

Reliable copying of chemical states: finite-molecule guarantees and molecular redundancy

Author name to be inserted

Affiliation to be inserted

13 September 2026

Abstract

Chemical multistability supplies candidate memory states, but a list of stable states does not say whether a finite population of molecules copies a state reliably through growth, division and recovery. We treat copying as a complete stochastic cycle: an admitted parent must reach division, its molecules must be allocated to two daughters, and both daughters must return to the admitted region of the same label, from which the same guarantee applies again. We prove three separate results and keep their scopes distinct. First, a constructed pair of interacting reversible modules stores four concentration-state labels with at most 874 resident molecules per newborn module; under ideal external pool and volume control, every admitted parent produces two correctly returned daughters at its first division within 20 seconds with probability at least 0.9901, uniformly over a positive interaction and growth-reversal interval. Second, for k interacting modules satisfying a local quadratic recovery inequality, bounded reaction activity and bounded total incident exchange, the one-generation failure probability is at most $kC(\gamma)e^{-cN}$ with an exponent and an interaction allowance independent of k ; for an explicit four-species module this gives sufficient copy numbers $N = O(\log(kG/\eta))$ for a designated lineage of G divisions and $O(G + \log(k/\eta))$ for a complete binary family, with matching necessary orders for immediate return of both daughters. Third, the nominal eight-species kinetic model of Semenov et al. supports a prescribed split, refill and 500-second recovery protocol at $39.853\ \mu\text{L}$ with joint daughter return probability at least 0.992. Uniform return implies repeated copying without any independence assumption between siblings or generations. The proofs combine finite-state generator certificates, exact rational arithmetic, the actual complementary allocation law and state-dependent offspring kernels. The finite-model theorems and the certified-generation kernel are checked in Lean 4; the remaining conventional steps are identified explicitly.

1 Introduction

A chemical system with several stable compositions can, in principle, store information in the composition it occupies. This idea underlies compositional models of prebiotic heredity [9, 13], analyses of inheritance in compartmentalized autocatalytic networks under growth and division [8, 11], and the question of how many heritable states a chemistry with a given number of species can support [14]. Bistable and oscillatory networks of small organic molecules have been realized experimentally in flow reactors [10]. None of this establishes that a *finite* population of molecules copies a stored state reliably, and the gap is not a technicality. Growth changes concentrations and may consume the very species that supports a state. Division allocates molecules at random, and the two daughters are complementary rather than independent. Any recovery after birth takes time and can itself fail. A stable composition is therefore only the starting point; what has to be controlled is the complete cycle.

This paper asks a quantitative version of the question. Under a fully specified stochastic growth-and-division protocol, can both daughters recover the parent’s chemical label and become admissible parents themselves, with an explicit probability and an explicit molecular budget? And when more labels must be stored, or more divisions must succeed, how does the required molecular redundancy grow?

Copying as a complete cycle. Throughout, a label ℓ is represented by an *admitted region* B_ℓ of newborn or restart states, chosen so that the copying guarantee is uniform over B_ℓ . One cycle succeeds when an admitted parent reaches the specified division event within the allowed time and resources, its molecules are distributed to two daughters by the protocol’s actual partition law, and both daughters satisfy the same label’s admission condition at the protocol’s inspection time. Success is deliberately stronger than a decoder assigning the correct label: a daughter can pass a threshold test while lying outside the set on which the next cycle’s theorem applies. The reward for this stronger definition is that repeated copying follows by conditioning alone, with no independence assumption between siblings or generations (Section 2).

Three results. We prove three separate results and keep their scopes distinct; Table 1 summarizes them.

- (A) *A small constructed source* (Section 3). Two interacting reversible two-species modules, coupled through shared consumptive growth and cross-catalysis, store four concentration-state labels with at most 874 resident molecules in each newborn module. Under an ideal external controller that prescribes the volume and holds two food and two waste pools at fixed activities, every admitted parent produces two correctly returned daughters at its first division within 20 seconds with probability at least 0.9901, uniformly over the interaction strength $\epsilon \in [0, 0.1]$ and the growth-reversal rate $\rho \in [0, 2 \times 10^{-7}]$, and in particular on a strictly positive rectangle. The proof is a finite-state generator certificate checked in exact integer arithmetic, together with a sharp stopped-counter estimate that makes the complete-cycle reservoir capacity per pool about 0.166 pmol.
- (B) *Molecular redundancy for many modules* (Section 4). For k interacting modules with d resident species each, sharing consumptive growth and exchanging a growth species through a graph with bounded total incident weight, a local quadratic recovery inequality and bounded reaction activity give a one-generation failure probability at most $kC(\gamma)e^{-cN}$, where N is the birth membrane count per module and the exponent c and interaction allowance do not depend on k . For an explicit four-species module the sufficient copy number for a designated lineage of G divisions is $O(\log(kG/\eta))$ at target error η , and $O(G + \log(k/\eta))$ for a complete binary family. A partition obstruction gives matching necessary orders for immediate return of both daughters, so these orders are sharp for that copying criterion. The constants are astronomically conservative, and the sufficient bound rescales physical rates with k ; we state exactly which costs the asymptotic statement hides.
- (C) *A nominal published chemistry* (Section 5). The eight-species kinetic model reconstructed from Semenov et al. [10], at its nominal rate constants, supports a prescribed protocol of fair splitting, refilling with fresh feed and 500 seconds of recovery in a continuously stirred reactor of 39.853 μL : both daughters return to the same rational return region with probability at least $124/125 = 0.992$. The certified regions are narrow, the recovery is imposed, and deterministic sensitivity diagnostics show why the theorem must not be read as an experimentally validated uncertainty range.

The conceptual link between the three is reliable return of a label under specified operations. They do not combine into one chemistry that is simultaneously small, experimentally identified

Table 1: The three results and their scopes. “Return” always means that both daughters lie in the admitted region of the same label; the inspection time differs between protocols.

	(A) Constructed source	(B) Redundancy theorem	(C) Nominal Semenov model
Chemistry	two two-species modules, hypothetical reversible rate table, shared growth marker, cross-catalysis	k modules of d species; explicit four-species and two-species instances; growth-species exchange graph	eight tracked species of Semenov et al. [10] in a flow reactor at nominal rates
Stored quantity	four concentration-state words $\sigma \in \{0, 1\}^2$	words $\sigma \in \{0, 1\}^k$ with a relative-composition readout	two labels, thiol concentration state (relative readout also valid)
Resource measure	≤ 874 residents per newborn module; birth marker $N = 53$; pools reset by the controller	birth membrane count N per module; $N \geq 1.4 \times 10^{20}$ floor for the four-species instance	$\Omega = 2.4 \times 10^{19}$ molecules per molar, $V = 39.853 \mu\text{L}$; feed quota 1.2×10^{19} per daughter
Success event	first division by 20 s, no monitor triggered, both daughters in the boxes at birth	certified generation, or immediate first division, with both daughters in the birth regions	both daughters in the return ellipsoid after split, refill and 500 s recovery
Guarantee	≥ 0.9901 per cycle, uniform in $\epsilon \in [0, 0.1]$, $\rho \in [0, 2 \times 10^{-7}]$	failure $\leq kC(\gamma)e^{-cN}$; $N = \Theta(\log(kG/\eta))$ lineage, $\Theta(G + \log(k/\eta))$ family	≥ 0.992 per cycle, uniform over the return regions
Main limitations	ideal instantaneous pool and volume control; hypothetical compounds	conservative constants; rates rescaled with k ; deadline $O(k/\gamma)$; necessity only for immediate return	nominal parameters; narrow regions; imposed recovery; known model/data discrepancy
Verification	Lean 4, exact integer tables	Lean 4 for the certified kernel and the necessity bound; one conventional transfer lemma	Lean 4, exact rational polynomial certificates

and autonomous: the 874-molecule bound belongs to the constructed source, the logarithmic orders belong to the redundancy theorem’s own chemistry, and the Semenov result uses roughly 10^{18} active molecules under an external protocol.

What is not claimed. None of the results establishes autonomous compartment growth, membrane geometry, selection, differential reproduction, or open-ended evolution; the guarantees concern copying under the stated external operations. The Semenov result is a prediction of a specified nominal model, not an experimental demonstration. Neither result claims to introduce chemical heredity or bistability as topics. We also do not claim that exponential residence at large volume is new; it has general antecedents in large-deviation theory for reaction networks [1] and in Lyapunov methods for stochastic reaction networks [5, 6]. The contributions are the explicit, uniform finite-copy bounds attached to literal stochastic protocols, the uniformity in the number of modules, and the fact that every quantitative constant is the output of an exactly checked certificate rather than an estimate from simulated trajectories.

Methods. All processes are continuous-time Markov jump processes on integer counts with mass-action propensities [2, 4, 7]. Probabilities of one cycle are bounded by comparison functions

on a finite stopped model: a subsolution v of the backward inequality and a nonnegative decaying witness w turn a generator inequality into a deadline bound (Lemma 3.2). The comparison functions are proposed numerically and then checked in exact integer or rational arithmetic; for the Semenov model they are time-dependent rational quadratic forms on 38 polynomial time pieces. Division is handled through the actual complementary binomial allocation, either as an exact terminal payoff or through union bounds, never through a product of sibling probabilities. The passage from finite stopped models to the actual count process uses nonexplosion and a pathwise coupling up to the first monitored event. The finite-model theorems and the certified-generation kernel of the redundancy theorem are checked in Lean 4 with Mathlib [3, 12]; Appendix A lists the declarations, the receipts and the exact boundary of what is machine-checked.

