit string within an N-bit sequence of π. Regarding the storing, to search for sequences of 40 consecutive bits, the upper bound suggests examining 5,063,438,167,262 bits, which entails significant storage requirements. This highlights the importance of analyzing the type of storage used, whether a hard drive or a solid-state drive, particularly for cryptographic applications.

Finally, although the present study focuses on the number π, the methodology developed here can be extended to analyze string searches in other numbers, such as transcendental numbers. Future research could expand this analysis to explore how the results vary with different numbers, identify the unique strengths of each number, and evaluate their potential applications. Such investigations could provide valuable options for utilizing these numbers in fields like Cryptography.

6. Conclusions

In this work, a theorem was presented and demonstrated, showing that any finite-length string A_n has a probability arbitrarily close to 1 of appearing somewhere with consecutive bits after the decimal point of the number π. From this theorem, a corollary

has been derived, demonstrating how to calculate the length of the chain A_N in π such that, with a probability p, the chain A_n appears at least once within A_N. In this research, $p = 0.99$ has been chosen, though it can approach 1 arbitrarily closely. The proof of this theorem relies on two hypotheses: firstly, when selecting a bit x_i from the right side of the decimal point, $P(x_i = 0) = P(x_i = 1) = 1/2$. Secondly, it establishes that when two different bit strings of lengths n and m are chosen from the decimal part of π, the probability of selecting A_n and B_m such that $A_n \neq B_m$ is independent; that is, $P(A_n \cap B_m) = P(A_n) \times P(B_m)$. Experimentation has been conducted to provide evidence supporting these hypotheses, as shown in Tables 1–4. Furthermore, the results derived from the images of π in Figures 1 and 2, when subjected to entropy and correlation parameters, resemble those of encrypted images known for their chaotic information. Finally, the upper bound N provides information about the number of bits of π to use for generating random sequences effectively.

Author Contributions: Conceptualization V.M.S.-G.; methodology M.A.C.-L.; formal analysis V.M.S.-G. and M.A.C.-L.; investigation M.A.C.-L., visualization M.A.C.-L., writing—review and editing M.A.C.-L. and R.F.-C.; data curation R.F.-C.; software R.F.-C.; validation R.F.-C.; writing—original draft preparation V.M.S.-G.; resources V.M.S.-G. and R.F.-C.; supervision V.M.S.-G.; project administration V.M.S.-G. and R.F.-C.; funding acquisition V.M.S.-G. and R.F.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded in part by the economic support program of the Comisión de Operación y Fomento de Actividades Académicas (COFAA) and the Secretaría de Investigación y Posgrado (SIP) of the Instituto Politécnico Nacional under grants SIP-20241356 and SIP-20241716.

Data Availability Statement: The data of this study are available from the corresponding author upon request and the digits of π were downloaded from: https://storage.googleapis.com/pi100t/index.html (accessed on 13 March 2024).

Acknowledgments: The authors would like to thank the Instituto Politécnico Nacional of México (Secretaría Académica, Comisión de Operación y Fomento de Actividades Académicas COFAA, SIP, and CIDETEC), and the CONAHCyT (SNI) for their support in the development of this work.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviation

The following abbreviation is used in this manuscript:

DFT Discrete Fourier Transform

References

1. Agarwal, R.P.; Agarwal, H.; Sen, S.K. Birth, growth and computation of pi to ten trillion digits. *Adv. Differ. Equ.* **2013**, *2013*, 100. [CrossRef]
2. Bailey, D.H.; Borwein, J.M. *Pi: The Next Generation: A Sourcebook on the Recent History of Pi and Its Computation*, 1st ed.; Springer International Publishing: Cham, Switzerland, 2016; Volume 5, pp. 1–501. [CrossRef]
3. Balasubramanian, K.; Davidson, E.R. Rational approximations to pie: Transcendental π and Euler's Constant *e*. *J. Math. Chem.* **2023**, *61*, 1471–1476. [CrossRef]
4. Bailey, D.; Borwein, P.; Plouffe, S. On the rapid computation of various polylogarithmic constants. *Math. Comput.* **1997**, *66*, 903–913. [CrossRef]
5. Frey, B.; Van Schie, A.; Zacheo, G. Pi-Experience—Making 62.8 Trillion Digits Come Alive. In Proceedings of the 15th International Symposium on Visual Information Communication and Interaction (VINCI), Chur, Switzerland, 16–18 August 2022; pp. 1–4. [CrossRef]

