

Finite-count chemical inheritance with reversible interactions and bounded food supplies

14 September 2026

Abstract

We construct a finite stochastic chemical system that transmits four labels, encoded in the normalized composition of its resident molecules, through a complete growth-and-division cycle with an explicit uniform probability. Two modules interact through reversible cross-catalytic channels with strictly positive rates in both directions. Each module conserves its resident-plus-food molecule count during a closed reaction batch; the only external operations are a scheduled fair split of every molecule and a label-blind food replenishment that reads module totals. With nine modelled molecules per module, every admitted newborn contains at most sixteen residents and each compartment holds eighteen residents plus food. Uniformly over all admitted states and all four labels, the probability that both complementary daughters return to the admitted region of the same label by time twenty is at least $8046297159/8112104000 > 0.99188782$; with twenty molecules per module the bound is 0.99998459 . The guarantee stays above 0.99025694 on an explicit three-parameter box around the nominal rates. The proof uses an empty-daughter observable that equals the fair-partition failure probability exactly, together with a finite-state generator inequality, and its constants are within 2.5×10^{-4} of the numerically evaluated one-cycle probability. Within the stated pure-species, two-module, fair-allocation architecture, eighteen peak residents are necessary and sufficient for a 99% guarantee. We also prove a sharp normalized-composition separation of $2/K$, two-sided bounds for a selected lineage and for a complete binary family of descendants, and an exact product-form stationary law. The uniform probability bound, the identification of the count chain with the literal twenty-four-channel chemistry, and the composition geometry are verified in Lean 4; the structural consequences are proved conventionally and are identified as such. The labels occupy disjoint species supports, the mechanism is an amplification-and-partition guarantee rather than multistability, and it has no defence against contamination by an absent species.

1 Introduction

1.1 The question

Can a small finite population of molecules transmit several distinguishable compositions through growth and division with a uniform, explicit probability? Compositional heredity has been studied as a possible precursor of genetic heredity in mutually catalytic assemblies [19, 24], in compartmentalized autocatalytic networks under growth and division [14, 21], and in the general question of how many heritable states a chemistry can hold [25]. Most of that work asks whether inheritance is possible in a deterministic or large-population limit. The question here is quantitative and finite: a stated number of molecules, a stated cycle, a stated probability, and a proof that does not pass through simulation or an asymptotic approximation.

A decoder alone does not answer such a question. A daughter can read as the correct label and still lie outside the set of states on which the next cycle’s guarantee applies. We

therefore treat copying as a complete cycle. A label σ is represented by an admitted region $\mathcal{C}_\sigma$ of restart states. One cycle succeeds when an admitted parent runs its reaction batch to the scheduled deadline, its molecules are allocated to two daughters by the protocol’s actual complementary partition law, and both daughters, after the declared label-blind replenishment, lie again in $\mathcal{C}_\sigma$. Every unsuccessful outcome stays in the probability denominator. The reward for the stronger definition is that repeated copying, along a lineage or across a whole family of descendants, follows by conditioning alone and needs no independence assumption between siblings or generations.

1.2 Main results

The construction has four resident species X_0, Y_0, X_1, Y_1 and two food species F_0, F_1 , arranged in two modules. Bit i of a word $\sigma \in \{0, 1\}^2$ is stored by which of X_i, Y_i is present in module i . Each module conserves its resident-plus-food count K during a batch. All twelve reversible reaction pairs have strictly positive rates in both directions, and the two modules interact through cross-catalysis: the residents of one module catalyse growth and reversal in the other. Section 2 gives the reaction table and the complete cycle.

- (i) *Uniform finite-time copying* (Theorem 3.4). For $K = 9$ and nominal rates $(a, b, \gamma) = (1000, 1, 1/10)$, every admitted state of every word returns both daughters to the same admitted region by time $T = 20$ with probability at least

$$q_9 = \frac{8046297159}{8112104000} > 0.99188782.$$

Admitted newborns carry at most 16 residents and a compartment never holds more than 18 residents plus food. For $K = 20$ the bound is $q_{20} > 0.99998459$ with 38 newborn and 40 peak residents.

- (ii) *Tightness* (Section 3.4). The numerically evaluated one-cycle probability at $K = 9$ is 0.99213, the proved bound is 0.99189, and the fair-partition ceiling that no dynamics can beat is 0.99220. The proof loses 2.5×10^{-4} and the construction is within 7×10^{-5} of the ceiling.
- (iii) *An explicit operating region* (Theorem 4.1). On $500 \leq a \leq 1500$, $\frac{1}{2} \leq b \leq 2$, $\frac{1}{20} \leq \gamma \leq \frac{2}{9}$ the uniform bound is $q_{\text{box}} > 0.99025694$.
- (iv) *A sharp budget* (Proposition 5.1). In any two-module pure-species architecture that splits individual resident molecules fairly, demands an immediate representative of each module in both daughters and permits no repair, a parent with at most 17 residents cannot achieve 99%; the construction achieves it with 18.
- (v) *Composition geometry* (Section 6). The four admitted regions have pairwise L^1 separation exactly $2/K$ in normalized resident composition, a nearest-region decoder tolerates measurement error below $1/K$, and a fixed decoder reads the word from which species has positive fraction.
- (vi) *Repeated copying* (Section 7). The actual successful-restart kernel is constructed explicitly. A selected lineage survives N cycles with probability between q_K^N and u_K^N , where $u_K = (1 - 2^{1-K})^2$, and a complete binary family of depth N survives with probability at least $1 - (2^N - 1)(1 - q_K)$. No lineage survives forever.
- (vii) *Stationary law and symmetry* (Section 8). Despite genuine interaction along sample paths, the count chain satisfies detailed balance and its unique stationary law is the product of two truncated binomial laws. The four words have identical count dynamics.

1.3 What the mechanism is, and is not

Readers familiar with compositional genomes [19] or bistable switches should note that inheritance here is not maintained by attractors in a shared composition space. The reverse autocatalytic step $2S \rightarrow S + F$ has propensity proportional to $n_S(n_S - 1)$, so it cannot remove a species’ last molecule, and no channel of the reaction table creates a species that is absent. The label is therefore an exact invariant of the batch dynamics. What the batch has to do is amplify each module’s count from as little as one molecule toward K fast enough that a fair random split is unlikely to leave either daughter without a representative. The result is an *amplification-and-partition* guarantee. Its strength is that no accurate transmission of a small fluctuating minority population is required, as a mixed-composition encoding would require. Its weakness, proved in Section 9, is the mirror image: a single molecule of an absent species can never be eliminated, so the mechanism tolerates dilution and partition losses but not contamination or mutation. In the vocabulary of Szathmáry [22] it is a limited-heredity replicator under an external division protocol, not an error-correcting one, and it does not address the error threshold of Eigen [6].

The two modules interact: for $\gamma > 0$ every enabled reaction in one module has a rate that depends on the count in the other. The interaction is visible in every sample path and in the transient law, even though the stationary count law factorizes (Section 8). All external operations, the scheduled split, the food refill and the volume reset, read only module totals and never the encoded species identity.

An earlier construction by the authors, formalized in the same repository but not published, stored the same four labels by keeping food at a maintained activity, quenching each module when it reached twenty residents and imposing an event quota; it reached 99% with 38 newborn residents. The present construction removes the quench, the maintained pool and the quota by conserving a finite food supply inside each batch, halves the budget, and is proved independently.