Organization. Section 2 defines the copying cycle, offspring kernels and the repeated-copying lemma, and distinguishes a designated lineage from a complete family. Section 3 states and proves the constructed-source theorem. Section 4 states the general transmission principle, its four-species instance, the immediate-copying bounds and the resulting molecular-redundancy orders. Section 5 treats the nominal Semenov protocol. Section 6 discusses what the results establish and what they leave open. The appendices give the verification map, the four-species source certificates and the Semenov certificate constants.

2 Copying events, offspring kernels and repeated copying

2.1 States, labels and admitted regions

A compartment state consists of the integer count vector n of its resident species, a volume-like variable (a membrane or marker count m in the growth protocols, a fixed volume V in the split-refill protocol), and whatever protocol variables affect the future law (an event counter, a feed inventory). A *label* ℓ ranges over a finite set; for k binary modules it is a word $\sigma \in \{0, 1\}^k$. Each label has an *admitted region* B_ℓ , a set of restart states. A *decoder* is a fixed map from states to labels that never inspects the true label; in every protocol below it is a coordinate threshold. Membership in B_ℓ is strictly stronger than decoder correctness: B_ℓ is the set on which the one-cycle theorem is proved uniformly, and a state that decodes correctly but lies outside B_ℓ carries no guarantee for the next cycle.

Which quantity a label measures is part of the statement. In Section 3 the admitted regions are boxes of resident *counts* at a prescribed volume, so the stored quantity is a concentration state (Proposition 3.5 shows that normalized composition alone does not separate the regions). In Sections 4 and 5 the regions additionally admit a relative-composition readout.

2.2 The copying cycle

Definition 2.1 (One copying cycle). Fix a protocol and a label ℓ . Starting from an admitted parent $n \in B_\ell$, the cycle *succeeds* when, in chronological order,

- (i) the parent reaches the protocol’s division event within the allowed time, without triggering any failure monitor the protocol prescribes (resource quota, domain exit, forbidden event);
- (ii) the resident molecules are assigned to two daughters by the protocol’s partition law, and the daughters’ volumes and any refill are set as prescribed;
- (iii) both daughters complete the prescribed recovery, if any, and lie in B_ℓ at the protocol’s inspection time;
- (iv) the state variables needed to run the protocol again are restored by the protocol’s label-blind reset, so that daughter A’s inspected state is again an admitted parent.

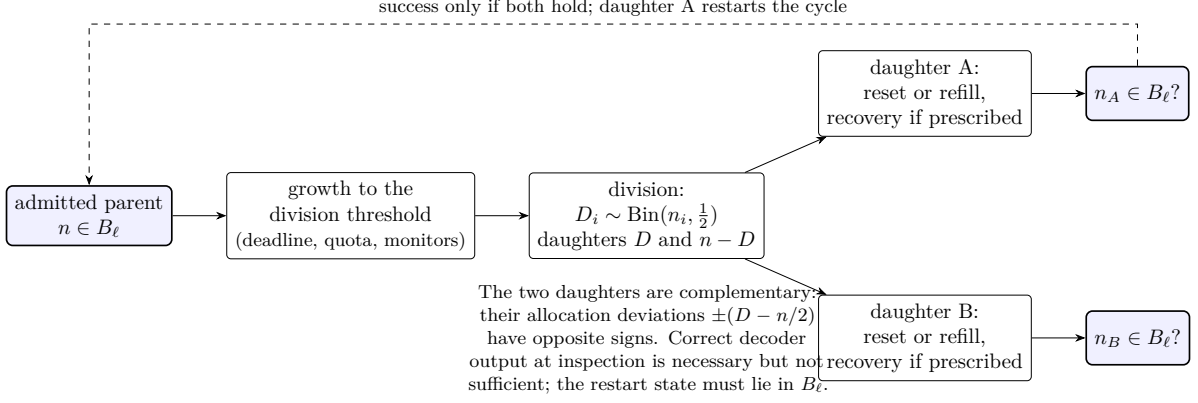


Figure 1: One copying cycle. Uniform return to B_ℓ is proved for every parent in B_ℓ ; a success at daughter A therefore restarts the same guarantee.

Two inspection conventions occur. In the growth–division protocols of Sections 3 and 4, daughters are inspected *at birth*, immediately after allocation, and the next cycle starts from the birth state; nothing after birth is required. In the split–refill–recovery protocol of Section 5, daughters are inspected at the end of a fixed recovery interval, and daughter A’s terminal state is the next parent. In both cases the inspected state is exactly the state from which the next cycle is run, so no additional residence estimate is needed. Figure 1 shows the cycle.

2.3 Offspring kernels and repeated copying

Let S_ℓ denote the set of possible restart states of daughter A (a finite set in all protocols below). One cycle defines a substochastic *offspring kernel*

$$K_\ell(n, n') = \mathbb{P}_n(\text{cycle succeeds, daughter A's inspected state is } n'), \quad n \in B_\ell, \quad n' \in S_\ell,$$

whose missing mass $1 - K_\ell(n, S_\ell)$ is the failure probability from n . Success forces $n' \in B_\ell$, so $K_\ell(n, S_\ell) = K_\ell(n, B_\ell)$. Because the protocol’s reset is label-blind and restores every non-resident variable to its prescribed value, the law of the next cycle depends on the past only through n' .

Lemma 2.2 (Uniform return implies repeated copying). *If $K_\ell(n, B_\ell) \geq p$ for every $n \in B_\ell$, then for every $n \in B_\ell$ and $G \geq 0$ the probability that all G inspected cycles along the designated daughter-A lineage succeed is at least p^G . Writing $p = 1 - \varepsilon$, it is also at least $\max\{0, 1 - G\varepsilon\}$.*

Proof. Let $S_0 \equiv 1$ on B_ℓ and $S_{g+1}(n) = \sum_{n' \in B_\ell} K_\ell(n, n') S_g(n')$, the probability of $g + 1$ successive successes from n . If $S_g \geq p^g$ on B_ℓ then $S_{g+1}(n) \geq p^g K_\ell(n, B_\ell) \geq p^{g+1}$. The linear bound is Bernoulli’s inequality, or a union bound over the G conditional failure events. No independence between generations is used; the multiplication is valid because the conditional guarantee is uniform over the restart set. $\square$

Lemma 2.2 inspects both daughters at every counted division but follows only daughter A. It says nothing about the other daughter’s descendants. For a complete family the number of inspected divisions grows exponentially with the depth.

Proposition 2.3 (A designated lineage and an entire family). *Let $K_\ell(n, (n_A, n_B))$ be the joint law of both inspected daughter states on success, with $\sum_{(n_A, n_B)} K_\ell(n, (n_A, n_B)) \geq p = 1 - \varepsilon$ for all $n \in B_\ell$. Let F_G be the event that every division in the complete binary family through depth G succeeds, the initial parent being at depth 0. There are $D_G = 2^G - 1$ such divisions, and*

$$\mathbb{P}_n(F_G) \geq p^{2^G - 1} \geq 1 - (2^G - 1)\varepsilon.$$

Proof. Define $f_0 \equiv 1$ and $f_{g+1}(n) = \sum_{(n_A, n_B)} K_\ell(n, (n_A, n_B)) f_g(n_A) f_g(n_B)$. This is the probability that the family through depth $g + 1$ succeeds, constructed with fresh compartment trajectories conditional on the *joint* daughter state; it does not factor the distribution of the sibling compositions. If $f_g \geq p^{D_g}$ on B_ℓ then $f_{g+1} \geq p^{2D_g} p = p^{D_{g+1}}$, since $D_{g+1} = 1 + 2D_g$. Equivalently, reveal divisions in breadth-first order while all earlier ones have succeeded; each next parent was born admitted, so its conditional success probability is at least p , and there are exactly D_G opportunities. $\square$

The terminal generation of the family is inspected with the same convention as in Definition 2.1: at birth in the growth protocols, after recovery in the split-refill protocol. A one-cycle bound of 0.9901 gives $0.9901^{10} > 0.905$ along ten lineage divisions but only $0.9901^{1023} \approx 3.8 \times 10^{-5}$ for a complete family of depth ten; the family bound is a lower bound and not evidence that most descendants fail.

2.4 Complementary allocation

In every protocol below, division allocates each resident molecule independently and fairly to one of the two daughters. Conditional on the parent’s count vector n , one draws $D_i \sim \text{Bin}(n_i, 1/2)$ independently across species and assigns $(D_i, n_i - D_i)$ to daughters A and B. The two daughter vectors are dependent: their deviations from $n/2$ are negatives of each other, and $\text{Cov}(D_i, n_i - D_i) = -n_i/4$. Consequently the probability that both daughters return is *not* the product of the two marginal return probabilities. Two devices replace that product. In Section 3 the joint return probability is an exact finite binomial sum and is used as the terminal payoff. In Sections 4 and 5 the daughters’ deviations are bounded by a common event (a sup-norm tail, or a union bound over the two marginals); the reflection $d \mapsto n - d$ shows that daughter B has the same marginal law as daughter A, so one marginal estimate serves both.

Remark 2.4 (What uniformity must include). The restart set must contain every variable that can change the future law. In the constructed source the controller resets the food, waste and marker pools and the event counter at each division, without inspecting the label or touching the resident counts; in the Semenov protocol the feed quota is renewed each cycle. In both cases the restart state is the resident count vector alone, and the bounds of Theorems 3.1 and 5.1 are uniform over it. A bound uniform only in resident concentration would not justify another generation if a depleted reservoir or a controller phase were carried across cycles; we therefore charge the reservoir explicitly (Proposition 3.4) and state where fresh resources enter.

