ON THE NATURE OF PRIMES: A DETERMINISTIC, ENDOGENOUS, NON-STATIONARY S-ADIC AUTOMATON FOR THE SIEVE OF ERATOSTHENES

BIRKE HEEREN

Abstract. We present a deterministic, endogenous, non-stationary S-adic automaton that realizes the Sieve of Eratosthenes as a symbolic dynamical system over a finite alphabet. Its evolution is governed by three operators—shift, expansion, and filtering—acting on a growing symbolic tape, and it reproduces the classical prime–composite classification for every integer $n \geq 2$.

A central result is the Stability Zone

$$\mathrm{SZ}_n = [n+1,\, 2n-1],$$

an interval in which the symbolic state is provably immune to all later filtering steps. This yields pointwise stability of the prime encoding and makes large-scale experimental verification possible through a Frozen Window technique, which we implement up to $n = 250{,}000$.

The tape also exhibits a canonical four-letter substructure $\{a, b, c, d\}$ governed by explicit prime-dependent substitution rules and an upper-triangular transition matrix M_p.

The automaton is not intended as an efficient prime generator, but as a symbolic research instrument in which arithmetic properties of the natural numbers become accessible to combinatorial and dynamical analysis.

Contents

2020 *Mathematics Subject Classification.* 11N05, 11B85, 37B10, 68Q45, 28A80.

Key words and phrases. Sieve of Eratosthenes; prime numbers; symbolic dynamics; S-adic systems; substitution morphisms; deterministic automata; fractal dimension; twin prime conjecture; Stability Zone; transition matrix; Hardy-Littlewood conjecture; non-stationary dynamical systems; primorial; Hydra effect.

Corresponding author. E-mail: birke.heeren@uni-rostock.de | ORCID: https://orcid.org/0000-0001-6713-583X.

1. Introduction

The distribution of prime numbers is one of the central subjects of number theory. Despite major advances—from the Prime Number Theorem to the Riemann Hypothesis—the local structure of the primes remains difficult to describe in a deterministic way. Statistical models and probabilistic heuristics, such as the Cramér model [8] and the Hardy–Littlewood conjectures [15], capture important large-scale features of prime gaps and prime constellations. However, they do not provide an explicit dynamical mechanism that generates the prime–composite pattern symbolically, step by step.

This paper takes a different point of view. Instead of asking how primes behave statistically, we ask whether the classical Sieve of Eratosthenes can be realized as a deterministic symbolic dynamical system, and what structural information becomes visible when the sieve is represented in that form.

We present a deterministic, endogenous, non-stationary S-adic automaton whose internal state evolves through three operators: a shift S_n, an expansion X_n, and a filter F_n (see Figure 1). The automaton maintains a cyclic symbolic tape over the binary alphabet $\Sigma_{\mathrm{CP}} = \{L, M\}$ and processes one natural number at each step. No external primality test is applied: the classification of a number as prime or composite is determined entirely by the symbolic state of the tape at the moment of encoding. In this sense, the automaton is endogenous. We show that it reproduces exactly the same prime–composite classification as the classical Sieve of Eratosthenes for every integer $n \geq 2$.

What distinguishes the automaton from the classical sieve is not what it computes, but what it records. At each step, the tape carries a periodic symbolic representation of the sieve state to the right of the current position. This representation is not introduced as a more efficient way to generate primes. Rather, it is a symbolic research instrument: a structured internal state in which arithmetical properties of the sieve become accessible to combinatorial and dynamical analysis (see Figure 2).

A central result of the paper is the *Stability Zone*,

$$\mathrm{SZ}_n = [n+1,\, 2n-1],$$

an interval in which the symbolic state is provably immune to all later filtering steps. Inside this zone, every surviving L-symbol is stable until it reaches the head of the tape. This gives a pointwise notion of stability for the prime encoding and provides the theoretical basis for the Frozen Window technique used in the experimental part of the paper.

A second structural feature is the emergence of a canonical four-letter substructure $\{a, b, c, d\}$ on letters of length 6. From this viewpoint, the binary tape admits a second-order symbolic description governed by explicit prime-dependent substitution rules and an upper-triangular transition matrix M_p. The letter

$$a = \langle LMLMMM \rangle$$

plays a special role as the symbolic template associated with candidate pairs at distance 2. Its global population evolves according to a simple multiplicative rule, which we call the Hydra Effect. This connects the automaton to classical local sieve patterns related to twin-prime heuristics, although the paper does not claim a proof of the Twin Prime Conjecture.

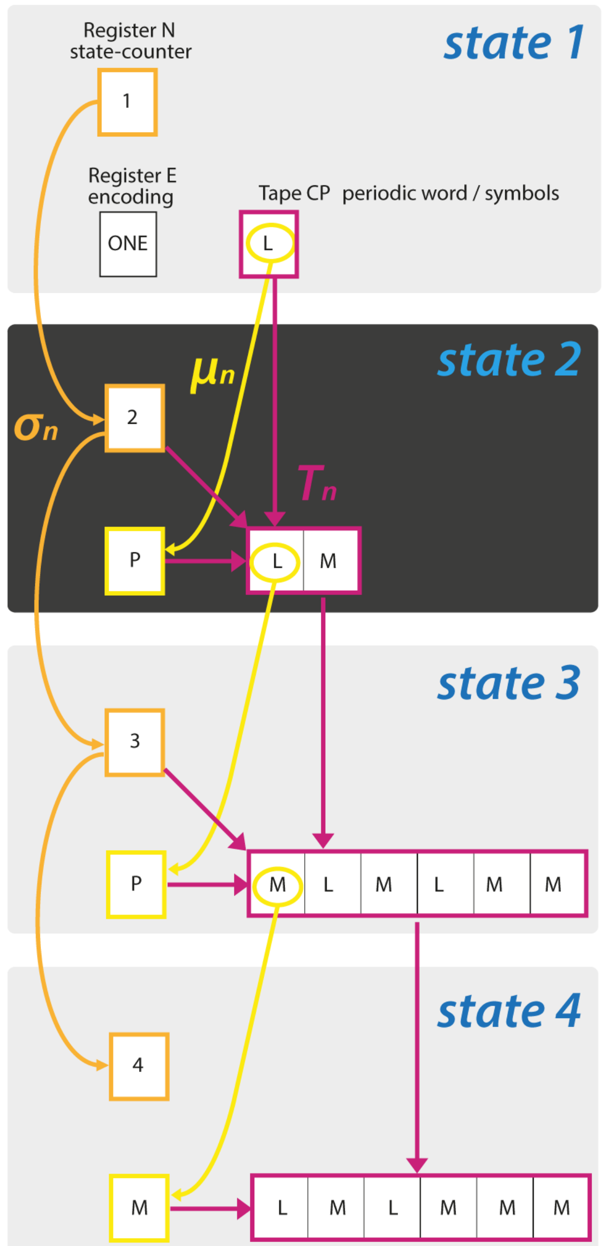

FIGURE 1. Dynamic development of the non-stationary S-adic system. The automaton evolves through four states, showing Register N (state counter), Register E (encoding), and Tape CP (binary periodic word). At state 2, the first prime $n = 2$ triggers the expansion operator X_n and filter operator F_n, growing the tape from width 1 to width 2. The three symbolic domains $\mathbb{N}$, $\Sigma_E = \{\mathrm{ONE}, P, M\}$, and $\Sigma_{\mathrm{CP}} = \{L, M\}$ are visually distinct.

E				
	P	L	M	CP
N	2	3	4	head
		5	6	
		7	8	
		9	10	
		11	12	
		13	14	
		15	16	
		17	18	
		19	20	
		21	22	
		23	24	
		25	26	
		27	28	
		29	30	
		31	32	
		33	34	
		35	36	
		etc. → ∞		
		width=2		

E								
	P	M	L	M	L	M	M	CP
N	3	4	5	6	7	8	9	head
		10	11	12	13	14	15	
		16	17	18	19	20	21	
		22	23	24	25	26	27	
		28	29	30	31	32	33	
		34	35	36	37	38	39	
		40	41	42	43	44	45	
		46	47	48	49	50	51	
		52	53	54	55	56	57	
		58	59	60	61	62	63	
		64	65	66	67	68	69	
		70	71	72	73	74	75	
		76	77	78	79	80	81	
		82	83	84	85	86	87	
		88	89	90	91	92	93	
		94	95	96	97	98	99	
		100	101	102	103	104	105	
		etc. → ∞						
		width=6						

E								
	P	L	M	L	M	M	M	CP
N	4	5	6	7	8	9	10	head
		11	12	13	14	15	16	
		17	18	19	20	21	22	
		23	24	25	26	27	28	
		29	30	31	32	33	34	
		35	36	37	38	39	40	
		41	42	43	44	45	46	
		47	48	49	50	51	52	
		53	54	55	56	57	58	
		59	60	61	62	63	64	
		65	66	67	68	69	70	
		71	72	73	74	75	76	
		77	78	79	80	81	82	
		83	84	85	86	87	88	
		89	90	91	92	93	94	
		95	96	97	98	99	100	
		101	102	103	104	105	106	
		etc. → ∞						
		width=6						

E																																
	P	M	L	M	M	M	L	M	L	M	M	M	L	M	L	M	M	M	L	M	M	M	M	M	L	M	L	M	M	M	M	CP
N	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	head
		36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	
		66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	
		96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	
		etc. → ∞																														
		width=30																														

FIGURE 2. Sieves from steps 2 to 5 of the automaton. M-columns (gray) contain composites only; L-columns (white) contain positions that have not yet been eliminated and therefore may contain primes or composites.

The purpose of the paper is therefore twofold. First, it gives an exact symbolic realization of the Sieve of Eratosthenes as an endogenous non-stationary S-adic system. Second, it develops structural tools—the four-letter substructure, the transition matrix, the Stability Zone, and the Frozen Window technique—for studying the internal symbolic organization of that system.

1.1. Related Work.

The distribution of prime numbers has been studied extensively from analytic, algorithmic, and probabilistic perspectives [2]. The Prime Number Theorem, proved independently by Hadamard [13] and de la Vallée Poussin [9], shows that the number of primes up to x is asymptotic to $x/\ln x$. The Cramér model [8] provides a probabilistic framework for prime gaps by modeling the probability that an integer x is prime as $1/\ln x$. Although highly influential, this model was shown by Maier [19] to make incorrect predictions in short intervals, illustrating the limitations of purely probabilistic approaches.