1.4 Related work

Stochastic mass-action kinetics on integer counts is the setting of Anderson and Kurtz [2], Gillespie [8], Kurtz [13]. Product-form stationary laws for complex-balanced networks are due to Anderson et al. [3]; the stationary law here is obtained by a direct detailed-balance calculation in the sense of Kelly [11], Schnakenberg [17], and the general theory is not needed. The proof technique is a finite-state Foster–Lyapunov drift estimate [15], applied to a deadline rather than to stationarity, in the spirit of certificate methods for stochastic reaction networks [9, 12]; the novelty is the choice of observable, which is the exact fair-partition failure probability and makes the certificate almost tight. Stochastic partitioning at division as a source of daughter heterogeneity is analysed by Huh and Paulsson [10]. The companion manuscript [1] treats copying of concentration states in larger systems, including an 874-molecule reversible source, a molecular-redundancy theorem and a protocol for a published bistable chemistry [20]; it shares the copying semantics and the finite-state verification infrastructure used here, but none of its probability theorems is used in the present proofs.

1.5 Methods and verification

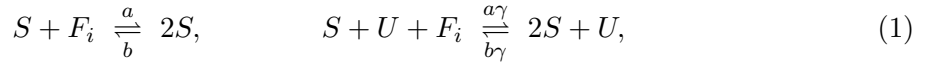
Every probability in this paper is an exact rational number produced by a generator inequality on a finite continuous-time Markov chain, an exact binomial partition identity and the elementary inequality $e^x \geq 1 + x$. No simulation, floating-point evaluation or truncation enters a proof. The uniform bound of Theorem 3.4, the operating-region bound, the identification of the count chain with the literal twenty-four-channel chemistry, the literal stoichiometry of every positive-rate transition, the refill arithmetic and the composition geometry are checked in

Lean 4 with Mathlib [5, 23]; Appendix A lists the declarations and the receipt. The sharp budget, the restart kernel and its iteration, the family bound, the stationary law and the contamination obstruction are conventional proofs given in full in the text. The numerical diagnostics of Section 3.4 and Figure 2 are evaluations, not certificates, and are labelled as such. The script `check_paper.py` distributed with the sources replays every constant printed below in exact arithmetic and produces the figures.

2 The finite-food source and the complete cycle

2.1 Species and reaction table

The resident species are X_0, Y_0, X_1, Y_1 and the food species are F_0, F_1 ; index $i \in \{0, 1\}$ is the module. For $i \in \{0, 1\}$, $j = 1 - i$, $S \in \{X_i, Y_i\}$ and $U \in \{X_j, Y_j\}$ the reaction table contains the reversible pairs



that is, four uncatalyzed autocatalytic pairs and eight cross-catalyzed pairs, twelve reversible pairs and twenty-four directional channels in all. The nominal coefficients are $a = 1000$, $b = 1$, $\gamma = 1/10$; every directional coefficient is strictly positive. Propensities follow the stochastic mass-action convention at a fixed normalized volume: the forward channels have propensity $a n_S f_i$ and $a\gamma n_S n_U f_i$, the reverse channels $b n_S (n_S - 1)$ and $b\gamma n_S (n_S - 1) n_U$, where n_S is the count of S and f_i the count of F_i . Rate constants absorb any combinatorial convention factor for the pair of identical reactants. The table is the same for every word: neither rates nor external operations receive σ .

The catalyst U belongs to the other module and appears unchanged on both sides, so cross-catalysis multiplies the rate of an existing channel without changing its stoichiometry. Two features of the table carry the whole argument. First, no channel produces a species that is absent: every product species also occurs as a reactant. Second, the reverse channels have propensity proportional to $n_S(n_S - 1)$, which vanishes at $n_S = 1$; the last molecule of a present species can never be consumed.

2.2 Conservation, pure faces and the count chain

Every channel that changes the count of a resident of module i changes the count of F_i by the opposite amount, so

$$n_{X_i} + n_{Y_i} + f_i = K_i \quad (2)$$

is conserved during a batch. We take $K_0 = K_1 = K$. A word $\sigma \in \{0, 1\}^2$ selects $S_i = X_i$ if $\sigma_i = 0$ and $S_i = Y_i$ if $\sigma_i = 1$. The *pure face* of σ is the set of states in which module i contains only S_i among its residents, with count $n_i \in \{1, \dots, K\}$ and food $f_i = K - n_i$. By the two features above the pure face is invariant under every positive-rate transition, and on it the twenty-four channels aggregate to a birth-and-death chain in each module: summing the uncatalyzed channel and the two possible neighbour-catalyzed channels (only one of which has a positive rate on the face),

$$B_i(n_i, n_j) = a n_i (K - n_i) (1 + \gamma n_j), \quad (3)$$

$$D_i(n_i, n_j) = b n_i (n_i - 1) (1 + \gamma n_j). \quad (4)$$

Each positive-rate jump has the literal stoichiometry of (1): one resident added and one food removed, or the reverse, with catalysts unchanged. At $n_i = K$ only births vanish, and reverse reactions regenerate food; at $n_i = 1$ deaths vanish. The selected count process

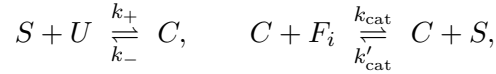
$n(t) = (n_0(t), n_1(t))$ is therefore a finite continuous-time Markov chain on $\{1, \dots, K\}^2$ with rates (3)–(4). It is non-explosive and needs no event quota. For $\gamma > 0$ the rates of module i depend on n_j whenever the corresponding reaction is enabled, so the modules interact.

The Lean development represents the literal chemistry as a jump model with 24 channels indexed by target module and direction, species identity and catalytic option, and proves that for every word its generator on the pure face equals the generator of the aggregated chain, so that the two finite-time laws coincide (`channel_sum`, `generator_identification`, `word_law`); the stoichiometry of every positive-rate channel is checked separately (`literal_next`, `food_transition`).

2.3 Thermodynamic consistency

All twelve reversible pairs have the common forward/reverse ratio a/b , whether catalyzed or not, as a catalyst must leave the equilibrium constant unchanged. Assigning activity 1 to each food species and activity $R = a/b$ to each resident species balances every pair, and unit resident and food masses balance every reaction. On the count chain this appears as detailed balance with respect to the product law of Section 8. At unit activities the free-energy change of converting one food molecule into one resident is $-k_B T \ln(a/b)$, which is about $-6.9 k_B T$ at the nominal ratio 1000 and lies between $-5.5 k_B T$ and $-8.0 k_B T$ on the operating box of Theorem 4.1. That free energy is supplied by the food that the controller replenishes after each split. These are consistency statements about the reaction table; they do not quantify the work of the split, the refill or the volume reset.

Remark 2.1 (Bimolecular realizations of the catalytic channel). The channel $S + U + F_i \rightarrow 2S + U$ is termolecular as written. A standard elementary realization is a binding step and a catalytic step,



where the complex C carries one S and one U . Under the quasi-steady-state reduction for the complex [4, 16, 18], $n_C \approx (k_+/k_-) n_S n_U$, and the effective catalyzed propensities are $k_{\text{cat}}(k_+/k_-) n_S n_U f_i$ and $k'_{\text{cat}}(k_+/k_-) n_S (n_S - 1) n_U$, so $a\gamma = k_{\text{cat}} k_+/k_-$ and $b\gamma = k'_{\text{cat}} k_+/k_-$ with $k_{\text{cat}}/k'_{\text{cat}} = a/b$ preserving the common ratio. Two structural properties survive the enlargement of the state space. The complex dissociates only into the species that formed it, so the label remains an exact invariant and the conserved quantity becomes (free S) + (bound S) + $F_i = K$. And the total number of S -bearing molecules, free or bound, is unchanged by binding and unbinding, so the empty-daughter observable of Section 3, evaluated on that total, is still the exact fair-partition failure probability provided the complex is split as one molecule. What does not transfer automatically is the local generator bound (Lemma 3.3), because birth propensities then depend on how the S count is divided between free and bound molecules. Re-establishing it on the enlarged chain is a separate task, and the theorems below are stated for the table (1) only.