3 A small constructed source with two interacting modules

This section gives the first result: a hypothetical but fully specified chemistry that copies four concentration-state labels with at most 874 resident molecules per newborn module. The mechanism is deliberately small so that its one-cycle probability can be certified in exact integer arithmetic on a finite state space. Its physical assumptions are equally explicit: an ideal external controller prescribes the volume, holds food and waste pools at fixed activities, and resets those pools at division without inspecting the stored label.

3.1 Chemistry

There are two resident modules $i = 0, 1$, each with two resident species X_i, Y_i , a shared growth marker with count m , and, for each module, a controlled food pool F_i and a controlled waste pool W_i . The resident reactions of each module are

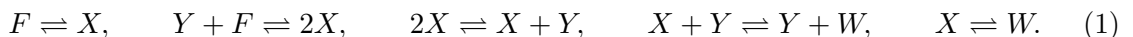


Table 2: Effective propensities per module in model time. x, y are the module’s resident counts and m is the shared marker count, which also sets the volume. The reverse coefficients are chosen so that the effective ratios $K_{FX} = 150000$, $K_{YX} = 1$ and $K_{XW} = 100000$ are consistent around every cycle of (1); these define a hypothetical chemistry.

Reaction pair	Forward propensity	Reverse propensity
$F \rightleftharpoons X$	$(193/5) m$	$(193/750000) x$
$Y + F \rightleftharpoons 2X$	$1555 y$	$(311/30000) x(x-1)/m$
$2X \rightleftharpoons X + Y$	$15 x(x-1)/m$	$15 xy/m$
$X + Y \rightleftharpoons Y + W$	$100 xy/m$	$y/1000$
$X \rightleftharpoons W$	$(105/2) x$	$(21/40000) m$

Read chemically: food is converted to X directly and, autocatalytically, by Y ; two X molecules convert one of them to Y ; Y catalyses the degradation of X to waste; and X leaks to waste. Every reaction is reversible. At the maintained food and waste activities the effective count propensities are those of Table 2; repeated reactants use falling factorials with no additional factorial divisor, so $2X \rightarrow X + Y$ has propensity $15x(x-1)/m$.

The modules interact in two ways. Each module contributes to growth through $X_i \rightarrow M$ at propensity $x_i/10$: one resident molecule is consumed and the shared marker m increases by one, which dilutes both modules. The reverse channel $M \rightarrow X_i$ has propensity ρm . Cross-catalysis $X_i + Y_j \rightleftharpoons Y_i + Y_j$ for $j \neq i$ has propensities $\epsilon x_i y_j/m$ forward and $\epsilon y_i y_j/m$ backward. The two parameters $\epsilon \geq 0$ and $\rho \geq 0$ are the interaction strength and the growth-reversal rate; the theorem below is uniform over an interval of each.

Assigning equal positive masses to F, X, Y, W and M makes every displayed reaction mass-balanced. The imposed food-to-waste affinity at the stipulated pool activities is $\log(K_{FX}K_{XW}) = \log(1.5 \times 10^{10}) \simeq 23.43$ in units of $k_B T$. This is the driving affinity of the reservoir pair, not a dissipation per copied bit.

Dimensional calibration. Take a concentration scale $C_0 = 1 \mu\text{M}$, one second per model-time unit, and food and waste concentrations $1000 C_0$. The controller prescribes the volume $V = m/(N_A C_0)$ and holds each local pool at $1000 m$ molecules after every event. A newborn has $m = N = 53$, corresponding to a volume of about 0.0880 fL , and divides at the first arrival at $m = 2N = 106$. The corresponding physical rate table has, for example, $Y + F \rightarrow 2X$ at $1.555 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ at 1 mM food and $X \rightarrow W$ at 52.5 s^{-1} . The marker is a chemical growth product coupled to the prescribed volume; it is not asserted to be a membrane, and exact instantaneous pool adjustment, prescribed volume and well-mixed equal-volume division are assumptions of the protocol rather than consequences of the chemistry.

3.2 Protocol, regions and the theorem

The protocol counts every model event, including an auxiliary rate-one self-event that keeps the total jump rate positive, and rejects the cycle when the counter reaches the quota $J = 10^8$. It monitors exits from the certified resident domain $\{x_i \leq 16m, y_i \leq 4m\}$ and every reverse-growth event; both are counted as failure. At division, the residents of each module undergo the complementary allocation of Section 2.4. The controller then sets each daughter’s marker to 53 and each local pool to 53000 molecules, without changing the allocated resident counts or inspecting the label.

At birth the admitted boxes of the two labels are

$$B_0 = \{6 \leq x \leq 106, 0 \leq y \leq 8\}, \quad B_1 = \{371 \leq x \leq 742, 6 \leq y \leq 132\}, \quad (2)$$

with integer counts. A word $\sigma \in \{0, 1\}^2$ means $(x_i, y_i) \in B_{\sigma_i}$ for both modules, so the admitted region of the word is $B_\sigma = B_{\sigma_0} \times B_{\sigma_1}$. The decoder is the threshold $x > 212$, applied to each module separately. The preparations $(53, 0) \in B_0$ and $(530, 53) \in B_1$ show that every word has admitted states. A newborn module in B_1 has at most $742 + 132 = 874$ resident molecules, and one in B_0 at most 114.

Theorem 3.1 (Finite-drive concentration-state copying). *For the quota-controlled source above, every admitted newborn of every word has probability at least 0.9901 of first division by $T = 20$ with both complementary daughters returned to the same word's admitted boxes and no monitor triggered. This holds uniformly for*

$$0 \leq \epsilon \leq 0.1, \quad 0 \leq \rho \leq 2 \times 10^{-7},$$

in particular on the strictly positive rectangle $10^{-4} \leq \epsilon \leq 0.1$, $10^{-7} \leq \rho \leq 2 \times 10^{-7}$. Each newborn module has at most 874 resident molecules. Ten inspected lineage cycles succeed with probability at least $0.9901^{10} > 0.905$.

The success event is Definition 2.1 with inspection at birth: the restart state of daughter A is its birth count vector, and the marker, pools and counter are reset by the controller. The first actual division can be preceded by histories that the conservative monitors reject; the theorem bounds the probability of the monitored successful subset, which is contained in the event “first division by T with both daughters in B_σ ”. The last sentence is Lemma 2.2.

3.3 The proof: a finite generator certificate

For a finite jump model with channels r , propensities $a_r(z)$ and jumps ν_r , write

$$\mathcal{L}f(z) = \sum_r a_r(z)(f(z + \nu_r) - f(z)).$$

The following comparison converts generator inequalities into a deadline bound. It is the only probabilistic tool needed for the constructed source.

Lemma 3.2 (Deadline comparison). *Let Z be a finite-state jump process with generator $\mathcal{L}$, and let h, v, w be functions on its state space with $v \leq h + w$, $w \geq 0$. If $\mathcal{L}v \geq -\delta$ and $\mathcal{L}w \leq -sw + \eta$ with $s > 0$, $\eta \geq 0$, then for every state n and $T \geq 0$,*

$$\mathbb{E}_n h(Z_T) \geq v(n) - T\delta - e^{-sT}w(n) - \eta/s. \quad (3)$$

Proof. On a finite state space $u(t) = \mathbb{E}_n v(Z_t)$ is differentiable with $u'(t) = \mathbb{E}_n \mathcal{L}v(Z_t) \geq -\delta$, so $\mathbb{E}_n v(Z_T) \geq v(n) - T\delta$. Likewise $g(t) = \mathbb{E}_n w(Z_t)$ satisfies $g' \leq -sg + \eta$, and Grönwall's inequality gives $g(T) \leq e^{-sT}w(n) + \eta/s$. Subtracting, $\mathbb{E}_n h(Z_T) \geq \mathbb{E}_n v(Z_T) - \mathbb{E}_n w(Z_T)$ gives (3). The finite law may be realized by uniformization or by the matrix exponential; the identity $u' = \mathbb{E} \mathcal{L}v$ holds for either. $\square$

Local certificates with the other module as a control. The state of the pair has four resident coordinates and the marker; a direct backward solve on the joint space is unnecessary. Instead, for each module and each label, rational backward certificates are computed on the volume layers $m = 53, \dots, 106$ with proposal caps $x \leq 16m$, $y \leq 4m$. The other module's influence, its cross-catalytic propensity and its contribution to growth, enters the local generator through affine controls whose ranges are the admissible values on the other module's certified domain. Checking the control extremes covers every admissible neighbouring state, so the local computation makes no independence assumption; it is a worst case over the other module.

Let h_i be the exact terminal payoff of module i at $m = 106$: the probability, under complementary allocation of that module's residents, that both daughters' module- i counts lie

in B_{σ_i} , an exact finite binomial sum. Let $h = h_0 h_1$ be the joint payoff, which is exactly the probability that both daughters return. Combine the local certificates as

$$v = v_0 + v_1 - 1, \quad w = w_0 + w_1.$$

This is valid because $h_0 + h_1 - 1 \leq h_0 h_1$ for $h_0, h_1 \in [0, 1]$, so $v \leq h + w$ follows from the local inequalities $v_i \leq h_i + w_i$, and the generator of the sum is controlled by the local generators evaluated at the control extremes. Exact exit comparisons handle joint killing: an exit of either module kills the pair, and the local certificates are compared with zero continuation at their own exits.