On the algorithmic side, the classical Sieve of Eratosthenes [10] provides a procedural method for identifying primes by iteratively removing multiples. More advanced sieve methods, including Brun's sieve and later developments surveyed by Halberstam and Richert [14], extend this basic idea to deeper questions about prime distribution. These methods are extremely powerful analytically, but they are not typically formulated as symbolic dynamical systems with an explicit internal state.

The Hardy–Littlewood conjectures [15] extend probabilistic reasoning to prime constellations, including twin primes and more general k-tuples, through singular series defined as products over primes. These conjectures remain open. In the present paper, we do not attempt to prove them. Rather, we study a deterministic symbolic framework whose local combinatorial structure is consistent with finite sieve patterns related to such heuristics.

From a different direction, symbolic dynamics and S-adic systems provide tools for studying sequences generated by substitutions over finite alphabets. Foundational treatments are given by Lind and Marcus [18] and Allouche and Shallit [1], while the non-stationary S-adic setting is developed by Fogg [12] and Berthé and Delecroix [4]. These frameworks have been applied to many combinatorial sequences, but the classical Sieve of Eratosthenes has not, to our knowledge, been developed as a deterministic symbolic dynamical system with an explicit evolving internal tape representation of the sieve state.

Connections between number theory and fractal geometry have also been explored by Lapidus and van Frankenhuijsen [16], where zeta functions and complex dimensions provide a geometric framework for number-theoretic structures. Our viewpoint is different in both method and scope. We do not identify the prime set with a classical fractal object. Instead, we associate the sieve with a non-stationary symbolic filtering process and a corresponding local scaling exponent.

A related but independent approach to primality without divisibility has been proposed by Bilokon [5], who introduces an imbalance metric on integer pairs together with a Möbius transformation of that metric. In that framework, primality is characterized by a novelty condition on imbalance values. Bilokon's work is conceptually parallel in that it seeks a deterministic description of primality outside the usual divisibility-based formulation. However, the present paper proceeds in a different direction: it models the Sieve of Eratosthenes itself as an endogenous symbolic dynamical system over a finite alphabet.

Against this background, the contribution of the present paper is not a new analytic estimate for primes, but an exact symbolic realization of the sieve together with structural tools—the four-letter substructure, the transition matrix, the Stability Zone, and the Frozen Window technique—for studying the internal organization of that realization.

1.2. **Organization of the Paper.** The paper is organized as follows. Section 2 states the main topic and explains the role of the automaton as a symbolic research instrument. Section 3 defines the automaton and its operators. Section 4 proves synchronization with the number line and establishes equivalence with the classical Sieve of Eratosthenes. Section 5 studies the associated substitution rules, transition matrix, and Hydra dynamics. Section 6 introduces the Stability Zone and discusses its arithmetical and symbolic consequences. Section 7 presents the Frozen Window experiment. Section 8 concludes.

At the end of the paper, we provide the bibliography, a declaration of AI-assisted tools used during preparation, and acknowledgments.

2. Topic

2.1. **Topic: A Symbolic Research Instrument for the Sieve of Eratosthenes.** The automaton $\mathcal{A}$ is not proposed as an efficient algorithm for generating primes, but as a *symbolic research instrument*: a deterministic, endogenous, non-stationary S-adic system whose internal state makes structural properties of the Sieve of Eratosthenes accessible to symbolic, combinatorial, and geometric analysis. The purpose of this section is not to prove the main results, but to summarize the role of the framework and the kinds of structures it reveals.

We consider an automaton

$$\mathcal{A} = (N, E, \mathrm{CP}; S_n, X_n, F_n)_{n\geq 1},$$

where N is a counter over the natural numbers, E is a finite encoding register, and CP is a cyclic symbolic tape over the binary alphabet $\Sigma_{\mathrm{CP}} = \{L, M\}$. The tape evolves through three operators: a shift S_n, an expansion X_n, and a filter F_n. The current encoding E_n is determined entirely by the symbolic state of the tape, so the automaton is endogenous in the sense that no external primality test or precomputed prime list is used.

The framework has five central features.

(i) **Exact symbolic realization of the sieve.** The automaton reproduces the same prime–composite classification as the classical Sieve of Eratosthenes. At each step, the tape provides a symbolic record of which positions have already been eliminated and which remain prime candidates.

(ii) **Periodic internal state.** At every finite step n, the tape is a periodic binary word whose width equals the primorial of the largest prime $p \leq n$:

$$|\mathrm{CP}_n| = \prod_{p\leq n} p.$$

Thus the sieve is represented not only procedurally, but also as an explicit symbolic state space.

(iii) **Emergent finite-alphabet structure.** From step $n = 4$ onward, the tape admits a canonical letter decomposition of width 6, yielding a second-order alphabet

$$\Sigma_2 = \{a, b, c, d\},$$

where

$$a = \langle LMLMMM\rangle, \quad b = \langle LMMMMM\rangle, \quad c = \langle MMLMMM\rangle, \quad d = \langle MMMMMM\rangle.$$

This four-letter structure provides a compressed symbolic description of the local sieve pattern.

(iv) **A built-in Stability Zone.** For each step n, the interval

$$\mathrm{SZ}_n = [n + 1,\ 2n - 1]$$

is provably stable under all later filtering steps. Inside this zone, surviving L-symbols correspond to confirmed primes. This gives the automaton a built-in notion of pointwise symbolic stability.

(v) **A structural viewpoint on local scaling.** The prime-dependent dynamics of the tape lead naturally to a local scaling exponent

$$D(n) = \frac{\log(a + b + c)}{\log(a + b + c + d)},$$

(with a, b, c, and d meaning the amount of the letters) which tends to 1 as $p \to \infty$. We interpret this as a symbolic signature of the progressively weaker local exclusion imposed by larger prime filters.

From this perspective, the automaton is best understood as a framework for studying the internal organization of the sieve rather than as a faster method for computing primes. Its main contribution is to turn the sieve into an explicit symbolic dynamical object whose structure can be analyzed at several levels: binary, letter-combinatorial, spectral, and experimental.

The following sections develop these aspects in detail: first by defining the automaton precisely, then by proving its equivalence to the classical sieve, and finally by analyzing the emergent symbolic structures that arise from its evolution.

3. The Automaton

3.1. **Symbolic Domains.** The automaton operates on three distinct domains, corresponding to its three components: the counter N, the encoding register E, and the symbolic tape CP.

Register N**.** Register N holds the current natural number. Its state at step n is simply $N_n = n$, generated by the successor operation [22]. Thus the automaton step and the represented natural number coincide.

Register E**.** Register E stores the qualitative encoding of the current number. Its alphabet is

$$\Sigma_E = \{\mathrm{ONE}, P, M\}, \tag{1}$$

where ONE marks the number 1, P marks a prime, and M marks a composite. These three cases are mutually exclusive and exhaustive.

Tape CP**.** The tape carries a symbolic representation of the sieve state on the number line to the right of the current step. Its alphabet is binary:

$$\Sigma_{\mathrm{CP}} = \{L, M\}, \tag{2}$$

where L indicates a surviving candidate position and M indicates a position already identified as composite. The tape records only this binary distinction.

This separation of roles is essential. Register E provides the current prime–composite encoding, whereas the tape CP records the evolving sieve state. In particular, the tape does not by itself distinguish between *candidate* and *confirmed* prime positions globally; that distinction becomes pointwise valid in the Stability Zone introduced later.

3.2. **Structure of the Automaton.** The automaton $\mathcal{A}$ consists of two single-cell registers and one cyclic symbolic tape.

Register N**.** The counter evolves by the successor rule

$$N_n = N_{n-1} + 1, \qquad N_1 = 1. \tag{3}$$

Register E**.** At step n, the encoding function reads the first tape symbol of the previous state and assigns

$$E_n = \begin{cases} \mathrm{ONE} & \text{if } \mathrm{CP}_{n-1}[1] \text{ does not exist}, \\ P & \text{if } \mathrm{CP}_{n-1}[1] = L, \\ M & \text{if } \mathrm{CP}_{n-1}[1] = M. \end{cases} \tag{4}$$

Thus the current classification is determined entirely by the symbolic state of the automaton. The value $E_n = \text{ONE}$ occurs only at $n = 1$.

Tape CP**.** The tape begins with one cell and grows during the evolution of the automaton. Its leftmost cell is denoted $\text{CP}_n[1]$, and its rightmost cell is denoted $\text{CP}_n[\text{last}]$. At every finite step n, the tape is a periodic word over $\Sigma_{\text{CP}} = \{L, M\}$. Its width is

$$|\text{CP}_n| = \prod_{p \le n} p, \tag{5}$$

where the product runs over all primes $p \le n$. The tape grows only at prime steps.

3.3. **Operators.** The evolution of the tape is governed by three operators acting on the binary alphabet $\Sigma_{\text{CP}} = \{L, M\}$.

Shift operator S_n**.** The first symbol CP[1] is removed from the left end of the tape and appended to the right end as the new CP[last]. This is a cyclic rotation by one position. The shift keeps the tape synchronized with the advance of the counter Figure 3.

Expansion operator X_n**.** This operator is applied only when $E_n = P$. It replaces the tape by n concatenated copies of itself. Thus the width is multiplied by n, but the relative pattern inside each copy is unchanged.

Filter operator F_n**.** This operator is also applied only when $E_n = P$. It traverses the tape with stride n, beginning at position $2n$, and replaces each visited L by M, leaving existing M-symbols unchanged. In this way, the filter marks all multiples of n represented on the tape.