2.4 The complete cycle

Definition 2.2 (Admitted regions). For $K \geq 2$ and a word σ , the admitted restart region $\mathcal{C}_\sigma$ consists of the pure-face states of σ with $1 \leq n_i \leq K - 1$ in both modules, food counts $f_i = K - n_i$, the fixed solvent and volume state, and a fresh cycle clock. Each $\mathcal{C}_\sigma$ is a nonempty set of integer states; $\mathcal{C}_\sigma \cap \mathcal{C}_\tau = \emptyset$ for $\sigma \neq \tau$ because the supports differ.

One cycle from a state $z \in \mathcal{C}_\sigma$ consists of four steps (Figure 1).

1. *Closed reaction batch.* The chemistry (1) runs for the scheduled time $T = 20$ at fixed volume with no exchange of matter; (2) holds throughout.

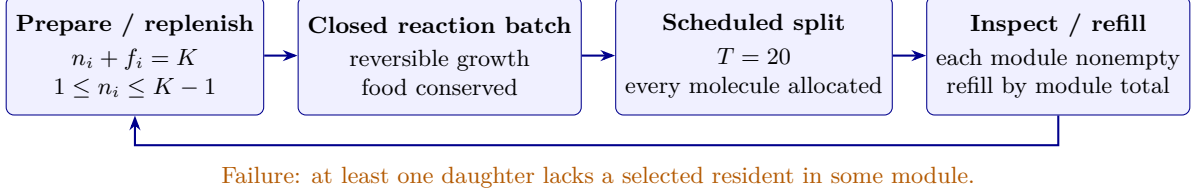


Figure 1: One finite-food cycle. Each module conserves K modelled molecules during the batch. Refill across both daughters adds exactly K food molecules per module. The controller reads totals, never the encoded species identity.

2. *Scheduled fair split.* At time T every molecule, resident or food, is assigned independently to daughter A with probability $1/2$ and to daughter B otherwise. The two daughters are complementary, not independent.
3. *Inspection.* The cycle is *successful* exactly when each daughter receives at least one resident of each module. If module i of the parent holds n_i residents at time T , the count r_i allocated to A must satisfy $1 \leq r_i \leq n_i - 1$; the complementary count $n_i - r_i$ then also lies in $\{1, \dots, K - 1\}$.
4. *Replenishment and reset.* Let M_i^A, M_i^B be the daughters' resident-plus-food totals in module i . Each daughter receives $K - M_i^A$ (respectively $K - M_i^B$) fresh molecules of F_i and the declared solvent and volume reset. The refill reads totals only and never alters residents.

Because $M_i^A + M_i^B = K$, the food supplied to both daughters in module i is

$$(K - M_i^A) + (K - M_i^B) = K, \quad (5)$$

exactly $2K$ new food molecules per cycle across the two modules, or 18 at $K = 9$. After refill each successful daughter has module totals K , the fixed volume and a fresh clock, so it is an element of $\mathcal{C}_\sigma$ and the same law governs its next cycle. Food allocation imposes no success condition: the binomial law of the food split sums to one and every allocation can be refilled as declared. Partition, refill and reset are instantaneous external maps in this model, so $T = 20$ is the complete-cycle deadline in the chosen time unit; it is not a statement about twenty seconds in a physical implementation. Failed partitions are retained as failure mass and are never conditioned away.

3 Uniform finite-time copying

3.1 The empty-daughter observable

For $n \geq 1$ put

$$g(n) = 2^{1-n}, \quad H(n_0, n_1) = g(n_0) + g(n_1), \quad Q(n_0, n_1) = (1 - g(n_0))(1 - g(n_1)). \quad (6)$$

Lemma 3.1 (Exact partition law). *Let a module hold $n \geq 1$ residents at the split. The probability that daughter A receives none or all of them is exactly $g(n)$. Conditionally on the counts (n_0, n_1) at time T , the probability that both daughters receive at least one resident of each module is exactly $Q(n_0, n_1)$, and $Q \geq 1 - H$.*

Proof. The events that A receives zero or all n molecules of the module are disjoint and each has probability 2^{-n} , so their union has probability $2^{1-n} = g(n)$. Molecules of different modules are allocated independently, so the two module successes are conditionally independent given the

counts and the joint success probability is the product Q . Finally $Q = 1 - H + g(n_0)g(n_1) \geq 1 - H$. No independence of the two count processes is used, only independence of the allocation of distinct molecules. $\square$

The observable H is thus an exact upper bound for the conditional failure probability, and a lower bound on $\mathbb{E}_z Q(n(T))$ follows from an upper bound on $\mathbb{E}_z H(n(T))$. The next lemma is the only piece of probability theory needed.

Lemma 3.2 (Finite-state deadline bound). *Let (Z_t) be a continuous-time Markov chain on a finite state space with generator $\mathcal{L}$, and let $w \geq 0$, $s > 0$ and $\eta \geq 0$ satisfy $\mathcal{L}w \leq -sw + \eta$ pointwise. Then for every start z and $t \geq 0$,*

$$\mathbb{E}_z w(Z_t) \leq e^{-st}w(z) + \frac{\eta}{s}(1 - e^{-st}). \quad (7)$$

Proof. On a finite state space $u(t) = \mathbb{E}_z w(Z_t)$ is differentiable with $u'(t) = \mathbb{E}_z \mathcal{L}w(Z_t) \leq -su(t) + \eta$ (Kolmogorov's backward equation, or Dynkin's formula [7]). Multiplying $u' + su \leq \eta$ by e^{st} and integrating from 0 to t gives (7). $\square$

In the formal development the finite-time law is defined as the matrix exponential $(e^{t\mathcal{L}}h)(z)$ of the rate matrix, and Lemma 3.2 is an instance of a residual certificate that combines a subsolution v with $\mathcal{L}v \geq -\varepsilon$, a decaying witness w and a cover $v \leq h + w$ (`finite_deadline_residual_certificate`); here $v \equiv 1$, $\varepsilon = 0$, $h = Q$ and $w = H$.

3.2 Local generator bounds

Write $\mathcal{L}$ for the generator of the count chain with rates (3)–(4). A birth in module i multiplies $g(n_i)$ by 1/2 and a death multiplies it by 2, so

$$\mathcal{L}g(n_i) = n_i g(n_i) (1 + \gamma n_j) \left[b(n_i - 1) - \frac{a}{2}(K - n_i) \right], \quad (8)$$

and the formula remains valid at $n_i = 1$ and $n_i = K$ because the corresponding rates vanish there.

Lemma 3.3 (Local generator bounds at the nominal rates). *Let $(a, b, \gamma) = (1000, 1, 1/10)$ and $1 \leq n_i, n_j \leq K$.*

$$(a) \text{ For } K = 9: \quad \mathcal{L}g(n_i) \leq -3900 g(n_i) + \frac{2523}{160}.$$

$$(b) \text{ For } K = 20: \quad \mathcal{L}g(n_i) \leq -9000 g(n_i) + \frac{2535}{131072}.$$

Both inequalities hold with equality at $n_i = n_j = K$.