The checked integer rows establish, for every state of the finite model,

$$v \leq 1, \quad w \geq 0, \quad v \leq h + w, \quad (4)$$

$$\mathcal{L}v \geq -2 \times 10^{-6}, \quad \mathcal{L}w \leq -\frac{5}{8}w + \frac{1}{50000}, \quad (5)$$

and at every admitted newborn n the birth row

$$500000 v(n) - 2w(n) \geq 495550. \quad (6)$$

The numerically proposed certificate is not assumed correct: every row of (4)–(6) is an integer inequality on exact rational data (about 2.2×10^7 transient rows), and the terminal payoffs are exact binomial sums. Appendix A states the trust boundary of the arithmetic.

Pre-monitor bound. Apply Lemma 3.2 with $T = 20$, $\delta = 2 \times 10^{-6}$, $s = 5/8$, $\eta = 1/50000$. Since $e^{-sT} = e^{-25/2} \leq 1/250000$, (6) gives $v(n) - e^{-sT}w(n) \geq 0.9911$, and hence

$$\mathbb{E}_n h(Z_T) \geq 0.9911 - 20 \cdot 2 \times 10^{-6} - \frac{8}{5} \cdot \frac{1}{50000} = 0.991028 \geq 0.991.$$

Reverse-growth monitor. Every reverse-growth event $M \rightarrow X_i$ is rejected. Before division $m \leq 2N$, so the combined hazard of the two reverse channels is at most $2\rho \cdot 2N = 4\rho N$, and the probability that such an event occurs by $T = 20$ is at most $80\rho N \leq 0.000848$ at $\rho = 2 \times 10^{-7}$. Killing at these events lowers the generator of v by at most the hazard, since $v \leq 1$ and the killed continuation is $0 \geq v - 1$; it can only improve the nonnegative witness w . No new backward solve is needed.

Event quota. The total event rate of the stopped model, including the self-event and the monitored reverse channels, is at most $\Lambda = 4 \times 10^6$ on the certified domain; rates after an already failed exit are immaterial because the process is frozen there.

Lemma 3.3 (Stopped exponential counter bound). *Let C_t count all events of a finite jump process whose total intensity is at most Λ , stopped at the quota J , with $C_0 = 0$. For every $a \geq 1$,*

$$\mathbb{P}(C_T = J) \leq a^{-J} \exp(\Lambda(a-1)T). \quad (7)$$

For $\Lambda = 4 \times 10^6$, $T = 20$, $J = 10^8$ and $a = 1 + 10^{-6}$ the right side is at most 2×10^{-7} .

Proof. Put $f(c) = a^c/a^J$. Before exhaustion, $\mathcal{L}f = \lambda(z)(a-1)f \leq \Lambda(a-1)f$, where $\lambda(z)$ is the total intensity, and at exhaustion the generator vanishes, so the inequality persists. Grönwall for the finite law gives $\mathbb{E}f(C_T) \leq e^{\Lambda(a-1)T}/a^J$, and $\mathbf{1}_{\{C_T=J\}} \leq f(C_T)$ proves (7), including state-dependent intensities and self-events. For the constants: Bernoulli's inequality gives $a^{10000} \geq 1.01$, exact rational arithmetic gives $1.01^{100} \geq 2.7$, hence $a^{10^8} \geq 2.7^{100}$; with $e < 11/4$, $e^{80}/a^{10^8} \leq (11/4)^{80}/2.7^{100} < 2 \times 10^{-7}$. $\square$

To attach the quota to the copying bound, lift v unchanged to the state augmented by the counter, including the frozen exhausted states; its drift lower bound is preserved. Set w to zero at exhaustion; its upper drift bound is preserved. The cover inequality $v \leq h + w$ then holds with h replaced by $h + \mathbf{1}_{\{C=J\}}$, because $v \leq 1$. Applying (3) to the augmented model and subtracting Lemma 3.3 gives the actual quota-controlled payoff

$$0.991 - 0.000848 - 0.0000002 = 0.9901518 > 0.9901.$$

The same finite model, with its monitors and quota, is the one whose first-division output defines the offspring kernel; Lemma 2.2 then gives the ten-cycle statement and completes the proof of Theorem 3.1.

3.4 Resource accounting

The 874-molecule figure counts residents of one newborn module. The other inventories are part of the protocol and are charged separately.

Proposition 3.4 (Complete-cycle reservoir capacity). *Each event changes a local pool and the marker by at most one molecule, so restoring a pool to $1000m$ after an event requires at most 1001 supplied or removed molecules. Including startup of the pool at $1000N$, the reset of both daughters' pools after an arbitrary complementary allocation, and decommissioning, a conservative gross supply capacity and a conservative gross disposal capacity per food or waste pool and per cycle are each*

$$4000N + 1001J = 100,100,212,000 \text{ molecules} \approx 0.1662 \text{ pmol}. \quad (8)$$

Four pools over ten inspected cycles need about 6.65 pmol of capacity. The marker allowance for startup, reset and removal is $4N = 212$ molecules per cycle. The initial local inventory is 213,801 molecules: at most 1748 residents, 212,000 food and waste molecules, and 53 marker molecules.

Proof. If a pool with parent count $2q$ is split as $(a, 2q - a)$, restoring both daughters to q needs at most q supplied and q removed molecules; with $q = 1000N$ this is the $2 \cdot 1000N$ reset term, and startup and decommissioning contribute $1000N$ each. Every counted event contributes at most 1001, and there are at most J events per cycle. The identities are checked in Lean (Appendix A). $\square$

These are worst-case capacities, not estimates of consumption. Compared with a crude linear exhaustion estimate, which would require a quota of 2×10^{12} for the same probability loss, Lemma 3.3 reduces the conservative capacity by a factor of about 2×10^4 . Finite capacity is compatible with the construction; it does not establish that the ideal controller is physically realizable.

3.5 What the labels measure

The decoder $x > 212$ is a count threshold at the prescribed birth volume, so it distinguishes concentration states. It is natural to ask whether the labels are also distinguished by the resident composition alone, that is, by the normalized proportions $x : y$. They are not.

Proposition 3.5 (Concentration memory, not ratio memory). *The states $(10, 1) \in B_0$ and $(400, 40) \in B_1$ have identical normalized resident compositions. Hence no decoder depending only on the proportions of X and Y separates all admitted low states from all admitted high states.*

Proof. $(400, 40) = 40 \cdot (10, 1)$, and both points satisfy (2). $\square$

Proposition 3.5 leaves Theorem 3.1 intact and fixes its interpretation: the constructed source stores which of two concentration states each module occupies at the controlled volume. An externally controlled reference pool does not change this. The Semenov regions of Section 5 and the redundancy construction of Section 4 do admit a relative-composition readout; the difference lies in the geometry of the admitted regions, not in the copying theorem.

4 Molecular redundancy for many interacting modules

The constructed source of Section 3 has two modules and fixed constants. This section asks how the required molecular redundancy grows when k modules must be copied and G divisions must succeed. The chemistry is different from Section 3: growth is consumptive with a membrane count rather than a controlled marker, there are no external pools to reset, and the interaction is exchange of a growth species through a graph. The answer is a uniform transmission principle: once local reaction quality is fixed, there are a positive exponent, a growth allowance and an interaction allowance that serve every k , every admissible exchange graph, every word and every admitted newborn.

4.1 Model

There are $k \geq 1$ disjoint modules, each with d resident species. The count state is (n, m) with $n_{ia} \in \mathbb{N}$ for module i and species a , and a common positive integer membrane count m , which is the physical volume. A newborn has $m = kN$ and divides at the first time $m = 2kN$; both daughters receive kN membrane units, and the residents are allocated complementarily as in Section 2.4, independently across all molecules. Reaction clocks after birth are fresh conditional on the newborn state.

A resident reaction ℓ of module i has consumption and production vectors $\alpha_{i\ell}, \beta_{i\ell} \in \mathbb{N}^d$, order $r_{i\ell} = \sum_a \alpha_{i\ell a}$, net change $\nu_{i\ell} = \beta_{i\ell} - \alpha_{i\ell}$ and reference coefficient $c_{i\ell} \geq 0$. Its physical propensity is

$$c_{i\ell} k^{r_{i\ell}-2} m^{1-r_{i\ell}} \prod_a (n_{ia})_{\alpha_{i\ell a}}, \quad (9)$$

with falling factorials $(n)_r = n(n-1)\cdots(n-r+1)$ and no factorial divisor. The designated growth species z is converted into membrane at physical propensity $\gamma n_{iz}/k$, changing (n_i, m) to $(n_i - e_z, m+1)$: growth consumes a resident molecule and dilutes every module. Directed exchange $i \rightarrow j$ transfers one z molecule at propensity $w_{ij} n_{iz}/k$, where

$$w_{ij} = w_{ji} \geq 0, \quad w_{ii} = 0, \quad \sum_j w_{ij} \leq \kappa. \quad (10)$$

The restriction is on the total incident exchange at a module, not on the graph size: a path or ring with edge weight $\kappa/2$ is admissible for every k (Corollary 4.3), whereas a complete graph with a fixed edge weight is not.