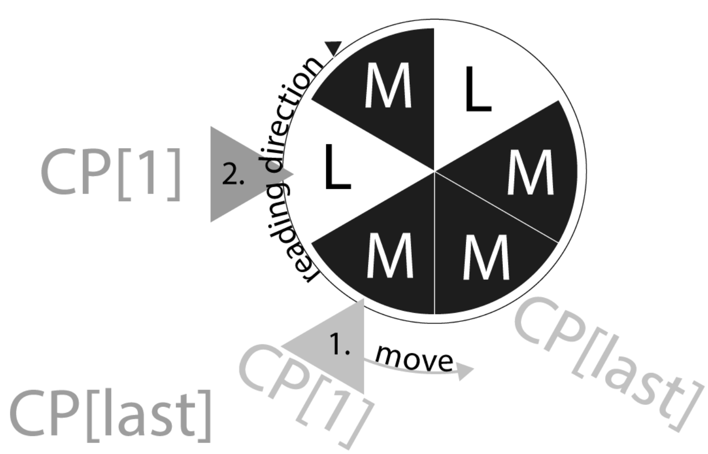

FIGURE 3. The tape CP_3 displayed as a cyclic structure. The shift operator rotates the tape by one position to produce CP_4, moving $\text{CP}_3[1]$ to $\text{CP}_4[\text{last}]$. This preserves periodicity and keeps Register N and the tape in step.

3.4. **Transition Function T_n.** The complete tape update at step n is

$$\mathrm{CP}_n = \begin{cases} F_n\big(X_n(S_n(\mathrm{CP}_{n-1}))\big) & \text{if } E_n = P, \\ S_n(\mathrm{CP}_{n-1}) & \text{if } E_n = M. \end{cases} \tag{6}$$

The initial state is

$$N_1 = 1, \qquad E_1 = \mathrm{ONE}, \qquad \mathrm{CP}_1 = \langle L \rangle.$$

When $E_n = P$, the order of the three operators is essential. The shift must come first so that the tape is aligned with the current number before replication. The expansion must precede the filter so that all copies are filtered consistently. If the filter were applied before expansion, the newly appended copies would remain unfiltered at that step.

3.5. **Note on the Binary Tape.** The choice of the binary alphabet $\Sigma_{\mathrm{CP}} = \{L, M\}$ is structural rather than merely notational. Each tape position records exactly one binary decision: either the corresponding residue class has already been eliminated or it has not.

This minimal representation has two consequences that matter later. First, the higher-order symbolic structure of the automaton must emerge from repeated patterns on the binary tape itself. Second, the resulting system fits naturally into the language of substitutions over finite alphabets, which makes the connection with S-adic methods precise.

3.6. **Illustrative Example: First Five Steps.** The first five states of the automaton are shown in Table 1.

n	E_n	CP_n	Width
1	ONE	$\langle L \rangle$	1
2	P	$\langle LM \rangle$	2
3	P	$\langle MLMLMM \rangle$	6
4	M	$\langle LMLMMM \rangle$	6
5	P	$\langle MLMMM\ LMLMMM\ LMLMMM\ LMMMMM\ LMLMMM\ M \rangle$	30

TABLE 1. The first five states of the automaton. Please compare with Figure 4. At $n = 2$, the tape expands from width 1 to width 2, and the filter marks position 4 as M. At $n = 3$, the tape expands from width 2 to width 6, and the filter marks position 9 as M. At $n = 4$, only the shift is applied. At $n = 5$, the tape expands from width 6 to width 30, and the filter marks positions 15 and 35 as M.

The same process can be viewed visually in Figure 4. The example shows how the initial one-symbol tape $\langle L \rangle$ develops into longer periodic tapes under repeated shifts, expansions, and filtering.

3.7. **Foundations of the Representation.** To avoid ambiguity, we distinguish between the absolute number line and its symbolic representation inside the automaton.

At step n, the state of the automaton is

$$(N_n, E_n, \mathrm{CP}_n),$$

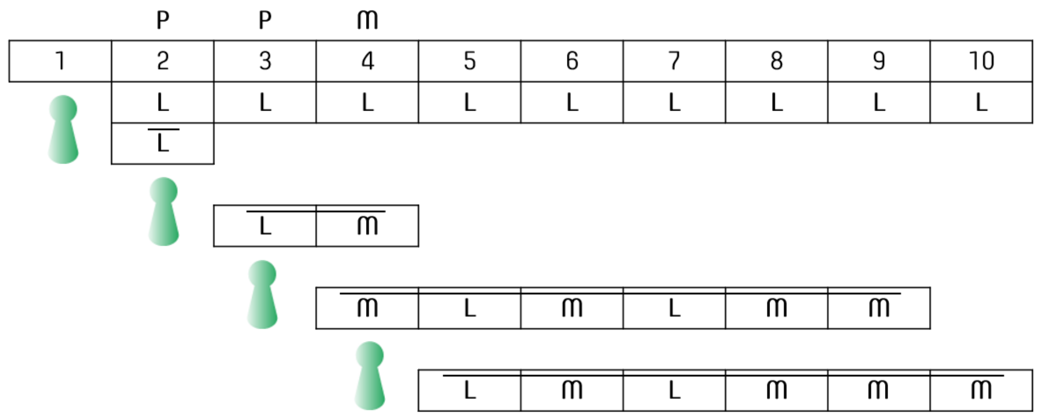

FIGURE 4. The algorithm in its first four steps. The green marker indicates the current position of Register E_n, that is, the natural number being processed. As the step number increases from $n = 1$ to $n = 4$, the tape grows and shifts accordingly.

where $N_n = n$, $E_n \in \{\text{ONE}, P, M\}$, and CP_n is a finite periodic binary word. The tape is not a second number line. Rather, it is the symbolic realization of the portion of the number line strictly to the right of the current step.

More precisely, the tape is always read from left to right starting at the integer $n+1$. Thus for every tape position $j \geq 1$, the cell $\text{CP}_n[j]$ canonically represents the integer

$$n + j.$$

This identification allows local tape patterns to be interpreted as local arithmetical patterns on the absolute number line.

At the same time, the tape is periodic Figure 2 with width $|\text{CP}_n|$, so it can also be studied internally as a finite symbolic object. These two viewpoints—canonical realization on the number line and periodic symbolic structure—are both used in the sequel. The first is needed for arithmetic interpretation; the second is needed for the combinatorial and S-adic analysis.

We also distinguish between formal statements proved within the framework and interpretive language used to explain them. Statements such as synchronization, equivalence with the classical sieve, and the existence of the Stability Zone are proved formally. Terms such as *research instrument* or *symbolic observatory* are explanatory descriptions of the role of the framework.

Definition 1 (Endogenous system)**.** A dynamical system is *endogenous* if its evolution is determined entirely by its internal state and internal rules, without reliance on external inputs, external tests, or precomputed arithmetic data.

For the automaton

$$\mathcal{A} = (N, E, \text{CP}; S_n, X_n, F_n)_{n \geq 1},$$

this means:

- the operators S_n, X_n, and F_n act only on the internal symbolic state of the tape;

- the classification of the current number is determined by $\mathrm{CP}_{n-1}[1]$ through (4);
- the state transitions depend only on the current state and the fixed update rules of the automaton.

The automaton is therefore introduced not as a faster way to compute primes, but as an exact symbolic model of the sieve whose internal organization can be studied directly.

4. Synchronization and Equivalence to the Sieve of Eratosthenes

4.1. **Synchronization.** The automaton processes one natural number at each step. The following lemma shows that the first tape position is always aligned with the next number to be encoded.

Lemma 1 (Synchronization)**.** *For every step $n \geq 1$, the symbol at position $\mathrm{CP}_{n-1}[1]$ corresponds to the natural number n and is the symbol read by Register E at step n.*

Proof. We argue by induction on n.

Base case. At step $n = 1$, the initial tape is $\mathrm{CP}_1 = \langle L \rangle$, and the counter holds $N_1 = 1$. By initialization, the first position $\mathrm{CP}_{n-1}[1]$ does not exist.

Inductive step. Assume that at step n, the first symbol of CP_{n-1} corresponds to the natural number n. We show that after the transition at step n, the first symbol of CP_n corresponds to $n+1$.

If $E_n = M$, then
$$\mathrm{CP}_n = S_n(\mathrm{CP}_{n-1}).$$
The shift operator removes the first tape symbol and appends it to the right end. Therefore the new first symbol is exactly the symbol that previously occupied position 2, which corresponds to the successor of the previously encoded number. Hence $\mathrm{CP}_n[1]$ corresponds to $n+1$.

If $E_n = P$, then
$$\mathrm{CP}_n = F_n\big(X_n(S_n(\mathrm{CP}_{n-1}))\big).$$
Again, the shift first moves the tape into the correct alignment, so after S_n the first position corresponds to $n+1$. The expansion operator X_n appends copies to the right and does not change the first position. The filter operator F_n begins at position $2n$, so it also does not alter the first position. Hence $\mathrm{CP}_n[1]$ still corresponds to $n+1$.

Thus in either case, the first symbol of CP_n corresponds to $n+1$. This completes the induction. ☐

Remark 1. The synchronization lemma makes the encoding function well defined: at each step, Register E reads the tape symbol corresponding to the current natural number.

4.2. **Equivalence to the Sieve of Eratosthenes.** We now prove that the automaton reproduces exactly the same prime–composite classification as the classical Sieve of Eratosthenes.

Theorem 1 (Equivalence to the Sieve of Eratosthenes)**.** *For every integer $n \geq 2$,*

$$n \in \mathbb{P} \iff \mathrm{CP}_{n-1}[1] = L, \tag{7}$$
$$n \notin \mathbb{P} \iff \mathrm{CP}_{n-1}[1] = M. \tag{8}$$

Therefore, the automaton produces exactly the same prime–composite classification as the classical Sieve of Eratosthenes.

Remark 2 (The number 1). The number 1 is processed at step $n = 1$ with encoding $E_1 =$ ONE. It is neither prime nor composite and is therefore excluded from Theorem 1.

Proof. By Lemma 1, the symbol $\mathrm{CP}_{n-1}[1]$ corresponds to the natural number n at step n.