Proof. (a) For $1 \leq n \leq 8$ the bracket in (8) is $-[500(9 - n) - (n - 1)]$ and the function $n \mapsto n[500(9 - n) - (n - 1)]$ is a concave quadratic with endpoint values 4000 at $n = 1$ and 3944 at $n = 8$, hence at least 3944 on the interior. Since $1 + n_j/10 \geq 1$, the interior generator is at most $-3944 g(n) \leq -3900 g(n)$. At $n = 9$ the bracket is 8, $g(9) = 2^{-8}$ and $1 + n_j/10 \leq 19/10$, so

$$\mathcal{L}g(9) + 3900 g(9) \leq \frac{9 \cdot 8 \cdot (19/10) + 3900}{256} = \frac{2523}{160}.$$

(b) For $1 \leq n \leq 19$ the quadratic $n[500(20 - n) - (n - 1)]$ has endpoint values 9500 and 9158, so the interior generator is at most $-9158 g(n) \leq -9000 g(n)$. At $n = 20$ the multiplier is at most $1 + 20/10 = 3$ and $\mathcal{L}g(20) + 9000 g(20) \leq (20 \cdot 19 \cdot 3 + 9000)/2^{19} = 2535/131072$. $\square$

3.3 The theorem

Theorem 3.4 (Finite-food copying). *Let $(a, b, \gamma) = (1000, 1, 1/10)$ and $T = 20$. For $K = 9$, every word σ and every $z \in \mathcal{C}_\sigma$, the complete cycle of Section 2.4 satisfies*

$$\mathbb{P}_z\{Z_A^{\text{restart}} \in \mathcal{C}_\sigma \text{ and } Z_B^{\text{restart}} \in \mathcal{C}_\sigma\} \geq q_9 := \frac{8046297159}{8112104000} > 0.99188782. \quad (9)$$

For $K = 20$ the same statement holds with

$$q_{20} := \frac{7077818258231}{7077927321600} > 0.99998459. \quad (10)$$

Admitted newborns carry at most $2(K - 1)$ residents, and each compartment carries exactly $2K$ residents plus food at every time. Both bounds are uniform in the word and in the admitted state.

Proof. Fix σ and $z \in \mathcal{C}_\sigma$. By the invariance of the pure face and the generator identification, the resident counts at time T are the counts $n(T)$ of the finite chain started at the selected counts of z . By Lemma 3.1 the conditional probability of a successful split is $Q(n(T))$; every successful split places both daughters on the pure face with counts in $\{1, \dots, K - 1\}$, and the refill (5) is always possible and returns both to $\mathcal{C}_\sigma$. Hence the joint-return probability equals $\mathbb{E}_z Q(n(T)) \geq 1 - \mathbb{E}_z H(n(T))$.

Let $\mathcal{L}g \leq -sg + C$ be the local bound of Lemma 3.3. Adding the two modules gives $\mathcal{L}H \leq -sH + 2C$, and Lemma 3.2 with $H(z) \leq 2$ yields

$$\mathbb{E}_z H(n(T)) \leq 2e^{-sT} + \frac{2C}{s}.$$

The inequality $e^x \geq 1 + x$ gives $e^{-20s} \leq (1 + 20s)^{-1}$. For $K = 9$, $(s, C) = (3900, 2523/160)$ and

$$\mathbb{P}_z(\text{joint return}) \geq 1 - \frac{2}{78001} - \frac{841}{104000} = q_9.$$

For $K = 20$, $(s, C) = (9000, 2535/131072)$ and the bound is $1 - 2/180001 - 169/39321600 = q_{20}$. No terminal count is prescribed: a parent with food remaining at time T is included with its actual partition probability. The molecule budgets follow from (2) and the definition of $\mathcal{C}_\sigma$. $\square$

Remark 3.5. The argument uses the deadline only through $e^{-20s} \leq 1/78001$; the bound is dominated by the residual term $2C/s$, which comes entirely from the boundary state $n_i = K$ where births have stopped and only reverse reactions act. Any deadline $T \geq 1/100$ would already give a bound above 0.99 at $K = 9$; the chosen $T = 20$ makes the decaying term negligible and matches the companion protocol.

3.4 How tight is the bound?

The bound has two losses: replacing Q by $1 - H$, which costs $\mathbb{E}g(n_0)g(n_1) \approx 1.6 \times 10^{-5}$ at stationarity, and the Lyapunov residual $2C/s$, which exceeds the stationary value $2\mathbb{E}_\pi g$ by a small amount because the boundary constant C is evaluated at the worst neighbour count. Table 1 compares the proved bounds with floating-point evaluations of the one-cycle probability $\mathbb{E}_z Q(n(T))$ from the matrix exponential of the $K^2 \times K^2$ rate matrix, and with the exact stationary value $(1 - \mathbb{E}_\pi g)^2$ computed from the product law of Section 8. Both are diagnostics, not certificates. At $T = 20$ every admitted start gives the same value to eleven digits, as expected for a chain whose rates are of order 10^3 to 10^4 . The proved bound is within 2.5×10^{-4} of the numerically evaluated probability at $K = 9$ and within 1.2×10^{-5} at $K = 20$; the construction itself is within 7×10^{-5} (respectively 8×10^{-8}) of the fair-partition ceiling $u_K = Q(K, K)$ that no dynamics on $\{1, \dots, K\}^2$ can exceed.

Table 1: Proved bounds against numerical diagnostics. The middle column is $\mathbb{E}_z Q(n(20))$ from the matrix exponential (minimum over admitted starts; the maximum agrees to 10^{-11}); the stationary value is exact rational arithmetic rounded. Only the first column is a theorem.

K	proved q_K	numerical one-cycle return	stationary $(1 - \mathbb{E}_\pi g)^2$	ceiling $u_K = Q(K, K)$
9	0.99188782	0.99213251	0.99213251	0.99220276
20	0.99998459	0.99999611	0.99999611	0.99999619

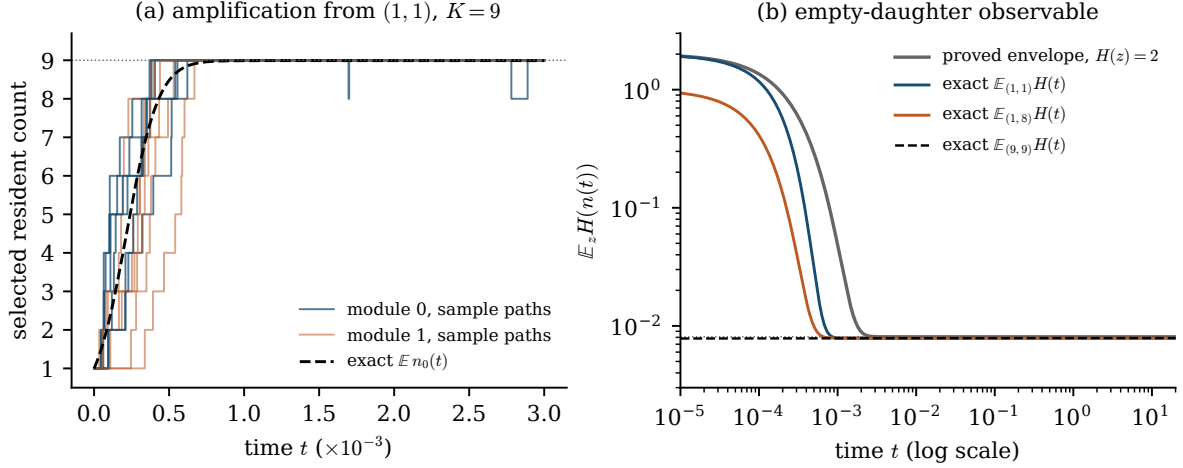


Figure 2: Amplification and the empty-daughter observable at $K = 9$, nominal rates. (a) Six Gillespie sample paths of the selected counts from the worst admitted start $(1, 1)$ on the first 3×10^{-3} time units, with the exact mean count $\mathbb{E} n_0(t)$ from the matrix exponential (dashed). (b) Exact $\mathbb{E}_z H(n(t))$ for the starts $(1, 1)$, $(1, 8)$ and $(9, 9)$ on a logarithmic time axis, against the proved envelope $2e^{-3900t} + (2C/s)(1 - e^{-3900t})$ and its floor $2C/s$ (dotted). Both panels are numerical illustrations of proved statements, not evidence used in a proof.

Figure 2 shows why the estimate is so close. Starting from the worst admitted state $(1, 1)$, amplification to $K = 9$ is complete within a few thousandths of a time unit: the birth rate at $n_i = 1$ is $a(K - 1)(1 + \gamma n_j) \geq 8800$. Thereafter the chain sits at $n_i = K$ with occasional single reverse steps, and $\mathbb{E}_z H$ settles at $2\mathbb{E}_\pi g = 0.00788$, just below the proved floor $2C/s = 0.00809$. Panel (a) shows six Gillespie sample paths [8] together with the exact mean count; panel (b) shows the exact $\mathbb{E}_z H(n(t))$ for three starts against the envelope (7).