The analysis uses the rescaled coordinates $v = m/k \in [N, 2N]$, $u_i = n_i/v$ and time $\tau = t/k$; in these coordinates the resident propensity per unit τ is $c_{i\ell} v^{1-r_{i\ell}} \prod_a (n_{ia})_{\alpha_{i\ell a}}$ and the physical concentration is u_i/k . These are choices of the theorem: the model has dk species, (9) rescales physical coefficients with k , and a certified generation lasts at most $k(t_0 + t_1) = O(k/\gamma)$ in physical time. Section 4.5 returns to what this means.

Regions and two success events. For each word $\sigma \in \{0, 1\}^k$, module i has a centre a_i^σ and a positive quadratic form $E_i^\sigma(y) = Q_i^\sigma(y, y)$; with birth energy b the closed newborn region is

$$\mathcal{B}_\sigma(N) = \{n : E_i^\sigma(n_i/N - a_i^\sigma) \leq b \text{ for every } i\}. \quad (11)$$

A *certified generation* remains in an outer energy tube until a deterministic recovery time t_0 (in τ units), has not divided by t_0 , is inside an inner recovery region at t_0 , then remains in a smaller parent tube until division within a further t_1 units, and both daughters lie in $\mathcal{B}_\sigma(N)$. Leaving either tube fails the certificate irreversibly even if later chemistry would repair it. An *immediate successful division* requires only that the first actual division occurs and both daughters lie in $\mathcal{B}_\sigma(N)$; nondivision counts as failure and repair after birth does not count. Write A_G for immediate success at all G divisions along a designated daughter, both daughters inspected each time, and F_G for immediate success at every division of the complete binary family through depth G .

4.2 The uniform transmission principle

All constants below are fixed before k , the graph, the word and N ; sup norms are used throughout, and module indices are suppressed on local quantities.

Assumption 4.1 (Local quality). (H1) *Geometry.* $c_0\|y\|_\infty^2 \leq Q(y, y)$ and $|Q(y, h)| \leq L\|y\|_\infty\|h\|_\infty$. There are $M, \lambda, \rho > 0$ and levels $M\rho < h_{\text{rec}} < h_{\text{par}} < b < h_{\text{out}} \leq c_0 R_0^2$. Centres are nonnegative and at most $C - R_0$ coordinatewise for a fixed integer cap C , and a fixed readout threshold separates each centre from the opposite label by at least R_0 .

(H2) *Local drift.* On every outer product domain, each displacement $y = u - a$ admits $r \in [0, r_{\text{max}}]$ with $\|y\|_\infty \leq r$, $\|u\|_\infty \leq U$, $E(y) \leq Mr^2$, and

$$2Q(y, f(u)) \leq -\lambda r^2, \quad |Q(y, \nu_\ell)| \leq Pr, \quad |Q(\nu_\ell, \nu_\ell)| \leq R_1, \quad f(u) = \sum_\ell c_\ell u^{\alpha_\ell} \nu_\ell,$$

uniformly over all admissible integer states, not only at the centres.

(H3) *Activity and finite-volume bias.* $\sum_\ell c_\ell U^{r_\ell} \leq A$ and $\sum_\ell c_\ell r_\ell^2 (U + 1)^{r_\ell} \|\nu_\ell\|_\infty \leq D_0$; the second bound controls the discrepancy between falling-factorial rates and fluid monomials by D_0/v .

(H4) *Growth and interaction.* On the outer domains $0 < \ell_0 \leq u_{iz} \leq Z$, and w satisfies (10).

Theorem 4.2 (Uniform certified transmission). *Under Assumption 4.1 there exist $\gamma_0, \kappa_0, c_* > 0$, $N_0 < \infty$ and $t_0 > 0$, independent of k , N , w and σ , such that for every $0 < \gamma \leq \gamma_0$, with $t_1 = 11/(10\gamma\ell_0)$, every architecture with row sums at most κ_0 , every $N \geq N_0$, every word and every integer newborn in its closed birth region, the certified-generation kernel satisfies*

$$K_{\text{cert}}(n, \text{failure}) \leq \varepsilon = kC_*(\gamma)e^{-c_*N}, \quad C_*(\gamma) = 7 + 2d + \frac{\lambda\rho}{2} \left(\frac{t_0}{h_{\text{out}} - b} + \frac{t_1}{h_{\text{par}} - h_{\text{rec}}} \right). \quad (12)$$

The integer birth regions are nonempty and have distinct word readouts. Along a designated lineage the certified success probability over G generations is at least $1 - G\varepsilon$, and at least $(1 - \varepsilon)^G$ when $\varepsilon \leq 1$.

Proof sketch. The proof has three parts; Appendix B gives the constants for the explicit source.

Shared growth cancels correctly. At a growth event in module j , module i sees the jump $\Delta u_i = -(u_i + k\mathbf{1}_{\{i=j\}}e_z)/(m+1)$. Summing over j with propensities $\gamma v u_{jz}$ before estimating gives the limiting growth force $-\gamma u_{iz}e_z - \gamma \bar{u}_z u_i$ with $\bar{u}_z = k^{-1} \sum_j u_{jz}$: consumption plus dilution by shared growth. For the quadratic variation the single consuming jump has size at most $(U + k)/(m+1)$ and the other $k-1$ jumps at most $U/(m+1)$, so the total is at most $\gamma Z(U+1)^2/v$, with no factor k . Exchange contributes a force of size at most $2Z\kappa$ and quadratic variation at most $2Z\kappa/v$; incoming and outgoing events both contribute variance. These two estimates are the dimension-uniform heart of the proof.

Exponential control. Let $V_i = \exp(\alpha N E_i)$ with the birth copy number N in the exponent. The identity $E_i(y + h) - E_i(y) = 2Q_i(y, h) + Q_i(h, h)$ and $e^s - 1 \leq s + s^2$ for $|s| \leq 1$ reduce every jump to (H1)–(H3) and give, for uniform constants F, K, B_1, C_1 built from $A, P, R_1, D_0, L, Z, \gamma, \kappa$,

$$\frac{\mathcal{L}V_i}{\alpha V_i} \leq -\lambda N r^2 + N F r + K + \alpha N B_1 r^2 + \frac{\alpha C_1}{N}.$$

Choose α with $\alpha B_1 \leq \lambda/4$, then small γ_0, κ_0 and large N_0 so that $F^2 \leq \lambda^2 \rho/8$ and $K + \alpha C_1/N \leq \lambda N \rho/8$; Young's inequality gives $\mathcal{L}V_i \leq \alpha N(-\frac{\lambda}{2}r^2 + \frac{\lambda \rho}{4})V_i$, negative outside $r^2 < \rho$ with rate at least $\alpha \lambda N \rho/4$ and bounded inside. Neither the exponent nor κ_0 shrinks with k .

Six errors. A nonnegative test function with generator at most β gives $\mathbb{E}V(X_{t \wedge \tau}) \leq V(X_0) + \beta t$, and at first exit through energy h its value is at least $e^{\alpha N h}$; this bounds each tube exit by an energy-gap exponential times $(1 + \lambda \rho t/(2g))$, summed over k modules. The affine drift inequality with killing at the outer exit gives recovery to h_{rec} by $t_0 = 1 + 4b/(\lambda \rho)$ up to $3ke^{-\alpha N(h_{\text{rec}} - M\rho)}$. Early division before t_0 has probability at most $e^{-kN/125}$ when $\gamma_0 Z t_0 \leq 2/5$, and late division after t_1 at most $e^{-9kN/4000}$, by the exponential generating function of the bounded-intensity membrane counter. At division, choose r_{par} and δ with $h_{\text{par}} \leq c_0 r_{\text{par}}^2$ and $h_{\text{par}} + 2Lr_{\text{par}}\delta + L\delta^2 < b$: if the sup norm of the allocation deviation $(D - n/2)/N$ is below δ , the quadratic identity puts *both* daughters in the birth region, since their deviations are opposite; a fair-binomial tail and a union bound over kd coordinates give $2kde^{-N\delta^2/C}$. No product over the two daughters is used. Adding outer exit, early division, failed recovery, parent-tube exit, late division and failed partition gives (12) with constant terms $1 + 1 + 3 + 1 + 1 + 2d$. Coordinatewise rounding of the centres gives integer newborns once N_0 absorbs the rounding error, and Lemma 2.2 gives the lineage statement. $\square$

Corollary 4.3 (Bounded-degree graphs). *If a symmetric graph has maximum degree at most $\Delta \geq 1$ and edge weights at most κ_0/Δ , it satisfies the interaction condition of Theorem 4.2. In particular paths and rings admit the fixed positive edge weight $\kappa_0/2$ for every k .*

4.3 An explicit four-species module

The resident species are A, B, z, H with the thirteen directed channels of Table 3. Growth consumes z and exchange transfers z . The fluid drift is

$$\begin{aligned} \dot{A} &= 6 - 2A + zB + 2e(B - A^2), & \dot{B} &= 27 + A - (1 + z)B - e(B - A^2), \\ \dot{z} &= A - Bz - 16z - 4z^2 + 3H, & \dot{H} &= 16z + 2z^2 - (2 + d_0)H, \end{aligned} \quad (13)$$

with $e = 10^{-5}$, $d_0 = 10^{-4}$. It has two stationary states whose z coordinates lie in the rational intervals $I_0 = [0.99579401232, 0.99579401233]$ and $I_1 = [2.97636724376, 2.97636724377]$; their existence follows from the intermediate value theorem applied to a one-variable residual (Appendix B), so no numerical root is a premise. The label is read from the relative composition: $10n_{iz} - n_{iB}$ is negative throughout the label-0 birth region and positive throughout the label-1 region, so these labels are distinguished by composition and not only by abundance. Appendix B gives the rational quadratic forms P_0, P_1 , the coercivity bounds $\frac{1}{200}|y|^2 \leq E_j(y) \leq \{24, 42\}|y|^2$, the local dissipation $2y^T P_j f(a_j + y) \leq -0.59|y|^2$ on $\|y\|_\infty \leq 1/400$, and the literal generator estimate. With $A_0 = 1/512000000$ and levels $(h_{\text{rec}}, h_{\text{par}}, b, h_{\text{out}}) = (A_0, 2A_0, 4A_0, 16A_0)$, the resulting constants are

$$N_{\min} = 1.4 \times 10^{20}, \quad c = \frac{1}{1.024 \times 10^{21}}, \quad C(\gamma) = 2703 + \frac{11}{5\gamma}, \quad \varepsilon = kC(\gamma)e^{-cN}, \quad (14)$$

valid for $0 < \gamma \leq 10^{-11}$, $0 \leq \kappa \leq 10^{-11}$, $N \geq N_{\min}$ and (10). At birth each module holds between $N/2$ and $140N$ residents. The certified physical generation lasts at most $k(2688 + 11/(5\gamma)) \leq 3k/\gamma$.