6. Bokari, N. *Piece of Pi: Wit-Sharpening, Brain-Bruising, Number-Crunching Activities With Pi*, 1st ed.; Taylor & Francis: New York, NY, USA, 2023; pp. 1–3. [CrossRef]
7. Wang, H.; Hsu, C.; Harn, L. A Lightweight and Robust Stream Cipher Based on PI for Intelligent Transportation Systems. *Wirel. Pers. Commun.* **2023**, *130*, 1661–1675. [CrossRef]
8. Mengdi, Z.; Xiaojuan, Z.; Yayun, Z.; Siwei, M. Overview of randomness test on cryptographic algorithms. *J. Phys. Conf. Ser.* **2021**, *1861*, 012009. [CrossRef]
9. Silva-García, V.M.; Ramírez-González, M.D.; Flores-Carapia, R.; Vega-Alvarado, E.; Escobar-Rodríguez, E. A novel method for image encryption based on chaos and transcendental numbers. *IEEE Access* **2019**, *7*, 163729–163739. [CrossRef]
10. Erkan, U.; Toktas, A.; Toktas, F.; Alenezi, F. 2D eπ-map for image encryption. *Inf. Sci.* **2022**, *589*, 770–789. [CrossRef]
11. Richards, N. *Questions in Dataviz: A Design-Driven Process for Data Visualization*, 1st ed.; AK Peters/CRC Press: Boca Raton, FL, USA, 2022; pp. 286–288. [CrossRef]
12. Émile Borel, M. Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Matem. Palermo* **1909**, *27*, 247–271. [CrossRef]
13. Wagon, S. Is π normal? *Math. Intell.* **1985**, *7*, 65–67. [CrossRef]
14. Lepore, A.; Palumbo, B.; Poggi, J.M. *Interpretability for Industry 4.0: Statistical and Machine Learning Approaches*, 1st ed.; Springer Nature: Cham, Switzerland, 2022; pp. 5–6. [CrossRef]
15. Bailey, D.H.; Borwein, J.M.; Calude, C.S.; Dinneen, M.J.; Dumitrescu, M.; Yee, A. An empirical approach to the normality of π. *Exp. Math.* **2012**, *21*, 375–384. [CrossRef]
16. Bailey, D.H.; Borwein, J.M.; Brent, R.P.; Reisi, M. Reproducibility in computational science: A case study: Randomness of the digits of Pi. *Exp. Math.* **2017**, *26*, 298–305. [CrossRef]
17. Gronau, Q.F.; Wagenmakers, E.J. Bayesian evidence accumulation in experimental mathematics: A case study of four irrational numbers. *Exp. Math.* **2018**, *27*, 277–286. [CrossRef]
18. Miklas-Kalczynska, M.; Drezner, T.D.; Drezner, Z. Experimenting with the generalized binomial distribution. *Commun. Stat. Case Stud. Anal. Appl.* **2024**, *10*, 27–46. [CrossRef]
19. Shapiro, S.S.; Zahedi, H. Bernoulli trials and discrete distributions. *J. Qual. Technol.* **1990**, *22*, 193–205. [CrossRef]
20. Li, C.; Lin, D.; Feng, B.; Lü, J.; Hao, F. Cryptanalysis of a chaotic image encryption algorithm based on information entropy. *IEEE Access* **2018**, *6*, 75834–75842. [CrossRef]
21. Zolfaghari, B.; Bibak, K.; Koshiba, T. The odyssey of entropy: Cryptography. *Entropy* **2022**, *24*, 266. [CrossRef] [PubMed]
22. Kumar, V.; Pathak, V.; Badal, N.; Pandey, P.S.; Mishra, R.; Gupta, S.K. Complex entropy based encryption and decryption technique for securing medical images. *Multimed. Tools Appl.* **2022**, *81*, 37441–37459. [CrossRef] [PubMed]