4 An explicit operating region

Theorem 4.1 (Operating box). *Let $K = 9$, $T = 20$, and suppose*

$$500 \leq a \leq 1500, \quad \frac{1}{2} \leq b \leq 2, \quad \frac{1}{20} \leq \gamma \leq \frac{2}{9}. \quad (11)$$

Then every admitted state of every word has joint-return probability at least

$$q_{\text{box}} := 1 - \frac{2}{36001} - \frac{31}{3200} = \frac{114080769}{115203200} > 0.99025694. \quad (12)$$

Proof. By (8), $\mathcal{L}g(n) = n g(n)(1 + \gamma m)[b(n - 1) - \frac{a}{2}(9 - n)]$ with m the neighbour count. For $1 \leq n \leq 8$ the bracket is negative and is bounded above by $2(n - 1) - 250(9 - n)$, obtained by replacing a by 500 and b by 2. The concave quadratic $n[250(9 - n) - 2(n - 1)]$ has endpoint values 2000 and 1888, so with $1 + \gamma m \geq 1$ the interior generator is at most $-1888 g(n) \leq -1800 g(n)$.

At $n = 9$ the bracket is $8b \leq 16$ and $1 + \gamma m \leq 1 + \frac{2}{9} \cdot 9 = 3$, so

$$\mathcal{L}g(9) + 1800g(9) \leq \frac{2 \cdot 9 \cdot 8 \cdot 3 + 1800}{256} = \frac{279}{32}.$$

The proof of Theorem 3.4 with $(s, C) = (1800, 279/32)$ gives (12). $\square$

The generator inequality is affine in a and in b for fixed counts and γ , and affine in γ for fixed counts, a and b ; the formal proof treats the three parameters symbolically over the whole box, and the exact replay of the vertex cases in `check_paper.py` is a diagnostic only. The uncertainty covered is within the compatible family (1): catalytic coefficients are $a\gamma$ and $b\gamma$, so the common ratio a/b is preserved, and the box does not cover independently perturbed channel constants, additional reaction channels or contamination. Within the box the equilibrium ratio ranges over $[250, 3000]$.

5 The minimum peak budget

The construction at $K = 9$ never holds more than 18 residents. The following proposition shows that, within the architecture actually used, 18 is also necessary, and that the constraint is combinatorial rather than dynamical: it comes from the fair split, not from the reaction kinetics.

Proposition 5.1 (Restricted sharp budget). *Consider any two-module pure-species architecture whose complete cycle allocates individual resident molecules independently and fairly to two complementary daughters, requires an immediate representative of each module in both daughters, and performs no resident repair. If the parent’s resident total at the split is always at most 17, the joint-return probability is strictly less than 0.99, whatever the preceding dynamics and whatever the law of the parent state. Eighteen peak residents suffice in this class.*

Proof. A parent with an empty module has zero success probability. Otherwise, by Lemma 3.1, the conditional success probability given counts (r, s) is $Q(r, s) = (1 - 2^{1-r})(1 - 2^{1-s})$, which is increasing in each argument. For a fixed total $r + s$ and $s \geq r + 2$, direct subtraction gives

$$Q(r + 1, s - 1) - Q(r, s) = 2^{-r} - 2^{1-s} > 0,$$

so the balanced split maximizes Q at a given total. Monotonicity lets any total below 17 be raised to 17, whose maximum is attained at $(8, 9)$ or $(9, 8)$:

$$Q(r, s) \leq \frac{127}{128} \cdot \frac{255}{256} = \frac{32385}{32768} = 0.988311767578125 < 0.99.$$

Taking the expectation over an arbitrary law of the parent counts preserves this pointwise ceiling. Sufficiency is Theorem 3.4 at $K = 9$, whose chain reaches $(9, 9)$ by a positive-rate path, so the capacity 18 is actually used. $\square$

Figure 3 places the two proved points against the fair-partition failure floor $1 - u_K = 1 - (1 - 2^{1-K})^2$ as a function of the peak capacity $2K$. The claim of Proposition 5.1 is restricted to the stated architecture; a mechanism with unequal or correlated allocation, with several representatives per label, or with repair after birth is outside its scope, as is any statement about a universal minimum molecule count for chemical memory.

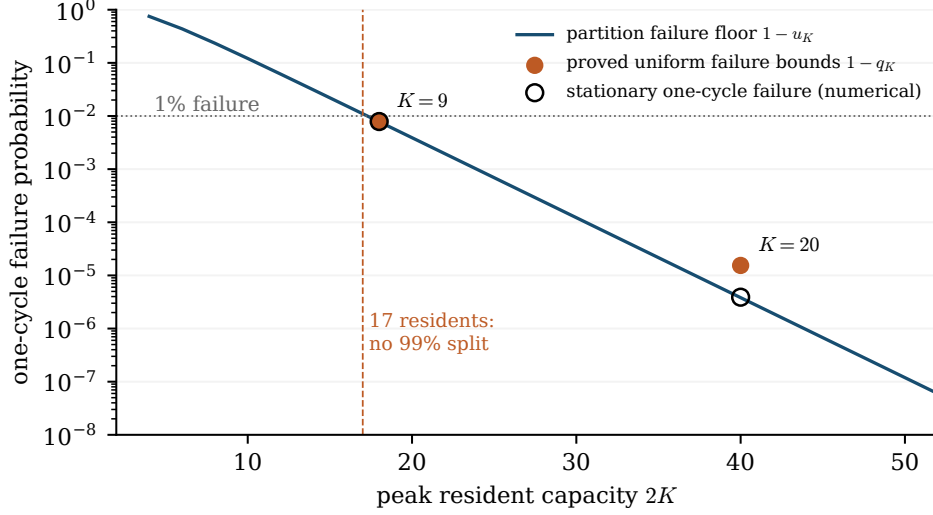


Figure 3: Budget and fidelity. The curve is the exact fair-partition failure floor $1 - u_K$ that no dynamics on $\{1, \dots, K\}^2$ can beat. Filled points are the proved uniform failure bounds $1 - q_K$ for $K = 9, 20$; open circles are the stationary one-cycle failures from Table 1 (numerical). The dashed line marks the seventeen-resident ceiling of Proposition 5.1. The gap between a filled point and the curve is the slack of the proof, not a simulation confidence interval.

6 Composition geometry

Normalize the four resident counts by their total, excluding food, to obtain the composition $p(z) \in \mathbb{R}^4$ of a state z . We show that the four admitted regions are separated in L^1 by exactly $2/K$ and that the readout is stable.

Proposition 6.1 (Sharp composition margin). *For every $K \geq 2$,*

$$\min_{\sigma \neq \tau} \inf_{z \in \mathcal{C}_\sigma, z' \in \mathcal{C}_\tau} \|p(z) - p(z')\|_1 = \frac{2}{K}, \quad (13)$$

and the minimum is attained. A nearest-region decoder returns the correct word from any measured composition within L^1 distance strictly less than $1/K$ of $p(z)$, $z \in \mathcal{C}_\sigma$.