($\Rightarrow$) Suppose n is prime. Then n has no prime divisor p with $2 \le p < n$. For every prime $p < n$, the filter F_p marks exactly those positions corresponding to multiples of p. Since $p \nmid n$, the position corresponding to n is never visited by F_p. Therefore no filter applied before step n changes that position from L to M, so

$$\mathrm{CP}_{n-1}[1] = L.$$

($\Leftarrow$) Suppose n is composite. Then there exists a prime p with $2 \le p < n$ such that $p \mid n$. At step p, the filter F_p visits the position corresponding to n and changes its symbol to M. Since $p < n$, this happens before step n. Hence

$$\mathrm{CP}_{n-1}[1] = M.$$

The two implications establish both equivalences. □

Remark 3. The symbol ONE in Register E ensures that step $n = 1$ is treated separately and that no filter is applied there.

4.3. **Width of the Tape.** The next lemma records the growth law of the tape.

Lemma 2 (Width of the tape)**.** *For every step $n \ge 1$,*

$$|\mathrm{CP}_n| = \prod_{p \le n} p, \tag{9}$$

where the product ranges over all primes $p \le n$. For $n = 1$, the empty product is understood to be 1.

Proof. At step $n = 1$, the tape has one cell, so $|\mathrm{CP}_1| = 1$, which agrees with the empty product.

Now assume $n > 1$. There are two cases.

If n is composite, then

$$\mathrm{CP}_n = S_n(\mathrm{CP}_{n-1}),$$

and the shift does not change the tape width. Since the set of primes $\le n$ is the same as the set of primes $\le n - 1$, the product remains unchanged.

If n is prime, then

$$\mathrm{CP}_n = F_n\big(X_n(S_n(\mathrm{CP}_{n-1}))\big).$$

The shift preserves width, the expansion multiplies the width by n, and the filter preserves width. Hence

$$|\mathrm{CP}_n| = n \cdot |\mathrm{CP}_{n-1}|.$$

Since n is a new prime factor at this step, the product over primes $\le n$ is obtained from the product over primes $\le n - 1$ by multiplication by n.

Thus the formula holds for all n. □

Remark 4 (Growth rate). By Lemma 2, the width of the tape is the primorial

$$\prod_{p \le n} p.$$

Equivalently,

$$\log |\mathrm{CP}_n| = \theta(n),$$

where θ is Chebyshev's function. In particular, the tape width Figure 5 grows on the order of e^n, which confirms that the automaton is not intended as a computationally efficient prime generator.

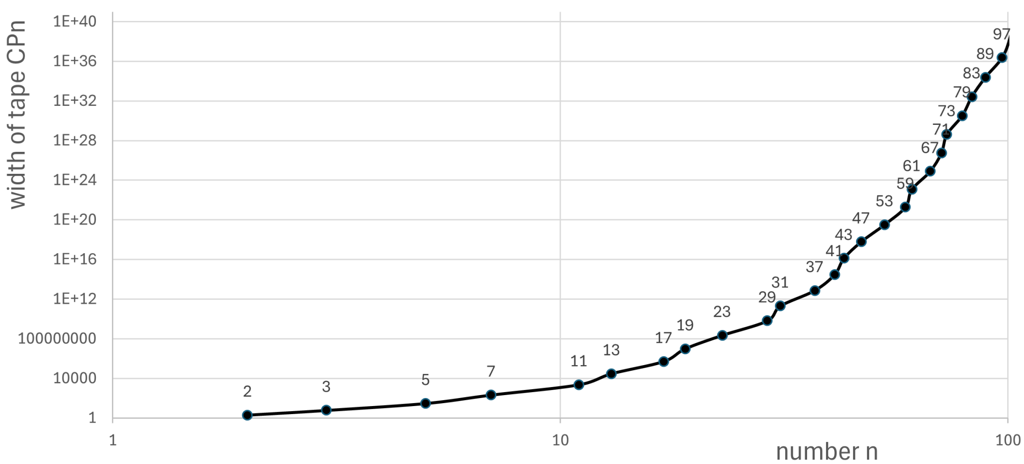

FIGURE 5. Width of the tape CP_n as a function of n on a log-log scale. By Lemma 2, the width equals the primorial $\prod_{p\le n} p$. The rapid growth reflects the expansion that occurs at each prime step.

5. The Four-Letter Substructure, Substitution Rules, and the Hydra Effect

5.1. **Emergence of the Substructure.** After the prime steps $n = 2$ and $n = 3$, the tape at step $n = 4$ has width 6 and carries the binary word

$$\mathrm{CP}_4 = \langle L\, M\, L\, M\, M\, M\rangle. \tag{10}$$

This word is the nontrivial periodic pattern produced by the automaton, and it serves as the fundamental letter for the higher-order symbolic structure that appears in all subsequent tape states.

From step $n = 4$ onward, the tape admits a canonical decomposition into letters of length 6. These letters belong to a finite second-order alphabet.

Definition 2 (Four-letter alphabet)**.** Define the following binary words of length 6:

$$a := \langle L\,M\,L\,M\,M\,M \rangle, \tag{11}$$

$$b := \langle L\,M\,M\,M\,M\,M \rangle, \tag{12}$$

$$c := \langle M\,M\,L\,M\,M\,M \rangle, \tag{13}$$

$$d := \langle M\,M\,M\,M\,M\,M \rangle. \tag{14}$$

The associated second-order alphabet is

$$\Sigma_2 = \{a, b, c, d\}.$$

Remark 5 (Recursive symbolic structure)*.* The four-letter alphabet is the first higher-order symbolic layer above the binary tape. At later prime steps, longer tape segments may themselves be viewed as composite symbolic units. In this sense, the four-letter structure is the first stage of a recursive symbolic organization rather than a final one.

5.2. **Structural Interpretation.** The four letters admit a natural local interpretation.

Letter $a = \langle LMLMMM \rangle$**.** This letter contains two surviving positions at distance 2, namely its first and third entries. It is therefore the unique letter type associated with a candidate pair at distance 2.

Letter $b = \langle LMMMMM \rangle$**.** This letter contains exactly one surviving position, at its first entry.

Letter $c = \langle MMLMMM \rangle$**.** This letter also contains exactly one surviving position, now at its third entry.

Letter $d = \langle MMMMMM \rangle$**.** This letter contains no surviving positions.

Thus the alphabet Σ_2 records, at letter scale 6, whether a local configuration contains two surviving positions, one surviving position on the left, one surviving position on the right, or none.

Remark 6 (Completeness)*.* Although there are $2^6 = 64$ binary words of length 6, only the four letters a, b, c, d occur in the tape from step $n = 3$ onward. This reflects the specific elimination rules of the sieve: the filter operators may remove surviving positions, but they never create new ones.

5.3. **Stability of the Substructure.**

Lemma 3 (Stability of Σ_2)**.** *For every step $n \geq 3$, the tape CP_n is a concatenation of letters from $\Sigma_2 = \{a, b, c, d\}$.*

Proof. At step $n = 4$,

$$\mathrm{CP}_4 = \langle LMLMMM \rangle = \langle a \rangle,$$

so the claim holds initially (step 4 is a shift from step 3).

Assume now that CP_{n-1} is a concatenation of letters from Σ_2. We consider the two cases of the transition function.

If $E_n = M$, then

$$\mathrm{CP}_n = S_n(\mathrm{CP}_{n-1}).$$

Because the tape is periodic and its width is a multiple of 6 for all $n \geq 4$ by Lemma 2, the cyclic shift preserves the decomposition into length-6 letters.

If $E_n = P$, then

$$\mathrm{CP}_n = F_n\big(X_n(S_n(\mathrm{CP}_{n-1}))\big).$$

The shift preserves the letter structure, and the expansion merely repeats the shifted tape. Thus after X_n, the tape is still a concatenation of letters from Σ_2.

Now consider the filter F_n. For prime steps with $n \geq 5$, the stride length is at least 5. In a letter of type a, the two surviving positions are separated by distance 2, so a single stride cannot eliminate both surviving positions within the same copy of the letter. More generally, the filter only changes symbols $L \to M$. Hence the only possible letter transitions are shown in Figure 6. Each image is again one of the four letters a, b, c, d. Therefore CP_n is a concatenation of letters from Σ_2.

By induction, the claim holds for all $n \geq 4$. □

5.4. **Transition Rules.** The proof of Lemma 3 yields the transition rules for the second-order alphabet. At any prime step $n \geq 5$, the filter may act on the four letter types as follows:

$$\begin{aligned} a &\to a,\ b,\ \text{or } c,\\ b &\to b \text{ or } d,\\ c &\to c \text{ or } d,\\ d &\to d. \end{aligned}$$

These transitions have three immediate consequences.

First, the process is irreversible at letter level: once a surviving symbol is changed from L to M, it is never restored.

Second, the letter d is absorbing.

Third, the letter a can only persist or lose one of its surviving positions under filtering. It is never created by a filter; it is only replicated by the expansion operator.

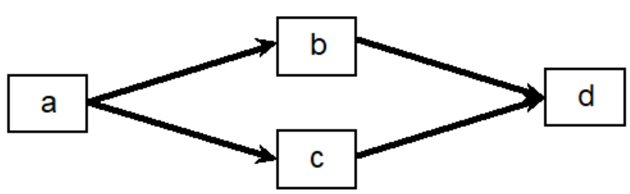

FIGURE 6. Transition rules for the four-letter alphabet $\Sigma_2 = \{a, b, c, d\}$. The letter $a = \langle LMLMMM \rangle$ may remain unchanged or lose one of its two surviving positions, producing b or c. The letters b and c may remain unchanged or lose their unique surviving position, producing d. The letter d is absorbing.

5.5. **Population Dynamics of Letter a: Growth and Elimination.** Let $G_n(a)$ denote the number of letter-a blocks present in the tape CP_n after step n. Let p_n be the current prime (i.e., $E_n = P$) and let $p_n^{\#}$ be the primorial width of the tape before the prime step.

- **Growth.** The expansion operator X_{p_n} first produces exactly p_n identical copies of every existing letter-a block. The intermediate population is therefore

$$G_{\text{temp}} = G_{n-1}(a) \cdot p_n. \tag{15}$$

- **Elimination.** The filter operator F_{p_n} traverses the tape with stride p_n. For any letter-a block whose leftmost L lies at position x (relative to the beginning of the tape), its p_n clones are located at positions

  $$x,\ x+w,\ x+2w,\ \ldots \quad \text{where } w = p_n^{\#}.$$

  Because p_n is coprime to the previous width w ($\gcd(w, p_n) = 1$), the Chinese Remainder Theorem guarantees that the congruence

  $$x + k \cdot w \equiv 0 \pmod{p_n}$$

  has *exactly one* solution $k \in \{0, 1, \ldots, p_n - 1\}$.