Table 3: The four-species module. Reference coefficients enter (9); the two bimolecular self-reactions use falling factorials.

Reaction pair or channel	Forward, reverse coefficients
$A \rightleftharpoons B + z$	1, 1
$z \rightleftharpoons H$	16, 1
$2z \rightleftharpoons H$	2, 1
$\emptyset \rightleftharpoons A$	6, 1
$\emptyset \rightleftharpoons B$	27, 1
$B \rightleftharpoons 2A$	10^{-5} , 10^{-5}
$H \rightarrow \emptyset$	10^{-4}

For this source the count process is constructed on the unbounded state space without artificial cutoffs. With $V(n, m) = m + \sum_{i,a} n_{ia}$, exchange and consuming growth preserve V and the resident reactions give $\mathcal{L}V \leq 33V$; stopping at mass level R and time T , $\mathbb{P}(V \text{ reaches } R \text{ by } T) \leq V(X_0)e^{33T}/R$, and letting $R \rightarrow \infty$ proves nonexplosion. The certified events of Theorem 4.2 are then events of this actual process: on a finite safe domain with division states frozen, the survival-history expectation of a bounded terminal payoff satisfies the first-jump renewal equation, whose unique solution coincides with the Poissonized finite-state expectation; no bound on rates outside the domain is needed because the continuation there is zero.

A second, two-species instance ($\emptyset \rightarrow X$, $Y \rightarrow 2X$, $2X \rightarrow X + Y$, $X + Y \rightarrow Y$, $X \rightarrow \emptyset$ with coefficients 6, 1, 6, $1/6$, 11 and centres $(1, 6)$, $(3, 54)$) also satisfies Assumption 4.1 with explicit rational forms, and has a relative readout $n_Y - 12n_X$; it is not identified with the four-reaction minimal bistable system of Wilhelm [15]. It also exhibits a dense-coupling obstruction: a module near $(3, 54)$ all of whose neighbours are near $(1, 6)$ has no stationary state in that mixed box once its exchange row sum reaches $1/100$, so a complete graph with any fixed positive edge weight eventually violates the interaction condition. This excludes a stationary mixed box; it is not a theorem of stochastic impossibility.

4.4 Immediate copying: matching upper and lower bounds

Theorem 4.2 concerns a certified generation with tubes and a recovery gate. The natural copying event is weaker: the first actual division occurs and both daughters are admitted. For the four-species source the two are related by inclusion, and the immediate event also has an unconditional lower bound on its failure probability.

Let X be the nonexplosive count process frozen when m reaches $2kN$, let J be the index of the first jump reaching that membrane count ($J = \infty$ if none), and at X_J draw the complementary partition. This defines a kernel K_{imm} from admitted newborns to pairs of admitted newborns or a failure symbol. Each admitted daughter has at most $140N$ residents per module, so two admitted daughters force $\sum_a n_{ia} \leq 280N$ for every parent module i ; the observed parent may be projected to this finite box, mapping everything else to failure, without losing any genuine success.

Lemma 4.4 (Uniform partition obstruction). *Set $q = 2^{-280N}$ and $r = (1 - q)^k$. For every parent, however produced, the probability that both daughters return immediately is at most r ; hence $K_{\text{imm}}(n, \text{failure}) \geq 1 - r$ for every admitted newborn n .*

Proof. If some parent module holds more than $280N$ molecules, both daughters cannot be admitted. Otherwise let $b_i \leq 280N$ be the count in module i ; daughter A is empty in module i with probability $2^{-b_i} \geq q$, and these events are independent across modules conditional on

the parent, so simultaneous nonemptiness has probability $\prod_i (1 - 2^{-b_i}) \leq r$. Every admitted region has a nonempty module at each coordinate, so immediate return is a subevent. Nothing requires the parent to remain in a tube. $\square$

Lemma 4.5 (Certificate inclusion). *For the four-species source, certified generation success is a subevent of immediate first-division success on a common probability space, and $K_{\text{imm}}(n, \text{failure}) \leq \varepsilon$ for every admitted newborn.*

Proof. On the certified event there is no division by t_0 , the recovery test passes, the retained path reaches the threshold by $t_0 + t_1$, and the partition returns both daughters. The membrane increases by single units and is frozen at the threshold, so the parent observed at the second deadline is exactly the first-division parent; applying the same allocation variables in both observations makes every certified output an immediate output. It remains to identify the probability of this common-path certificate with the two-phase law of Theorem 4.2. At the deterministic time t_0 , condition on the finite jump history and the current state x ; the residual holding time is exponential with the rate of x by memorylessness, the next mark is drawn from the rates at x , and subsequent marks and holding times are fresh, so the conditional future path has the law of a new trajectory from x . Nonexplosion ensures finitely many past jumps almost surely; partitioning over the jump index, integrating the holding times and applying Tonelli (all continuation payoffs lie in $[0, 1]$) identifies the common-path expectation with the first safe-interval expectation composed with the recovery gate and the second safe-interval expectation. Inclusion then gives the bound. $\square$

The deterministic-time argument in Lemma 4.5 is a conventional proof; the certified two-phase law and the necessity bound are compiled, and Appendix A records that the inclusion itself is not.

Theorem 4.6 (Immediate-copy scaling). *Under (14) and the stated source restrictions, for every word, admitted newborn and $G \geq 0$,*

$$1 - G\varepsilon \leq \mathbb{P}(A_G) \leq (1 - q)^{kG}, \quad (15)$$

and $\mathbb{P}(A_G) \geq (1 - \varepsilon)^G$ when $\varepsilon \leq 1$. Consequently, for $0 < \eta < 1$ and $G \geq 1$, the integer choice

$$N \geq \max\left\{N_{\min}, \left\lceil \frac{1}{c} \log \frac{kGC(\gamma)}{\eta} \right\rceil \right\} \quad (16)$$

gives $\mathbb{P}(A_G) \geq 1 - \eta$, while $\mathbb{P}(A_G) \geq 1 - \eta$ requires

$$N \geq \frac{1}{280 \log 2} \log \frac{kG}{-\log(1 - \eta)}. \quad (17)$$

For the complete binary family, (15)–(17) hold with $D_G = 2^G - 1$ in place of G .

Proof. Given any successful past, the designated daughter is an admitted newborn of the original word, so Lemmas 4.4 and 4.5 apply to its next attempt; conditional multiplication gives the geometric bounds and Bernoulli's inequality the linear one. Substitution gives (16). For (17), $(1 - q)^{kG} \leq e^{-kGq}$, so $1 - \eta \leq e^{-kG2^{-280N}}$; take logarithms. The family statement is Proposition 2.3 together with the same conditional upper iteration on the joint kernel. $\square$

Proposition 4.7 (Unequal allocation and total budget). *If a parent holds $b_i \geq 0$ molecules in module i with $\sum_i b_i \leq B$, then under independent fair allocation $\prod_i (1 - 2^{-b_i}) \leq (1 - 2^{-B/k})^k$. Hence if every accepted newborn has at most M residents in total and every module nonempty, reliability at least $1 - \eta$ across $D \geq 1$ inspected divisions requires*

$$M \geq \frac{k}{2 \log 2} \log \frac{kD}{-\log(1 - \eta)}. \quad (18)$$

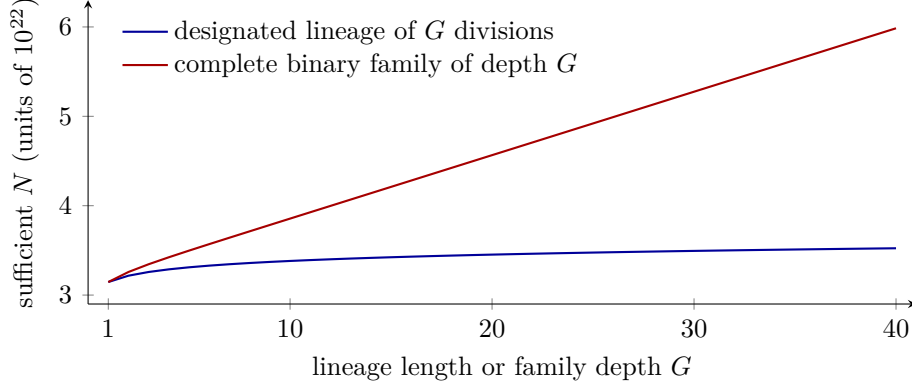


Figure 2: The sufficient copy number (16) for the four-species source at $k = 1$, $\gamma = 10^{-11}$, $\eta = 0.01$, against lineage length or family depth G . These are evaluations of a proved bound, not measured performance; the necessary bound (17) is of order 10^{20} on this scale. The two curves differ only in the horizon term, $\log G$ against $\log(2^G - 1)$.