Proof. In an admitted state module i holds at least one resident and the other module at most $K - 1$, so the occupied coordinate of module i satisfies $p_{S_i}(z) \geq 1/K$. If σ and τ differ in bit i , the coordinate occupied by σ in module i is zero in every state of $\mathcal{C}_\tau$, and conversely, so these two coordinates contribute at least $1/K$ each to the L^1 distance, giving the lower bound $2/K$. Equality holds for two words differing only in module i , with $n_i = 1$ and $n_j = K - 1$ in both states: the two module- i coordinates are $1/K$ and 0 (in either order) and the module- j coordinates coincide. For the decoder, if a measured composition p' with $\|p' - p(z)\|_1 < 1/K$ were at least as close to some $z' \in \mathcal{C}_\tau$, $\tau \neq \sigma$, the triangle inequality would give $\|p(z) - p(z')\|_1 < 2/K$, contradicting (13). $\square$

For $K = 9$ the gap is $2/9$ and the tolerance $1/9$; for the count band $1, \dots, 19$ of the earlier quenched construction the sharp gap is $1/10$. Word pairs differing in both modules may be farther apart; (13) is the global minimum. A fixed exact decoder simply tests which of the two species of each module has positive fraction (**decode**); on the admitted regions the local alternative-species fractions are exactly zero or one. These are measurement guarantees and say nothing about a reaction system removing contaminating molecules.

Table 2: Selected-lineage and family bounds. Percentages are rounded evaluations of the exact bases q_K and u_K .

Selected lineage, Proposition 7.1

K	cycles N	lower bound q_K^N	upper bound u_K^N
9	100	44.28%	45.71%
20	1000	98.47%	99.62%

Complete family, Corollary 7.2

K	depth N	cycles $2^N - 1$	lower bound	upper bound $u_K^{2^N - 1}$
9	3	7	94.32%	94.67%
9	5	31	74.85%	78.47%
20	10	1023	98.42%	99.61%
20	14	16383	74.76%	93.94%

7 Repeated copying

7.1 The actual successful-restart kernel

Let $\mathcal{A}_K = \{1, \dots, K-1\}^2$ be the admitted selected counts and $P_T(x, n)$ the transition probability of the count chain over one batch. For $x, y \in \mathcal{A}_K$ define

$$W(x, y) = \sum_n P_T(x, n) \prod_{i=0}^1 \left[2^{-n_i} \binom{n_i}{y_i} \mathbf{1}\{1 \leq y_i < n_i\} \right]. \quad (14)$$

$W(x, y)$ is the probability that, starting from admitted counts x , the split is successful and daughter A restarts with counts y ; the indicator tests both complementary daughters, food allocation has been summed out because its binomial law sums to one, and the refill fixes daughter A's full next-cycle state. Hence $\sum_y W(x, y) = \mathbb{E}_x Q(n(T))$ is exactly the joint-return probability, and the missing mass is failure. By symmetry of the split the same kernel describes daughter B. This constructs the restart law from the protocol rather than positing a kernel with a desired reliability.

7.2 A selected lineage

Proposition 7.1 (Finite-lineage retention). *Run the cycle repeatedly with fresh label-blind supplies and resets, inspect both daughters at every split, and follow daughter A. Let q_K be a uniform one-cycle bound from Theorem 3.4 or 4.1 and $u_K = (1 - 2^{1-K})^2$. The probability $S_N(x)$ that the first N cycles all succeed satisfies*

$$q_K^N \leq S_N(x) \leq u_K^N \quad (x \in \mathcal{A}_K), \quad (15)$$

and the probability that the selected lineage never fails is zero.

Proof. By the Markov property at the restart states, $S_0 = 1$ and $S_{N+1}(x) = \sum_y W(x, y) S_N(y)$. Every row sum of W lies between q_K (the uniform theorem) and u_K (since $n_i \leq K$ and Q is increasing, $Q \leq Q(K, K) = u_K$). Induction gives (15). The events {first N cycles succeed} decrease in N , so their probabilities converge to the probability of eternal success, which is at most $\lim u_K^N = 0$. No unconditional independence between generations is used. $\square$

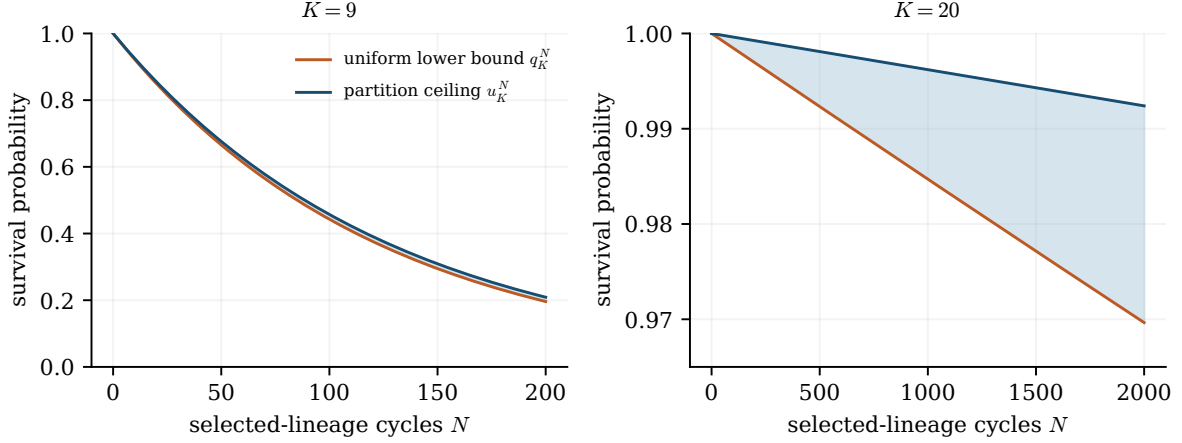


Figure 4: Selected-lineage bounds (15). The shaded band lies between the uniform lower guarantee q_K^N and the partition ceiling u_K^N ; axis ranges differ to show the two time scales. Neither curve assumes independence between generations.

7.3 A complete family

In the multi-generation protocol each successful daughter becomes a separate compartment with its own supplies, and separate compartments evolve independently given their restart states. A *family of depth N* consists of the $2^N - 1$ cycles run by the founder and all its descendants down to depth $N - 1$; it succeeds when every one of these cycles succeeds, so that all 2^N descendants at depth N are admitted states of the founder's word.

Corollary 7.2 (Family retention). *Under the assumptions of Proposition 7.1, the probability that a family of depth N succeeds satisfies*

$$1 - (2^N - 1)(1 - q_K) \leq \mathbb{P}(\text{family of depth } N \text{ succeeds}) \leq u_K^{2^N - 1}. \quad (16)$$

Proof. Order the $2^N - 1$ internal nodes so that each node follows its ancestors. The family fails only if some node v fails while all its strict ancestors succeeded; on that event v is in an admitted state, and conditionally on the history up to the start of v 's cycle the failure probability of v is at most $1 - q_K$ by the uniform theorem and the Markov property at v 's restart state. Hence each such event has probability at most $1 - q_K$, and the union bound gives the lower bound. For the upper bound, condition on all cycles but the last in the order: the last node's success has conditional probability at most u_K , so the family success probability is at most u_K times the success probability of the remaining nodes, and induction gives $u_K^{2^N - 1}$. $\square$

Table 2 and Figure 4 give the numbers. Two objectives are visibly distinct: $K = 9$ minimizes molecules while $K = 20$ retains a label across thousands of cycles, and neither is close to indefinite retention. The family bound concerns all descendants; the lineage bound concerns one selected line while both daughters are inspected at each split.

8 Stationary law and label symmetry

Proposition 8.1 (Product-form stationary law). *Fix a word and put $R = a/b > 0$. The one-module distribution*

$$\pi_K(n) = \frac{\binom{K}{n} R^n}{(1 + R)^K - 1}, \quad 1 \leq n \leq K, \quad (17)$$

is a probability law, and $\pi_K \otimes \pi_K$ is the unique stationary law of the count chain with rates (3)–(4), for every $\gamma \geq 0$.