  Consequently, exactly one clone of the left L of each a-block is marked M, and exactly one clone of the right L (distance 2) is marked M. Since the filter stride $p_n \geq 5$ exceeds the twin distance 2, no single stride can hit both L's of the same template simultaneously. Thus, for every original a-block, *exactly two* clones are destroyed (converted to b and c), and precisely $p_n - 2$ clones survive as a.

The population of letter-a blocks therefore evolves according to the recurrence

$$G_n(a) = G_{n-1}(a) \cdot (p_n - 2), \tag{16}$$

with initial condition $G_4(a) = 1$.

This relation (16) is the *Hydra Equation*. It directly determines the (a, a)-entry of the transition matrix M_p (see subsection 5.7 below) and is equivalent to the substitution morphism $\sigma_p(a) = a^{p-2}bc$ introduced in the next subsection.

5.6. **Substitution Rules.** The transition behavior can be summarized by prime-dependent substitution rules. At a prime step p, each letter is first replicated p times by the expansion operator and then partially reduced by the filter.

This yields the following substitutions on Σ_2:

$$\sigma_p(a) = a^{p-2}bc, \tag{17}$$

$$\sigma_p(b) = b^{p-1}d, \tag{18}$$

$$\sigma_p(c) = c^{p-1}d, \tag{19}$$

$$\sigma_p(d) = d^p. \tag{20}$$

These formulas should be read as net symbolic transition rules for one full prime step: the expansion replicates each letter p times, and the filter removes the appropriate surviving positions among those copies.

Remark 7. The substitution rule for a reflects a simple residue-class argument. Across the p copies created by expansion, exactly one copy loses its left surviving position and exactly one loses its right surviving position. This is the symbolic reason for the appearance of one b, one c, and $p - 2$ surviving copies of a.

5.7. **Transition Matrix.** The substitutions σ_p are encoded by the transition matrix

$$M_p = \begin{pmatrix} p-2 & 1 & 1 & 0 \\ 0 & p-1 & 0 & 1 \\ 0 & 0 & p-1 & 1 \\ 0 & 0 & 0 & p \end{pmatrix}. \tag{21}$$

Here the rows and columns are ordered as (a, b, c, d), and the (i, j)-entry counts how often letter j appears in the image of letter i.

The matrix has three important features.

First, it is upper triangular, reflecting the irreversibility of the letter transitions.

Second, its diagonal entries

$$p-2, \quad p-1, \quad p-1, \quad p$$

are its eigenvalues.

Third, the roles of b and c are symmetric, as seen in the analogous middle rows.

5.8. **The Hydra Effect.** We now focus on the letter type a, which is the only letter carrying two surviving positions at distance 2.

Let $G_n(a)$ denote the total number of letter-a letters on the tape after step n. At a prime step p, each existing a-letter is replicated into p copies. The filter then transforms exactly one copy into b and exactly one copy into c, while the remaining $p-2$ copies stay of type a. Thus the total number of a-letters evolves by the recurrence (16) whenever step n is a prime step with prime p.

The name reflects the following structural phenomenon: at each prime step, the expansion operator produces many new copies of the twin-prime template a, while the filter destroys only two of the p copies associated with each preexisting template. Thus the absolute population of a continues to grow even though filtering is always active.

Corollary 1. *Starting from $G_4(a) = 1$, the total number of letter-a letters at prime level p is*

$$G(a) = \prod_{\substack{5 \le q \le p \\ q \text{ prime}}} (q-2).$$

Remark 8. The Hydra Effect is a statement about the *global population* of the symbolic template a on the periodic tape. By itself, it does not determine where those a-letters are located on the absolute number line. That later distribution problem is addressed separately in the discussion of the Q-Zone.

5.9. **Lemma on Three Consecutive Primes.** The four-letter structure also yields an immediate symbolic exclusion result.

Lemma 4 (No three consecutive odd primes beyond $3, 5, 7$)**.** *For every $n \ge 3$, the pattern*

$$\langle LMLML \rangle$$

does not occur in $\mathrm{CP}_{n\ge3}$.

Proof. A pattern of the form $\langle LMLML\rangle$ would require three surviving positions at pairwise distance 2.

Among the four letter types, only

$$a = \langle LMLMMM\rangle$$

contains two surviving positions, and its fifth symbol is M, not L. The letters b, c, and d contain at most one surviving position. Therefore no single letter contains $\langle LMLML\rangle$.

Since the tape is a concatenation of letters from Σ_2, the same remains true across letter boundaries: no adjacent pair of letters produces this pattern as a contiguous subword. Hence $\langle LMLML\rangle$ never occurs in CP_n. □

Remark 9. This is consistent with the classical fact that the only prime triple of the form $(p, p+2, p+4)$ is $(3, 5, 7)$. Within the automaton, the exclusion appears as a structural consequence of the four-letter letter system.

6. The Stability Zone, the Q-Zone, and an Open Distribution Problem

This section has two purposes. First, it introduces the Stability Zone and the Q-Zone as arithmetically safe regions of the tape. Second, it explains how the symbolic twin-prime template

$$a = \langle LMLMMM\rangle$$

interacts with these regions. The formal results concern stability and primality inside these zones. The later connection to twin primes is heuristic and is formulated as an open distribution problem.

6.1. **Definition and Motivation.**

At step n, the filter operator F_n acts with stride n. Its first new hit occurs at the position corresponding to the integer $2n$. Therefore the interval

$$[n+1,\, 2n-1]$$

is untouched by the current filter and all future filters.

More generally, if an integer z lies in this interval, then any composite structure of z must already involve a prime factor $p \le z/2 < n$. Hence any filter that could eliminate z as composite must already have acted before step n. This motivates the following definition.

Definition 3 (Stability Zone)**.** At step n, the *Stability Zone* is the interval

$$\mathrm{SZ}_n = [n+1,\, 2n-1]. \tag{22}$$

Remark 10. For $n = 1$, the interval $\mathrm{SZ}_1 = [2, 1]$ is empty. This is consistent with the fact that step 1 produces no filtering.

6.2. **Immutability of the Stability Zone.**

Theorem 2 (Stability of the Stability Zone)**.** *For every step $n \ge 2$ and every integer $z \in \mathrm{SZ}_n = [n+1, 2n-1]$, if the tape cell canonically corresponding to z carries the symbol L at step n, then that symbol remains L until z reaches the head of the tape. Equivalently,*

$$\mathrm{CP}_{z-1}[1] = L.$$

Proof. Let $z \in [n+1, 2n-1]$, and suppose that the cell corresponding to z carries the symbol L at step n.

A later filter F_q with $q > n$ could change that symbol only if $q \mid z$. Suppose such a prime q existed with $n < q < z$. Then

$$z \geq 2q > 2n,$$

which contradicts $z \leq 2n - 1$. Therefore no prime q with $n < q < z$ divides z.

Since filters are applied only at prime steps, no filter applied between steps $n + 1$ and $z - 1$ can affect the symbol corresponding to z. By Lemma 1, that symbol reaches the head position at step z, so

$$\mathrm{CP}_{z-1}[1] = L.$$

□

Remark 11. The index $z - 1$ is exact. Because the tape shifts left by one position at each step, the cell corresponding to z reaches the first position precisely at step z.

6.3. **Advancement of the Stability Zone.** The Stability Zone moves to the right as the automaton advances.

Lemma 5 (Advancement)**.** *For every $n \geq 2$,*

$$\mathrm{SZ}_n = [n + 1,\ 2n - 1] \quad \longrightarrow \quad \mathrm{SZ}_{n+1} = [n + 2,\ 2n + 1].$$

Thus the left boundary advances by 1, the right boundary advances by 2, and the width of the Stability Zone is

$$|\mathrm{SZ}_n| = n - 1.$$

Proof. This follows directly from the definition:

$$\mathrm{SZ}_{n+1} = [(n + 1) + 1,\ 2(n + 1) - 1] = [n + 2,\ 2n + 1].$$

The width is

$$|\mathrm{SZ}_n| = (2n - 1) - (n + 1) + 1 = n - 1.$$

□

Remark 12 (Maximality)*.* The interval $[n+1, 2n-1]$ is maximal with the property that it is untouched by the current filter F_n and all future filters. If the right endpoint were increased to $2n$, then $2n$ would be hit immediately by the stride n.

Remark 13 (Location of the Stability Zone)*.* The Stability Zone grows linearly:

$$|\mathrm{SZ}_n| = n - 1 \to \infty.$$

By contrast, the tape width grows as the primorial $\prod_{p \leq n} p$. Hence

$$\frac{|\mathrm{SZ}_n|}{|\mathrm{CP}_n|} \to 0.$$

Thus the Stability Zone occupies a vanishing fraction of the full periodic tape, even though its absolute size tends to infinity.

6.4. **Confirmed Primes in the Stability Zone.** The Stability Zone is not merely stable. It is also arithmetically decisive.

Corollary 2 (Confirmed primes in the Stability Zone)**.** *Let* $z \in \mathrm{SZ}_n = [n+1, 2n-1]$*. Then the tape cell canonically corresponding to* z *carries the symbol* L *at step* n *if and only if* z *is prime.*

Proof. If the corresponding cell carries L at step n, then by Theorem 2 that symbol survives until z reaches the head of the tape. By Theorem 1, this means that z is prime.

Conversely, if z is prime, then by Theorem 1 its corresponding tape cell must remain L until it reaches the head. In particular, it is L at step n. □

Remark 14. Thus the Stability Zone provides a moving interval of confirmed primality on the tape: inside SZ_n, the symbols L are no longer merely candidates, but actual primes. This can be used for experiments.

6.5. **The Q-Zone.** The Stability Zone has a purely dynamical definition. It is useful to supplement it with a second interval that is arithmetically safe for the classical sieve reason.