Proof. $x \mapsto \log(1 - 2^{-x})$ is increasing and concave, so $\sum_i \log(1 - 2^{-b_i}) \leq k \log(1 - 2^{-B/k})$. Two accepted daughters force parent total at most $2M$, so every attempt succeeds with conditional probability at most $(1 - 2^{-2M/k})^k$, and D attempts with probability at most $e^{-kD2^{-2M/k}}$. Rearranging gives (18). Unequal distribution of a fixed budget cannot evade the obstruction. $\square$

Proposition 4.8 (Lifetime at fixed redundancy). *For $0 < \varepsilon < 1$, let T be the index of the first unsuccessful copying attempt in a designated lineage. Then $1/\varepsilon \leq \mathbb{E}T \leq 1/(1 - r)$ and $\mathbb{P}(T < \infty) = 1$.*

Proof. $\mathbb{P}(T > g) = \mathbb{P}(A_g) \in [(1 - \varepsilon)^g, r^g]$ with $r < 1$; sum the tails. These count generations, not physical time. $\square$

4.5 Orders of the molecular requirement

For fixed positive γ and fixed local quality, and $0 < \eta \leq 1/2$, Theorem 4.6 gives the sufficient and necessary orders

$$\text{designated lineage: } N = \Theta(\log(kG/\eta)), \quad \text{complete family: } N = \Theta(G + \log(k/\eta)), \quad (19)$$

the second because $\log(2^G - 1) = G \log 2 + O(1)$. The fixed floor $N_{\min}$ remains in both sufficient statements, and $C(\gamma)$ grows as $1/\gamma$: slower growth lengthens the exposure to fluctuations, which worsens the sufficient bound without proving actual loss of heredity. Figure 2 evaluates (16) for the four-species source.

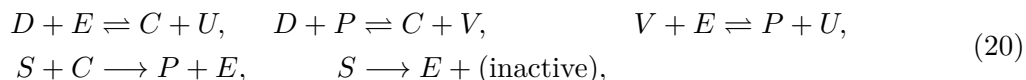
Three costs are hidden by a molecule-only asymptotic statement and should be read together with (19). Physical reaction coefficients are rescaled with k by (9), the species count is dk , and the certified generation time is $O(k/\gamma)$. The necessity statements concern *immediate* return under fair independent allocation with nonempty modules; a failure to satisfy admission at birth need not remain a failure when post-birth repair is allowed, as in the protocol of Section 5, and neither (17) nor (18) applies to that protocol or to the constructed source of Section 3 without a separate argument. The sufficient constants of (14) are conservative existence guarantees; at $\gamma = 10^{-11}$ the threshold for a single lineage division at $\eta = 0.01$ is about 3.1×10^{22} , and no claim is made that these are laboratory scales. The value of the result is the precisely stated trade-off and its uniformity in k , not the constants.

5 A nominal published chemistry under a split–refill–recovery protocol

The third result applies the operational definition of Section 2 to a reaction mechanism that has been realized experimentally: the thiol–thioester network of Semenov et al. [10], which is bistable in a continuously stirred tank reactor (CSTR). We analyse the nominal kinetic model reconstructed from that paper, at its reported rate constants and feed, and prove that a prescribed protocol of splitting, refilling and recovering copies the two stable states with joint daughter return probability at least 0.992. Three features distinguish this from Section 3: the volume is macroscopic (about 40 μL), so the count space is enormous and the proof uses moving quadratic metrics rather than an enumerated certificate; the protocol allows repair after splitting, so the success event is return after a fixed recovery time; and the model, not the laboratory system, is what the theorem is about.

5.1 The nominal model

The tracked species, in the order (S, C, P, E, D, U, V, I) , are the thioester AlaSEt, cysteamine, the thiol-bearing ligation product, ethanethiol, cystamine, the two mixed disulfides and the inhibitor maleimide. The reconstructed reactions are



together with the inhibition of C , P and E by I to inactive products. The six directional disulfide-exchange coefficients are $0.65 \text{ M}^{-1}\text{s}^{-1}$, ligation is $0.41 \text{ M}^{-1}\text{s}^{-1}$, hydrolysis is $9.26 \times 10^{-6} \text{ s}^{-1}$ and inhibition is $150 \text{ M}^{-1}\text{s}^{-1}$. The CSTR dilution rate is $s_v = 0.002 \text{ s}^{-1}$ and the feed is

$$c_f = (0.05, 0, 0, 0, 0.1, 0, 0, 0.00231) \text{ M}.$$

Autocatalysis arises because the thiols C , P , E liberate further thiol from cystamine and from the thioester, while maleimide in the feed removes thiol at low concentration; the readout is total free thiol $C + P + E$. The reported conditions are pH 7.5, 25°C and 0.5 M phosphate; they are provenance, not a certified uncertainty interval. The reconstructed deterministic field reproduces the published bistability at three flow rates; at $s_v = 0.002$ the model’s high thiol state is near 35 mM whereas the experimental comparison in the source is near 23 mM, a discrepancy we return to in Section 5.4.

The stochastic model at volume V has $\Omega = N_A V$ molecules per molar. Distinct-reactant propensities are $kn_i n_j / \Omega$, unimolecular propensities kn_i , feed propensities $\Omega s_v (c_f)_i$ and out-flow propensities $s_v n_i$; all 11 chemical, 8 feed and 8 outflow channels are retained, and the deterministic invariant planes are not imposed on the fluctuations. Inactive products, water and buffer lie outside the eight-species count.

5.2 Protocol and return regions

Set $\Omega = 24 \times 10^{18}$, so $V = \Omega / N_A = 39.85294 \mu\text{L}$. One cycle is:

1. start from a parent in the return region of its label at volume V ;
2. split fairly into two half-volume daughters by one complementary allocation vector;
3. refill each daughter to V with fresh feed: the added count of species i is Poisson with mean $\Omega (c_f)_i / 2$, independently across species and daughters;
4. run each daughter for 500 s under the nominal CSTR law at volume V , with a per-daughter feed quota of $K = 12 \times 10^{18}$ active molecules and a coordinate cap of $B = 20 \times 10^{18}$; exceeding either is absorbing failure;

5. inspect both terminal states; daughter A continues, and the feed quota is renewed.

This is a prescribed external protocol that permits repair after splitting. It is not immediate newborn return, and the necessity bounds of Section 4.4 do not transfer to it.

The two return regions are rational ellipsoids

$$B_\ell = \{n : (n/\Omega - z_\ell)^T P_\ell (n/\Omega - z_\ell) \leq \eta_\ell\}, \quad \eta_0 = 6 \times 10^{-12}, \quad \eta_1 = 9 \times 10^{-14}, \quad (21)$$

within the finite count domain. The centres z_ℓ and positive matrices P_ℓ are exact rationals extracted from the terminal pieces of the checked polynomial certificates; they are distributed with the sources as `terminal_regions.json`, and Appendix C lists their approximate values. The thiol values at the centres are about $1.4488 \mu\text{M}$ (low) and 34.5098 mM (high), and the certified coordinate radii are below $4.703 \mu\text{M}$ and $1.626 \mu\text{M}$ respectively. Exact endpoint checks show that the fixed decoder $n_C + n_P + n_E > 1.2 \times 10^{17}$, that is $C + P + E > 5 \text{ mM}$ at V , separates the entire regions; so does the volume-free ratio rule $100(n_C + n_P + n_E) > \sum_i n_i$, since thiol is below 0.001 of the active total throughout the low region and above 0.17 throughout the high region. The regions therefore admit a relative-composition readout, unlike the boxes of Section 3; this is a consequence of their geometry and does not extend the theorem to other volumes or preparations. The explicit preparations $[\Omega z_\ell]$ lie in the regions by exact membership checks.

Theorem 5.1 (Nominal split–refill–recovery copying). *For each label and every parent in (21), the nominal stochastic protocol returns both daughters to the same region with probability at least $124/125 = 0.992$. Ten inspected cycles succeed with probability at least $0.992^{10} > 0.922$.*

5.3 Proof by moving rational metrics

Write $u = n/\Omega$ for the concentration vector. The count space at $\Omega = 2.4 \times 10^{19}$ cannot be enumerated, so the certificate is a time-dependent quadratic energy along the deterministic recovery.

Reference curves and metrics. Numerically proposed centre curves $z(t)$ and positive matrices $P(t)$ are represented by degree-16 rational polynomials on 16 (low) and 22 (high) time pieces covering exactly $[0, 500]$, with centres and metrics joining exactly at the piece boundaries. They need not solve the differential equations: the residual $P(f(z) - \dot{z})$ of the field f and the matrix residual of $\dot{P} + J^T P + P J + I$, with J the Jacobian at z , are explicitly bounded on each piece by rational coefficient inequalities. Rational congruence and Gershgorin bounds certify positive definiteness. On each piece define

$$D(t, n) = \frac{(u - z(t))^T P(t) (u - z(t))}{\eta}.$$

The matrix inequality, the residual bounds, the exact quadratic remainder of the field, and the literal finite-count jump increments of all 27 channels yield a generator bound for the sixth power: $(\partial_t + \mathcal{L})D^6 \leq 10^{-9}$ throughout $D \leq 1$, for both labels. The sixth power, rather than D itself, is what balances the initial disturbance from splitting against the drift cost of 500 seconds. Table 4 lists the sufficient constants.