Proof. Normalization is the binomial theorem. Since $\pi_K(n+1)/\pi_K(n) = R(K-n)/(n+1)$, for every fixed neighbour count m

$$\pi_K(n)\pi_K(m) \cdot a n(K-n)(1+\gamma m) = \pi_K(n+1)\pi_K(m) \cdot b(n+1)n(1+\gamma m),$$

so every edge of the chain satisfies detailed balance with respect to $\pi_K \otimes \pi_K$ [11]. The chain on $\{1, \dots, K\}^2$ is irreducible for $a, b > 0$, so the stationary law is unique. $\square$

The catalytic factor $1 + \gamma m$ multiplies both directions of every edge and cancels in the balance equation. Cross-catalysis therefore changes rates, trajectories and the transient law without changing the stationary law, and the factorization of π does not imply independence of the two transient count processes. The calculation is consistent with the product-form theory for complex-balanced networks [3], which is not needed here. The stationary expectation $\mathbb{E}_{\pi} g = \sum_n \pi_K(n) 2^{1-n}$ is the exact rational number used in Table 1; at $K = 9$ it equals 0.0039415, so the stationary one-cycle success is $(1 - \mathbb{E}_{\pi} g)^2 = 0.9921325$.

Permuting X_i and Y_i in either module preserves the reaction table, the partition rule and the refill. At matched count preparations the four words therefore have identical count-process and copying laws: the construction supplies inheritance without intrinsic fitness differences between its labels.

9 Worked cycles, limitations and scope

9.1 Two complete cycles at $K = 9$

Example 9.1. For word 01 prepare $(4X_0, 4Y_1; 5F_0, 5F_1)$: eight residents and ten food molecules. A possible parent state at the split has nine selected residents in each module and no food. One successful allocation is

$$A = (4X_0, 5Y_1), \quad B = (5X_0, 4Y_1).$$

Add food $(5, 4)$ to A and $(4, 5)$ to B. Each daughter restarts with module totals nine, and the total food added is eighteen.

Example 9.2. Complete conversion of food is not required. A parent $(7X_0, 6Y_1; 2F_0, 3F_1)$ can split as

	X_0	Y_1	F_0	F_1
A	3	2	1	2
B	4	4	1	1

with all other resident counts zero. Module totals are $(4, 4)$ in A and $(5, 5)$ in B; add food $(5, 5)$ and $(4, 4)$ respectively. The final food counts are $(6, 7)$ and $(5, 5)$, each module again holds nine modelled molecules, both daughters carry word 01, and the total replenishment is still eighteen.

9.2 A precise support limitation

Proposition 9.3 (Contamination is permanent). *Add one molecule of a previously absent resident species to a pure module and run the table (1). Then the count of that species remains positive for all time, and the probability that both daughters return to the original pure-label region in that cycle is zero.*

Proof. A channel decreases the count of a species only through reverse autocatalysis, whose propensity contains the factor $n(n-1)$ and vanishes at $n=1$; catalytic appearances in other channels do not consume the species. Hence the count never returns to zero. At the split, at least one daughter receives at least one molecule of the alternative species, and the refill removes none, so that daughter is outside the original region. $\square$

The mechanism thus lacks any active error correction: it tolerates dilution and partition losses, but not contamination through impure food, side reactions or mutation. The statement is restricted to the table (1); protocols with resident removal, conversion channels, mutation correction or a different admitted geometry are outside its scope, and designing a small-budget mechanism with common positive support and a positive mutation channel is the natural next question (Section 10).

9.3 External operations and physical scope

The eighteen-molecule accounting includes residents and food, not solvent, membranes, apparatus or the controller. The split is scheduled at a deterministic time, the allocation is perfectly fair and molecule-wise, and the refill requires an agent that measures the module totals M_i^A, M_i^B and injects exactly the missing food without disturbing the residents. The volume reset is likewise ideal. The controller never reads which species is present, but it is external machinery, so the result is a benchmark for compartmentalized chemical memory under a stated protocol and not a model of an autonomous protocell. The trimolecular channels need an elementary realization before an experimental claim can be made (Remark 2.1); the rate constants and the time unit are hypothetical, and the equilibrium ratio $a/b = 1000$ corresponds to a substantial free-energy drop per birth (Section 2.3) that the refilled food must carry. Thermodynamic consistency of the table does not quantify heat, work or the feasibility of the external operations.

10 Discussion and open questions

The paper answers a finite, quantitative question: two interacting reversible modules with nine modelled molecules each transmit four composition labels through a complete split-and-refill cycle with a uniform guarantee above 0.9918, within 2.5×10^{-4} of the numerically evaluated probability and within 7×10^{-5} of the combinatorial ceiling, and no fair-split architecture of this kind can reach 0.99 with fewer than eighteen peak residents. The proof is short because the observable is exact: $g(n) = 2^{1-n}$ is the fair-partition failure probability, and the reaction table is arranged so that g decays under the generator at rate thousands of times faster than the deadline.

Three limitations define the next questions. First, the labels occupy disjoint species supports and contamination is permanent; a mechanism whose four labels share a common positive support and survive a positive mutation or immigration channel would need a genuinely different restoring dynamics, and the balance between noise and restoration should be tested on a small model before any budget is claimed. Second, the split, the refill and the volume reset are external; replacing them by growth-driven division and a food-uptake channel changes the deadline, the conservation law and the admitted regions, and the immediate-return semantics used here would no longer apply unchanged. Third, the termolecular channels are effective rate laws; Remark 2.1 shows which parts of the argument survive an explicit complex and which must be re-proved. None of these questions is answered by the present theorem, and none of the present bounds depends on their answers.

A Formal verification and reproducibility

The critical statements are checked in Lean 4 (toolchain `leanprover/lean4:v4.30.0`) against a frozen Mathlib manifest [5, 23]. Verification treats compiler warnings as errors and rejects any declaration depending on `sorry`. The root file `proofs/SmallResidentCompositionCopying/Publication.lean` exports three declarations. All three live in the namespace `SmallResidentCompositionCopying.FiniteFood`, and their content is as follows.

- **publication**: for every word and admitted state at $K = 9$, the actual return probability is at least $8046297159/8112104000$; on the operating box it is at least $114080769/115203200$; distinct words have composition distance at least $2/9$; newborns have at most 16 and all states at most 18 residents;
- **return20**: the $K = 20$ bound $7077818258231/7077927321600$;
- **literal_next**: every positive-rate channel of the twenty-four-channel chemistry has the literal molecular stoichiometry of (1).

The machine-readable receipt records exit code zero, 18 transitive dependency modules with matching source and artifact hashes, a verification time of 60.031 seconds, and an axiom inventory consisting only of `Classical.choice`, `Quot.sound` and `propext`; no native evaluation is used. The root source has SHA-256 `d111998e...`; the frozen manifest has SHA-256 `a8df0b60...`. The earlier quenched construction and its receipts are not in the dependency graph of the root.