Definition 4 (Q-Zone)**.** At step n, define

$$Q_n = [2n,\, n^2].$$

Lemma 6 (Primality in the Q-Zone)**.** *For every step* $n \geq 2$ *and every integer* $z \in Q_n = [2n, n^2]$*, if the tape cell canonically corresponding to* z *carries the symbol* L *at step* n*, then* z *is prime.*

Proof. Suppose $z \in [2n, n^2]$ and the corresponding tape cell is labeled L at step n. If z were composite, then it would have a prime divisor $p \leq \sqrt{z} \leq n$. The filter F_p would already have acted by step n, so the corresponding cell would have been changed to M, a contradiction. Therefore z is prime. □

Remark 15. The Q-Zone is justified differently from the Stability Zone. In the Stability Zone, primality follows from dynamical immutability. In the Q-Zone, it follows from the classical sieve fact that every composite $z \leq n^2$ has a prime factor $\leq n$. The interval $Q_n = [2n, n^2]$ is therefore best viewed as a second arithmetically safe window, adjacent to and extending beyond the Stability Zone.

6.6. **Twin-Prime Templates: Heuristic Interpretation.** We now connect the symbolic letter

$$a = \langle LMLMMM \rangle$$

with twin-prime patterns.

Lemma 7 (Canonical realization of an a-letter)**.** *Let an occurrence of the letter* $a = \langle LMLMMM \rangle$ *begin at tape position* j *of* CP_n*. Then, in the canonical realization of the tape, this letter corresponds to the six consecutive integers*

$$k,\ k+1,\ k+2,\ k+3,\ k+4,\ k+5 \qquad \text{with } k = n + j.$$

In particular, the two surviving positions in the letter correspond to the candidate pair

$$(k,\ k+2).$$

Proof. By the canonical realization of the tape, the cell $CP_n[j]$ corresponds to the integer $n+j$. Therefore the six cells of the letter correspond to

$$n+j,\ n+j+1,\ \ldots,\ n+j+5.$$

Setting $k = n+j$, the surviving entries of the letter $a = \langle LMLMMM \rangle$ occur at offsets 0 and 2, so the corresponding candidate pair is $(k, k+2)$. □

Corollary 3. *If a letter-a letter has canonical starting position in $Q_n = [2n, n^2]$, then its two surviving positions correspond to an actual twin-prime pair, up to the obvious finite boundary restriction that both surviving positions must still lie inside Q_n.*

Proof. If the starting position of the letter lies in Q_n, then its surviving positions k and $k+2$ also lie in Q_n, except for the trivial case near the right boundary. By Lemma 6, any surviving L-position in that interval is prime. Hence both k and $k+2$ are prime. □

Remark 16 (Heuristic density estimate). Let p_k be the k-th prime. At prime level p_k, the global number of letter-a letters on the periodic tape is

$$G_k(a) = \prod_{5 \le p_i \le p_k} (p_i - 2).$$

Since the tape has period $p_k^{\#}$, the quotient

$$\frac{G_k(a)}{p_k^{\#}}$$

is the global density of letter-a letters per period. This suggests the heuristic estimate

$$\mathcal{A}(p_k) \approx G_k(a) \frac{p_k^2}{p_k^{\#}}$$

for the number of letter-a letters whose canonical starting positions fall in an interval of length p_k^2.

This is only a density heuristic. It does not prove that any specific interval, including Q_n, must contain such a letter.

Remark 17 (Heuristic significance of the Q-Zone). The Q-Zone is arithmetically natural because any surviving L-symbol in Q_n already represents a prime. Thus an a-letter in Q_n is not merely a symbolic twin-prime template; it corresponds to an actual twin-prime pair.

The unresolved issue is therefore not how to interpret such a letter once it lies in Q_n, but whether such letters occur there for arbitrarily large n.

6.7. An Open Distribution Problem.

The Hydra Effect shows that the total population of letter-a letters on the periodic tape remains positive and grows globally. What remains open is their distribution relative to canonically anchored safe windows such as Q_n.

Definition 5 (Permanent multi-scale skew)**.** We say that the distribution of letter-a exhibits a *permanent multi-scale skew away from* Q_n if there exists a finite step t such that for all $n \ge t$, no letter-a letter has its canonical starting position in

$$Q_n = [2n, n^2].$$

Question 1 (Open distribution problem for twin-prime templates). Do the canonical starting positions of letter-a letters intersect the windows

$$Q_n = [2n, n^2]$$

for arbitrarily large n? Equivalently, does the symbolic template $a = \langle LMLMMM \rangle$ avoid a permanent multi-scale skew away from Q_n?

Remark 18 (Heuristic interpretation). Within the automaton, the Twin Prime Conjecture can be viewed heuristically as a distribution problem for the symbolic template

$$a = \langle LMLMMM \rangle.$$

The automaton proves global persistence of this template on the periodic tape, but not its occurrence in the specific safe windows Q_n for arbitrarily large n.

The present paper therefore does not claim a proof of the Twin Prime Conjecture. Instead, it isolates a symbolic-dynamical question whose positive resolution would connect the automaton directly to infinitely many twin primes. The nested structure of CP together with the Hydra Effect (by CRT) leads to maximal distibuted elimination of L to M. The nested periodical structure of CP itself leads to a maximal distribution of letter a by Expansion.

6.8. **Visual Representation.** The advancement of the Stability Zone is illustrated in Table 2.

Step n	$\mathrm{SZ}_n = [n+1, 2n-1]$	Confirmed primes in SZ_n
2	$[3, 3]$	3
3	$[4, 5]$	5
4	$[5, 7]$	5, 7
5	$[6, 9]$	7
6	$[7, 11]$	7, 11
7	$[8, 13]$	11, 13
8	$[9, 15]$	11, 13
9	$[10, 17]$	11, 13, 17
10	$[11, 19]$	11, 13, 17, 19
11	$[12, 21]$	13, 17, 19

TABLE 2. The Stability Zone $\mathrm{SZ}_n = [n+1, 2n-1]$ and the confirmed primes it contains in the first several steps. By Corollary 2, every L-symbol in SZ_n corresponds to a confirmed prime.

7. THE FROZEN WINDOW EXPERIMENT

7.1. **Motivation and Computational Challenge.** The theoretical results established so far describe the symbolic structure of the automaton exactly. In particular, the tape admits a stable four-letter substructure, the Stability Zone provides a moving interval of confirmed primes, and the population of the letter type a can be tracked globally through the Hydra recurrence. These results suggest a natural computational question: how far can the symbolic structure be observed directly?

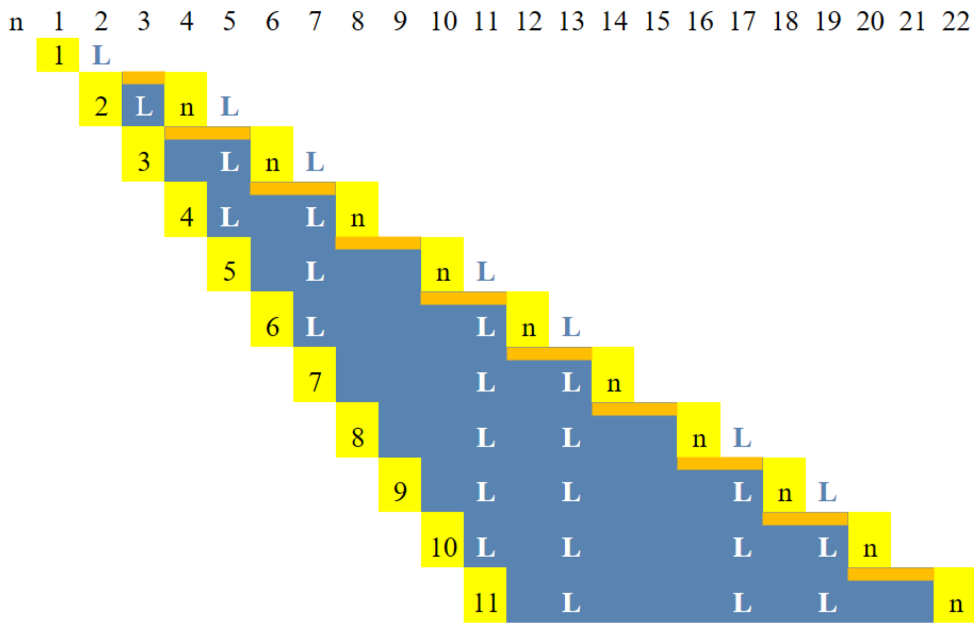

FIGURE 7. The Stability Zone shown in blue, its forward advance in orange, and the first stride of the current filter in yellow. The figure visualizes why the interval $[n+1, 2n-1]$ remains untouched by the current filtering step and all future filtering steps.

A naive simulation of the full automaton quickly becomes infeasible. By Lemma 2, the width of the tape at step n is

$$|\mathrm{CP}_n| = \prod_{p \le n} p.$$

Equivalently,

$$\log |\mathrm{CP}_n| = \theta(n),$$

so the tape width grows on the order of e^n. Already for moderate values of n, the full periodic tape is far too large to store explicitly. Thus a direct computation of CP_n is not a realistic way to study the symbolic structure at large scales.

The key observation is that the experimental questions considered in this paper concern the Stability Zone rather than the full tape. Since the Stability Zone has width $n-1$, one does not need to store the entire periodic tape in order to observe its local structure there. This motivates the Frozen Window technique.

7.2. **The Frozen Window Technique.** To study the Stability Zone up to a prescribed range of n, it is enough to maintain a finite initial segment of the tape large enough to contain the zone throughout the experiment.

For our implementation up to $n = 250{,}000$, a window of width

$$W = 510{,}510 = 17\#$$

is sufficient, since

$$510{,}510 > 2 \cdot 250{,}000.$$

Thus one full period of the tape at step $n = 18$ already covers the range needed to observe the Stability Zone for all subsequent steps up to the chosen bound.

Definition 6 (Frozen Window)**.** The *Frozen Window* is a fixed finite segment of width

$$W = 510{,}510,$$

initialized as an exact copy of CP_{18}. For all later steps, the experiment updates this finite window rather than the full tape.

The experiment runs in two modes.

Mode 1 (Open Window). For $n < 19$, the automaton is run exactly, including shift, expansion, and filtering, until the tape CP_{18} has been constructed.