23. Wu, Y.; Zhou, Y.; Saveriades, G.; Agaian, S.; Noonan, J.P.; Natarajan, P. Local Shannon entropy measure with statistical tests for image randomness. *Inf. Sci.* **2013**, *222*, 323–342. [CrossRef]
24. Alghamdi, Y.; Munir, A. Image Encryption Algorithms: A Survey of Design and Evaluation Metrics. *J. Cybersecur. Priv.* **2024**, *4*, 126–152. [CrossRef]
25. Diaconu, A.V.; Dascalescu, A.C. Correlation distribution of adjacent pixels randomness test for image encryption. *Proc. Rom. Acad. Ser. A* **2017**, *18*, 351–360.
26. Luengo, E.A.; Olivares, B.A.; Villalba, L.J.G.; Hernandez-Castro, J. Further analysis of the statistical independence of the NIST SP 800-22 randomness tests. *Appl. Math. Comput.* **2023**, *459*, 128222. [CrossRef]
27. Cirrone, G.; Donadio, S.; Guatelli, S.; Mantero, A.; Mascialino, B.; Parlati, S.; Pia, M.; Pfeiffer, A.; Ribon, A.; Viarengo, P. A goodness-of-fit statistical toolkit. *IEEE Trans. Nucl. Sci.* **2004**, *51*, 2056–2063. [CrossRef]
28. Brereton, R.G. The chi squared and multinormal distributions. *J. Chemometr.* **2014**, *29*, 9–12. [CrossRef]
29. Zhang, X.; Liu, G.; Zou, C. An image encryption method based on improved Lorenz chaotic system and Galois field. *Appl. Math. Model.* **2024**, *131*, 535–568. [CrossRef]
30. Wen, H.; Xie, Z.; Wu, Z.; Lin, Y.; Feng, W. Exploring the future application of UAVs: Face image privacy protection scheme based on chaos and DNA cryptography. *J. King Saud Univ.-Comput. Inf. Sci.* **2024**, *36*, 101871. [CrossRef]
31. Singh, L.D.; Thingbaijam, R.; Patgiri, R.; Singh, K.M. Cryptanalysis of cross-coupled chaotic maps multi-image encryption scheme. *J. Inf. Secur. Appl.* **2024**, *80*, 103694. [CrossRef]
32. Bailey, D.H.; Crandall, R.E. Random generators and normal numbers. *Exp. Math.* **2002**, *11*, 527–546. [CrossRef]
33. Ponuma, R.; Amutha, R. Compressive sensing based image compression-encryption using novel 1D-chaotic map. *Multimed. Tools Appl.* **2018**, *77*, 19209–19234. [CrossRef]
34. Iqbal, N.; Hussain, I.; Khan, M.A.; Abbas, S.; Yousaf, S. An efficient image cipher based on the 1D scrambled image and 2D logistic chaotic map. *Multimed. Tools Appl.* **2023**, *82*, 40345–40373. [CrossRef]
35. Du, L.; Teng, L.; Liu, H.; Lu, H. Multiple face images encryption based on a new non-adjacent dynamic coupled mapping lattice. *Expert Syst. Appl.* **2024**, *238*, 121728. [CrossRef]
36. Xian, Y.; Wang, X. Fractal sorting matrix and its application on chaotic image encryption. *Inf. Sci.* **2021**, *547*, 1154–1169. [CrossRef]

37. Wang, C.; Chong, Z.; Zhang, H.; Ma, P.; Dong, W. Color image encryption based on discrete memristor logistic map and DNA encoding. *Integration* **2024**, *96*, 102138. [CrossRef]
38. Holzer, M.; Kutrib, M. Descriptional and computational complexity of finite automata—A survey. *Inf. Comput.* **2011**, *209*, 456–470. [CrossRef]