Paper statement	Lean declarations and status
Reaction table, count chain, generator identification, Section 2	<code>FiniteFoodSource</code> : <code>conserved</code> , <code>channel_sum</code> , <code>generator_identification</code> , <code>word_law</code> . The chemistry is a jump model on $\text{Fin } 4 \times \text{Bool} \times \text{Option Bool}$ (24 channels) with literal propensities; its finite-time law equals that of the aggregated chain for every word.
Literal stoichiometry, invariance of the pure face Lemma 3.1	<code>FiniteFoodInvariants</code> : <code>literal_next</code> , <code>food_transition</code> , <code>absent_stays_absent</code> , <code>resident_never_zero</code> .
Lemma 3.2	<code>FiniteFoodPartition</code> : <code>single_return</code> (the fair binomial sum minus its two endpoints equals $1 - 2^{-n}$), <code>actual_payoff</code> , <code>actual_return</code> .
Lemma 3.3 and Theorem 3.4	<code>CompositionalMemory</code> . <code>finite_deadline_residual_certificate</code> on the matrix-exponential law <code>finiteTimeExpectation</code> ; specialized in <code>FiniteFoodDrift</code> . <code>deadline</code> with the bound $1 - 2/(20s + 1) - 2C/s$ and <code>exp_bound</code> for $e^{-20s} \leq 1/(20s + 1)$.
Theorem 4.1	<code>FiniteFoodDrift</code> : <code>generator_H</code> , <code>reward_cover</code> , <code>local9</code> , <code>local20</code> , <code>nominal9</code> , <code>nominal20</code> ; <code>FiniteFoodPartition</code> : <code>return9</code> , <code>return20</code> , <code>refill</code> , <code>admitted_daughters</code> , <code>food_refill</code> .
Proposition 6.1, budgets, decoder	<code>FiniteFoodRobustness</code> : <code>OperatingBox</code> , <code>robust_local</code> , <code>robust9</code> , <code>positive_coefficients</code> , <code>common_ratio</code> ; <code>FiniteFoodPartition</code> . <code>return_box</code> . Parameters are symbolic over the box.
	<code>FiniteFoodGeometry</code> : <code>budgets</code> , <code>occupied</code> , <code>separated</code> (lower bound $2/K$), <code>decode</code> . The attainment of $2/K$ is the elementary example in the text.

Paper statement	Lean declarations and status
Proposition 5.1; kernel (14); Proposition 7.1; Corollary 7.2; Proposition 8.1; label symmetry; Proposition 9.3; Remark 2.1	<i>Conventional proofs</i> in the text, not separately Lean-verified. Their constants are replayed in exact arithmetic by <code>check_paper.py</code> .
Table 1, Figure 2, open circles in Figure 3	Floating-point matrix exponentials and Gillespie sample paths in <code>check_paper.py</code> ; diagnostics, not certificates.

Reproduction. From the repository root, through the proof bridge that fixes the frozen Mathlib project and does not update dependencies:

```
python scripts/verify_proof.py \
  proofs/SmallResidentCompositionCopying/Publication.lean --timeout 600 \
  --declaration SmallResidentCompositionCopying.FiniteFood.publication \
  --declaration SmallResidentCompositionCopying.FiniteFood.return20 \
  --declaration SmallResidentCompositionCopying.FiniteFood.literal_next
```

The manuscript sources include `check_paper.py`, which asserts every printed constant (the generator inequalities on all 81, 400 and 648 state or vertex cases, the rational bounds q_9 , q_{20} , q_{box} , the seventeen-resident ceiling, the sharp gaps $2/9$ and $1/10$, detailed balance on every edge, the lineage and family percentages and the two worked cycles) in exact rational arithmetic, evaluates the diagnostics of Table 1, and draws Figures 2–4. Its last run reported every check passed.

References

- [1] Author name to be inserted. Reliable copying of chemical states: finite-molecule guarantees and molecular redundancy. Companion manuscript, 2026; arXiv preprint in preparation, 2026.
- [2] David F. Anderson and Thomas G. Kurtz. *Stochastic Analysis of Biochemical Systems*, volume 1.2 of *Mathematical Biosciences Institute Lecture Series*. Springer, 2015. doi: 10.1007/978-3-319-16895-1.
- [3] David F. Anderson, Gheorghe Craciun, and Thomas G. Kurtz. Product-form stationary distributions for deficiency zero chemical reaction networks. *Bulletin of Mathematical Biology*, 72(8):1947–1970, 2010. doi: 10.1007/s11538-010-9517-4.
- [4] George Edward Briggs and John Burdon Sanderson Haldane. A note on the kinetics of enzyme action. *Biochemical Journal*, 19(2):338–339, 1925. doi: 10.1042/bj0190338.
- [5] Leonardo de Moura and Sebastian Ullrich. The Lean 4 theorem prover and programming language. In *Automated Deduction – CADE 28*, volume 12699 of *Lecture Notes in Computer Science*, pages 625–635. Springer, 2021. doi: 10.1007/978-3-030-79876-5_37.
- [6] Manfred Eigen. Selforganization of matter and the evolution of biological macromolecules. *Die Naturwissenschaften*, 58(10):465–523, 1971. doi: 10.1007/BF00623322.
- [7] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, New York, 1986. doi: 10.1002/9780470316658.
- [8] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977. doi: 10.1021/j100540a008.

- [9] Ankit Gupta, Corentin Briat, and Mustafa Khammash. A scalable computational framework for establishing long-term behavior of stochastic reaction networks. *PLoS Computational Biology*, 10(6):e1003669, 2014. doi: 10.1371/journal.pcbi.1003669.
- [10] Dann Huh and Johan Paulsson. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nature Genetics*, 43(2):95–100, 2011. doi: 10.1038/ng.729.
- [11] Frank P. Kelly. *Reversibility and Stochastic Networks*. Wiley, Chichester, 1979.
- [12] Juan Kuntz, Philipp Thomas, Guy-Bart Stan, and Mauricio Barahona. The exit time finite state projection scheme: bounding exit distributions and occupation measures of continuous-time Markov chains. *SIAM Journal on Scientific Computing*, 41(2):A748–A769, 2019. doi: 10.1137/18M1168261.
- [13] Thomas G. Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics*, 57(7):2976–2978, 1972. doi: 10.1063/1.1678692.
- [14] Yoshiya J. Matsubara, Sandeep Ameta, Shashi Thutupalli, Philippe Nghe, and Sandeep Krishna. Conditions for Darwinian evolution in compartmentalized autocatalytic reaction networks, 2026. arXiv:2211.03155v2, revision of 21 July 2026.
- [15] Sean P. Meyn and Richard L. Tweedie. Stability of Markovian processes III: Foster–Lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, 25(3): 518–548, 1993. doi: 10.2307/1427522.
- [16] Christopher V. Rao and Adam P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm. *The Journal of Chemical Physics*, 118(11):4999–5010, 2003. doi: 10.1063/1.1545446.
- [17] Jürgen Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. *Reviews of Modern Physics*, 48(4):571–585, 1976. doi: 10.1103/RevModPhys.48.571.
- [18] Lee A. Segel and Marshall Slemrod. The quasi-steady-state assumption: a case study in perturbation. *SIAM Review*, 31(3):446–477, 1989. doi: 10.1137/1031091.
- [19] Daniel Segré, Dafna Ben-Eli, and Doron Lancet. Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. *Proceedings of the National Academy of Sciences*, 97(8):4112–4117, 2000. doi: 10.1073/pnas.97.8.4112.
- [20] Sergey N. Semenov, Lewis J. Kraft, Alar Ainla, Mengxia Zhao, Mostafa Baghbanzadeh, Victoria E. Campbell, Kyungtae Kang, Jerome M. Fox, and George M. Whitesides. Autocatalytic, bistable, oscillatory networks of biologically relevant organic reactions. *Nature*, 537:656–660, 2016. doi: 10.1038/nature19776.
- [21] Angad Yuvraj Singh and Sanjay Jain. Multistable protocells can aid the evolution of prebiotic autocatalytic sets, 2023. arXiv:2310.19744.
- [22] Eörs Szathmáry. The origin of replicators and reproducers. *Philosophical Transactions of the Royal Society B*, 361(1474):1761–1776, 2006. doi: 10.1098/rstb.2006.1912.
- [23] The mathlib Community. The Lean mathematical library. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020)*, pages 367–381. ACM, 2020. doi: 10.1145/3372885.3373824.

- [24] Vera Vasas, Chrisantha Fernando, Mauro Santos, Stuart Kauffman, and Eörs Szathmáry. Evolution before genes. *Biology Direct*, 7:1, 2012. doi: 10.1186/1745-6150-7-1.
- [25] Nathaniel Virgo and Nicholas Guttenberg. Heredity in messy chemistries. In *Proceedings of the European Conference on Artificial Life 2015*, pages 325–332. MIT Press, 2015. doi: 10.7551/978-0-262-33027-5-ch061.