Mode 2 (Frozen Window). For $19 \le n \le 250{,}000$, the full expansion operator is suppressed, and the usual cyclic shift is replaced by a truncated shift that removes the first symbol without appending it at the right edge. The filter continues to act on the remaining finite window exactly as before.

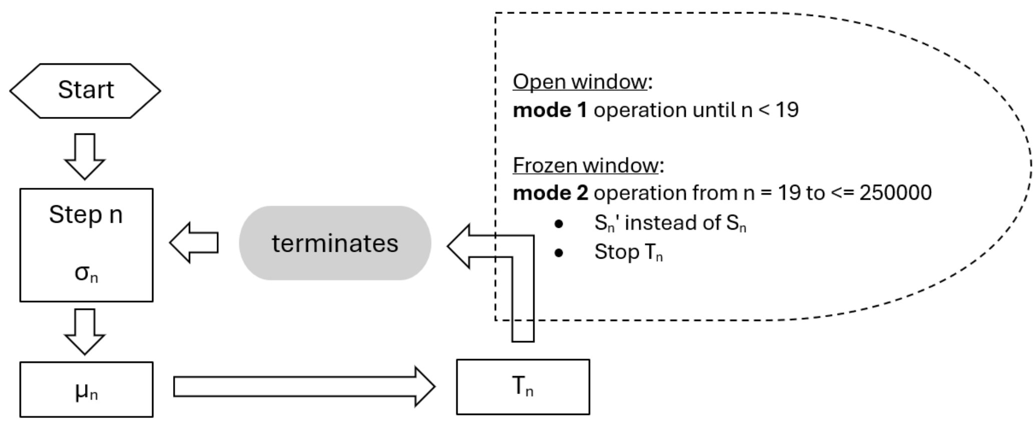

FIGURE 8. The Frozen Window experiment. In Mode 1, the full automaton is run until the tape CP_{18} of width $17\# = 510{,}510$ has been constructed. In Mode 2, the full expansion is suppressed, the shift is replaced by a truncated left shift, and the filter continues to act on the fixed window. This makes it possible to track the Stability Zone up to $n = 250{,}000$ without storing the full periodic tape.

The purpose of this procedure is not to reproduce the entire automaton, but to preserve exactly the part of the symbolic state needed to study the Stability Zone.

7.3. **Validity of the Frozen Window.** The Frozen Window technique is valid because it combines two structural facts proved earlier.

First, the tape is periodic, and one full period contains the complete symbolic information of the sieve state at that level.

Second, the Stability Zone is immune to all later filtering steps once its symbols have reached that region. Therefore, for the purpose of observing the symbolic structure inside the Stability Zone, it is enough to retain a finite segment that already contains all relevant future positions.

These observations justify the following claim.

Proposition 1 (Validity of the Frozen Window)**.** *For all steps $n \leq 250{,}000$, the Frozen Window experiment reproduces the same symbolic pattern inside the Stability Zone as the full automaton.*

Proof. The window is initialized as an exact copy of CP_{18}, which contains one full period of the tape at that stage. Since $W = 510{,}510 > 2n$ for all $n \leq 250{,}000$, the Stability Zone $\mathrm{SZ}_n = [n+1, 2n-1]$ always lies within the retained segment.

By Theorem 2, the symbols in SZ_n are not altered by any later filter before they reach the head of the tape. Therefore, to recover the symbolic pattern in SZ_n, it is enough to preserve the correct symbols in the relevant finite prefix of the tape. The truncated shift moves this preserved segment forward exactly as needed, while the filter continues to act on the window at the same positions as in the full automaton whenever those positions lie inside the retained segment.

Hence the symbolic pattern observed in SZ_n agrees with that of the full automaton throughout the experimental range. □

7.4. **Termination and Determinism.** Both the full automaton and the Frozen Window experiment are deterministic. At each step, the state update is uniquely determined by the current state and the fixed transition rules.

Termination is also immediate. Each step applies only finitely many operations to finite data structures: the counter, the encoding register, and either the full finite tape or the finite frozen window. Thus every step terminates after a finite number of symbolic updates.

7.5. **Boundary Conditions for Letter Counts.** The four-letter classification uses letters of length 6. Inside the Stability Zone, the left or right boundary of the zone may cut through such a letter. In that case, the partial letter is excluded from the letter count.

At most one incomplete letter can occur at each boundary, so at most 10 positions are discarded in total. This has a noticeable effect only for small n, where the Stability Zone itself is still narrow. As n grows, the zone has width $n-1$, so the proportion of excluded positions decays like

$$O(1/n).$$

Thus the boundary effect becomes negligible over the main experimental range.

Remark 19. The initial fluctuations of measured quantities in the experiment are therefore best understood as finite-size boundary effects rather than as evidence of qualitatively different symbolic behavior.

7.6. **Implementation.** The experiment was implemented in Java. The program first constructs the full tape up to step 18, stores the period CP_{18}, and then continues in Frozen Window mode up to $n = 250{,}000$.

At each step, the program identifies the current Stability Zone and scans it in letters of length 6, counting the occurrences of the four letter types a, b, c, d. These data are then used to track structural quantities such as

- the number of a-letters in the Stability Zone,
- the relative frequencies of the four letter types,

- the local density of letter a,
- and the experimental scaling quantity associated with the local letter distribution.

The source code is available at

https://github.com/cerebrummi/stabilityzone

7.7. **Results.** The experiment was run for all steps to $n = 250{,}000$. The data support the structural picture developed in the earlier sections.

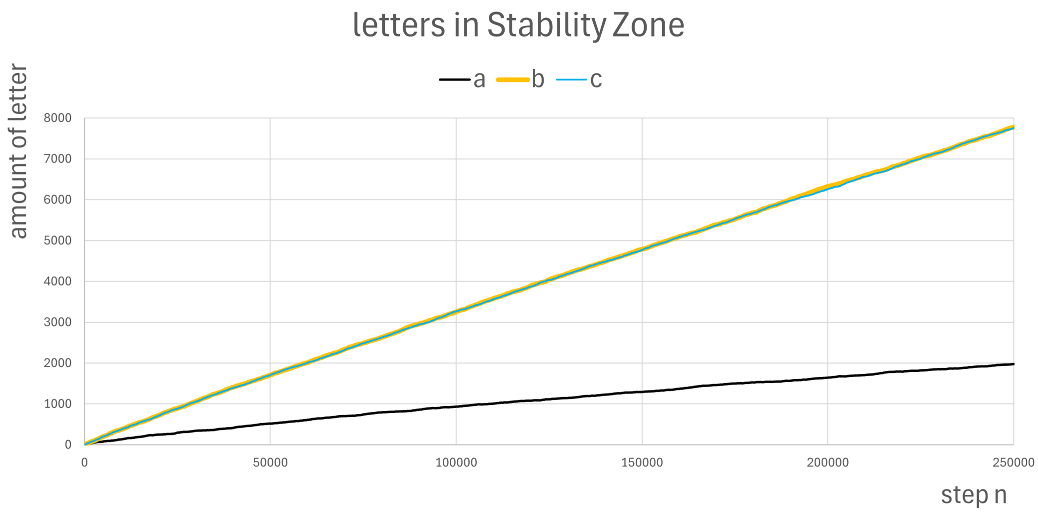

FIGURE 9. Counts of the letter types a, b, and c inside the Stability Zone up to $n = 250{,}000$, measured by the Frozen Window experiment. The persistence of a and the near symmetry of b and c are visible throughout the observed range.

Result 1: Persistence of letter a. The letter type a appears throughout the observed range inside the Stability Zone. Thus the symbolic twin-prime template remains visible locally across all steps tested.

Result 2: Symmetry of b and c. The counts of b and c remain close and exhibit the symmetry predicted by the letter-level transition rules and the structure of the transition matrix.

Result 3: Density of letter a. The density of letter a in the Stability Zone is not monotone. It exhibits small local fluctuations, but the overall trend across the observed range is downward. Thus the local relative frequency of the twin-prime template decreases even though the template itself persists.

Result 4: Experimental scaling behavior. The experimental scaling quantity derived from the letter counts Figure 10 shows finite-size oscillations, especially at smaller n, but this is from edge effects.

7.8. **Limitations and Scope.** The Frozen Window experiment verifies the symbolic structure of the automaton only within the Stability Zone and only up to the chosen range $n = 250{,}000$. It does not simulate the entire periodic tape at large scales, nor does it establish asymptotic results on its own.

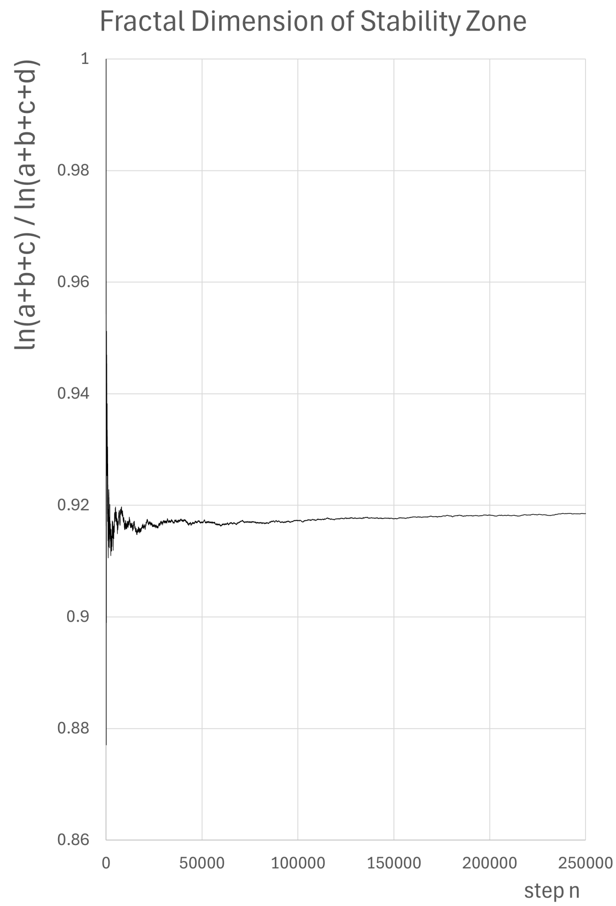

FIGURE 10. Experimental scaling quantity measured in the Stability Zone up to $n = 250{,}000$. The graph shows small finite-scale oscillations together with a broad trend.

Accordingly, the experiment should be understood as a large-scale structural check of the theory rather than as a substitute for proof. Its value lies in confirming that the symbolic phenomena proved earlier remain visible far beyond the range accessible to direct full-tape simulation.

Remark 20. The experimental observation that letter a continues to appear in the Stability Zone up to $n = 250{,}000$ is consistent with the heuristic discussion in subsection 6.6, but it does not prove that such occurrences continue indefinitely.

7.9. **Experimental Interpretation.** The Frozen Window experiment provides finite-scale data on the local symbolic distribution inside the Stability Zone.

The experimental quantities exhibit two important features.

First, they show finite-size oscillations. This is consistent with the boundary effects discussed in subsection 7.5 and with the fact that the experiment measures local letter statistics in a moving finite window rather than a stationary asymptotic object.

Second, despite these local fluctuations, the data are broadly compatible with the interpretation that later prime filters have a weaker local relative effect than earlier ones.

8. Conclusion

8.1. **Summary of Results.** This paper has introduced a deterministic, endogenous, non-stationary S-adic automaton that realizes the Sieve of Eratosthenes as a symbolic dynamical system over a finite alphabet. The automaton evolves by three operators—shift, expansion, and filtering—acting on a binary periodic tape together with two registers.

The main results may be summarized as follows.

Exact realization of the sieve. The automaton reproduces the same prime–composite classification as the classical Sieve of Eratosthenes. The symbolic state of the tape determines the encoding of each integer without any external primality test.

Periodic symbolic state space. At every finite step, the tape is a periodic binary word whose width equals the primorial of the largest prime processed so far. In this way, the sieve is represented not only procedurally, but also as an explicit symbolic state.

Four-letter substructure and letter dynamics. From step $n = 4$ onward, the tape decomposes into a canonical letter alphabet $\Sigma_2 = \{a, b, c, d\}$. This second-order description reveals an internal combinatorial structure that is invisible in the usual formulation of the sieve. Its prime-dependent substitution rules are encoded by an upper-triangular transition matrix M_p.

Hydra Effect. The letter type

$$a = \langle LMLMMM \rangle$$

plays a special role as the symbolic template associated with candidate pairs at distance 2. Its global population evolves by a simple multiplicative rule, which we called the Hydra Effect.

Stability Zone and Q-Zone. The Stability Zone

$$\mathrm{SZ}_n = [n + 1, 2n - 1]$$

provides a dynamically protected interval in which surviving L-symbols are provably stable until they reach the head of the tape. The larger safe region

$$G_n = [n + 1, n^2]$$

and its subinterval

$$Q_n = [2n, n^2]$$

provide an arithmetically natural setting for interpreting the symbolic twin-prime template.

Frozen Window experiment. The Frozen Window technique makes it possible to study the symbolic structure of the Stability Zone far beyond the range accessible to direct simulation of the full periodic tape. The experiment up to $n = 250{,}000$ confirms that the structural phenomena proved earlier remain visible at large finite scales.

8.2. **The Automaton as a Research Instrument.** The automaton is not proposed as a faster way to generate primes. Its value lies elsewhere: it turns the sieve into an explicit symbolic object whose internal organization can be studied directly.

In the classical formulation, the sieve is usually presented as a procedure for eliminating composites. In the present framework, it also becomes a structured symbolic state space. This makes it possible to analyze the sieve at several levels simultaneously: binary, letter-combinatorial, spectral, and experimental.

From this perspective, the main contribution of the paper is conceptual as well as technical. It shows that the sieve admits an exact endogenous symbolic realization and that this realization carries nontrivial internal structure.

8.3. **An Open Distribution Problem.** The paper does not prove the Twin Prime Conjecture. What it does provide is a precise symbolic framework in which one aspect of that problem can be reformulated.

The global persistence of the symbolic template

$$a = \langle LMLMMM \rangle$$

is proved by the internal letter dynamics of the automaton. What remains open is a distribution question: do the canonical starting positions of these letter-a letters intersect the safe windows

$$SZ_n = [n+1, 2n-1]$$
$$Q_n = [2n, n^2]$$

for arbitrarily large n, or can a permanent multi-scale skew away from these windows occur?

This is the main open problem that emerges naturally from the present framework. A positive answer would connect the symbolic dynamics developed here directly to infinitely many twin primes. The current paper isolates this question but does not resolve it.

Other natural directions for future work include a more systematic higher-order symbolic renormalization of the tape, a fuller formal embedding of the automaton into the theory of non-stationary S-adic systems, and larger-scale computational experiments beyond the present Frozen Window range.

8.4. **Code and Data Availability.** The Java implementation of the automaton, the Frozen Window experiment, and the associated output data are publicly available at

https://github.com/cerebrummi/stabilityzone

8.5. **Closing Remark.** The Sieve of Eratosthenes is one of the oldest and most familiar constructions in mathematics. In its classical form, it is an algorithm. In the present paper, it has also been treated as a symbolic dynamical system with an explicit evolving internal state.

This change of viewpoint does not alter which numbers are prime. What it changes is what becomes visible while the sieve unfolds. The automaton records the sieve as a structured symbolic process, and in doing so it opens a different way of studying familiar arithmetic phenomena.

The purpose of the framework is therefore not to replace classical number theory, but to complement it with a symbolic model in which the internal organization of the sieve can be analyzed directly.

Declaration of AI-Assisted Tools. During the preparation of this manuscript, the author used several AI-assisted tools, including ChatGPT, Claude, Gemini, Grok, Copilot, Mistral Le Chat, and Perplexity, for language feedback, wording suggestions, structural discussion, and general editorial support. After using these tools, the author reviewed and revised all material as needed and takes full responsibility for the final content of the paper.

The automaton construction, the underlying mathematical ideas, the Java implementation, the experimental results, and the figures are the author's own.

Acknowledgments. The author thanks Prof. Michel L. Lapidus (University of California, Riverside) for encouragement and for suggesting the inclusion of recent results related to Bertrand's postulate. The author also thanks Prof. Carlo Cattani (University of Tuscia, Viterbo, Italy) for helpful comments on mathematical presentation and paper structure.

References

[1] Allouche, J.-P., Shallit, J. *Automatic Sequences: Theory, Applications, Generalizations.* Cambridge University Press, 2003. 6

[2] Apostol, T. M. *Introduction to Analytic Number Theory.* Springer, New York, 1976. 5

[3] Axler, C. Explicit results on primes in short intervals. *Bulletin of the Australian Mathematical Society*, 106(3):448–457, 2022.

[4] Berthé, V., Delecroix, V. Beyond substitutive dynamical systems: S-adic expansions. In *Numeration and Substitution 2012*, RIMS Kôkyûroku Bessatsu, B46, pp. 81–123, 2014. 6

[5] Bilokon, P. A. Symmetric Zeros in the Imbalance Sequence via Möbius Obstructions. SSRN Working Paper, August 14, 2025. Available at https://ssrn.com/abstract=5391465, DOI: https://doi.org/10.2139/ssrn.5391465. 6

[6] Brun, V. La série $\frac{1}{5}+\frac{1}{7}+\frac{1}{11}+\frac{1}{13}+\cdots$ où les dénominateurs sont "nombres premiers jumeaux" est convergente ou finie. *Bulletin des Sciences Mathématiques*, 43:100–104 and 124–128, 1919.

[7] Chebyshev, P. L. Mémoire sur les nombres premiers. *Journal de mathématiques pures et appliquées*, 17:366–390, 1852.

[8] Cramér, H. On the order of magnitude of the difference between consecutive prime numbers. *Acta Arithmetica*, 2:23–46, 1936. 3, 5

[9] de la Vallée Poussin, C. Recherches analytiques sur la théorie des nombres premiers. *Annales de la Société Scientifique de Bruxelles*, 1896. 5

[10] Dickson, L. E. *History of the Theory of Numbers, Vol. I: Divisibility and Primality.* Carnegie Institution of Washington, 1919. 5

[11] Erdős, P. Beweis eines Satzes von Tschebyschef. *Acta Sci. Math. (Szeged)*, 5:194–198, 1932.

[12] Fogg, N. P. *Substitutions in Dynamics, Arithmetics and Combinatorics.* Lecture Notes in Mathematics, Vol. 1794. Springer, Berlin, 2002. 6

[13] Hadamard, J. Sur la distribution des zéros de la fonction $\zeta(s)$. *Bulletin de la Société Mathématique de France*, 1896. 5

[14] Halberstam, H., Richert, H.-E. *Sieve Methods.* Academic Press, London, 1974. 5

[15] Hardy, G. H., Littlewood, J. E. Some problems of Partitio Numerorum III. *Acta Mathematica*, 44:1–70, 1923. 3, 6

[16] Lapidus, M. L., van Frankenhuijsen, M. *Fractal Geometry, Complex Dimensions and Zeta Functions.* 2nd ed. Springer, New York, 2013. 6

[17] Li, R. Primes in almost all short intervals. arXiv:2407.05651, 2024.

[18] Lind, D., Marcus, B. *An Introduction to Symbolic Dynamics and Coding.* Cambridge University Press, 1995. 6

[19] Maier, H. Primes in short intervals. *Michigan Mathematical Journal*, 32(2):221–225, 1985. 5

[20] Mertens, F. Ein Beitrag zur analytischen Zahlentheorie. *Journal für die reine und angewandte Mathematik*, 78:46–62, 1874.

[21] Nagura, J. On the interval containing at least one prime number. *Proceedings of the Japan Academy*, 28(4):177–181, 1952.
[22] Peano, G. *Arithmetices principia, nova methodo exposita*. Bocca, Turin, 1889. 8
[23] Ramanujan, S. A proof of Bertrand's postulate. *Journal of the Indian Mathematical Society*, 11:181–182, 1919.
[24] Wolfram, S. *A New Kind of Science*. Wolfram Media, 2002.

University of Rostock, Germany
Email address: birke.heeren@uni-rostock.de
URL: https://orcid.org/0000-0001-6713-583X