Finite feed inventory. Let C_t be the cumulative feed input of a daughter, with initial refill mean $m_0 = \Omega \sum_i (c_f)_i / 2 = 1.82772 \times 10^{18}$ and inflow rate $\lambda = \Omega s_v \sum_i (c_f)_i$. The centred inventory energy

$$E(t, n, C) = D(t, n)^6 + \frac{4(C - m_0 - \lambda t)^2}{K^2}$$

Table 4: Sufficient scalar bounds in the nominal recovery certificate. L_0 bounds the initial metric’s absolute row sums; v_s bounds the allocation-plus-refill variance coefficient; r bounds the inherited parent energy after half dilution. These are certificate constants, not measured kinetic uncertainties.

Quantity	Low	High
η	6×10^{-12}	9×10^{-14}
L_0	1601	3818
v_s	0.115	0.114
r	0.09	0.26
Sixth-power drift bound	10^{-9}	10^{-9}
Polynomial time pieces	16	22

has exact inventory drift $4\lambda/K^2$: predictable consumption cancels and only fluctuations around it are charged. The mean input through 500 s is $5.48316 \times 10^{18} < K/2$, so exhaustion forces $E \geq 1$. The coordinate cap cannot intervene inside the controlled domain, because every admitted parent has total active concentration at most 0.2 M and the parent plus the entire quota is below B .

Global comparison through a concave cap. Let $\phi(e) = 2e - e^2$ for $0 \leq e \leq 1$ and $\phi(e) = 1$ for $e \geq 1$, which is C^1 , concave, bounded by 1, and equal to 1 wherever the certificate has failed. The tangent inequality $\phi(b) \leq \phi(a) + \phi'(a)(b - a)$ and $0 \leq \phi' \leq 2$ turn the local drift estimate on $E \leq 1$ into a global bound $(\partial_t + \mathcal{L})\phi(E) \leq 2 \cdot (10^{-9} + 4\lambda/K^2)$ on the whole finite model, including the absorbing failed state, without any estimate outside the tube. The indicator of terminal failure (outside the region, quota exhausted, or cap exceeded) is bounded by $\phi(E)$ at time 500. Applying the actual finite-time law on each polynomial piece and composing the joins, a daughter that starts with energy E_0 fails with probability at most $\mathbb{E}\phi(E_0) + 1000(10^{-9} + 4\lambda/K^2)$.

Allocation and refill. For species i let $\xi_i = D_i - n_i/2 + Q_i - \mathbb{E}Q_i$ with $D_i \sim \text{Bin}(n_i, 1/2)$ the allocation and $Q_i \sim \text{Pois}(\Omega(c_f)_i/2)$ the refill. Its exact centred second moment is $v_i = n_i/4 + \mathbb{E}Q_i$ and its fourth moment is $3v_i^2 + \mathbb{E}Q_i - n_i/8$. The exact endpoint comparison $P(0)/4 \preceq rP(500)$ bounds the deterministic displacement of a parent anywhere in the region after halving and refilling by r in normalized energy. With $m_1 = L_0 v_s / (\Omega\eta)$ and $m_2 = L_0^2 \eta^{-2} (24v_s^2/\Omega^2 + 8v_s/\Omega^3)$, Taylor’s inequality for $g(D) = \phi(D^6)$, whose derivative bounds are 4 and 12 on the relevant range, gives $\mathbb{E}g(D_{\text{newborn}}) \leq \phi(r^6) + (4 + 48r)m_1 + 12m_2$; the moment bounds use only marginal moments and Cauchy’s inequality, so no cross-species independence beyond the allocation law itself is required. Altogether each daughter fails with probability at most

$$\phi(r^6) + (4 + 48r)m_1 + 12m_2 + \frac{64m_0}{K^2} + 1000\left(10^{-9} + \frac{4\lambda}{K^2}\right) \leq \frac{1}{250}, \quad (22)$$

where $64m_0/K^2$ charges the initial refill variance of the inventory term. All comparisons are exact rationals. The exact prospective values of twice the per-daughter bound are about 2.5×10^{-5} (low) and 7.9×10^{-3} (high); the theorem states the simpler uniform bound.

Both daughters. Daughter B’s allocation is $n - D$, and the reflection $d \mapsto n - d$ shows that its marginal law equals daughter A’s, so (22) applies to each. The union bound gives joint failure at most $2/250$, retaining the dependence between siblings. Integrating the actual recovery transitions of daughter A against the joint allocation-and-refill law, weighted by the

Table 5: Deterministic sensitivity diagnostics at $\pm 0.1\%$ changes of one rate group, from the nominal centre after mean split and refill. These are floating-point diagnostics, not stochastic bounds and not empirical uncertainty intervals. Every run retains the correct thiol decoder.

Changed group	Low terminal D	High terminal D
Disulfide exchange	0.145	2.17×10^4
Ligation	7.0×10^{-4}	1.43×10^6
Hydrolysis	7.44	20.8
Inhibition	0.152	0.028
Dilution rate	3.08	1.09×10^6

success of daughter B, gives the offspring kernel, whose row sum is exactly the joint return probability; Lemma 2.2 proves Theorem 5.1.

5.4 Preparation tolerance and rate sensitivity

Passing the thiol decoder is much weaker than lying in the return ellipsoid, and the ellipsoids are narrow. For a preparation error e in concentration units, a sufficient simultaneous tolerance is

$$\|e\|_\infty \leq b, \quad b^2 \sum_{i,j} |(P_\ell)_{ij}| \leq \eta_\ell, \quad (23)$$

since $e^T P_\ell e \leq \sum_{i,j} |(P_\ell)_{ij}| |e_i| |e_j|$. Exact evaluation of the stored matrices gives $b = 16.366$ nM (low) and $b = 1.788$ nM (high). These are sufficient cube tolerances about the exact centres; other directions permit larger deviations, and no claim is made that they are experimentally achievable.

To quantify parameter sensitivity without a new certification campaign, we integrated the same deterministic field for the 500-second recovery from the nominal centre after mean split and refill, changing one kinetic group at a time by $\pm 0.1\%$ and $\pm 1\%$ and keeping the accepted terminal ellipsoid fixed. Table 5 reports the larger terminal energy D of the two signs at 0.1% ; return requires $D \leq 1$. The integration used a BDF method with relative tolerance 2×10^{-10} and absolute tolerance 2×10^{-14} M, and a tighter nominal replay moved the endpoints by at most 6.4×10^{-13} M (low) and 1.2×10^{-11} M (high).

Several 0.1% changes move the high-state endpoint far outside the nominal ellipsoid while the decoder still reads correctly. A deterministic endpoint outside the old region does not prove loss of bistability, loss of decoder memory, or stochastic failure of a suitably recentred protocol: the regions are centred on the nominal trajectory, and a perturbed model has its own recovery trajectory around which a new certificate could be built. What the table shows is that Theorem 5.1 certifies the nominal model and cannot be read as robustness to unspecified kinetic uncertainty. Two source-motivated variations make the same point: replacing the nominal 50/100 mM feeds by the reported experimental 47/92 mM moves the high-state thiol to about 32.1 mM, and replacing the flow-fitted hydrolysis by the batch value $7 \times 10^{-6} \text{ s}^{-1}$ also leaves both nominal regions. Together with the 35-versus-23 mM discrepancy already present in the source, this is the concrete reason to present Theorem 5.1 as a nominal-model prediction and not as an experimental fidelity claim. An experimentally supported statement would require measured rates and feeds at the stated conditions, a resolution of the high-state discrepancy, and certificates regenerated at those parameters.

6 Discussion

What is established. Reliable copying of a chemical state is here a property of a complete stochastic cycle, proved uniformly over an admitted region, rather than a consequence of deterministic multistability or a favourable fraction of simulated trajectories. That uniformity is what makes iteration legitimate: Lemma 2.2 needs no independence between siblings or generations, only that every successful daughter is again a state to which the one-cycle theorem applies. Three quite different chemistries meet this standard. The constructed source shows that the standard can be met with fewer than a thousand resident molecules per module when the environment is ideally controlled, and that two modules can be coupled through shared growth and cross-catalysis without losing it. The redundancy theorem shows that local recovery against bounded total incident interaction is the mechanism that keeps the error exponent and the interaction allowance independent of the number of modules, and that for immediate return the logarithmic dependence on the number of modules, the lineage length and the target error is sharp, with the family depth entering linearly. The Semenov result shows that the proof machinery reaches a published eight-species mechanism at a macroscopic volume, through moving rational metrics instead of state enumeration.

The role of external operations. None of the three protocols is autonomous. The constructed source relies on an ideal controller for volume, pool activities and post-division reset; the redundancy model prescribes division at a membrane threshold and equal daughter volumes; the Semenov protocol imposes splitting, refilling and a fixed recovery interval. These operations are part of the theorem statements, and the resource costs they entail are stated separately from the resident counts: reservoir capacities of order 0.17 pmol per pool and cycle for the constructed source, a growth deadline of order k/γ and rate rescaling with k for the redundancy theorem, and a feed quota of 1.2×10^{19} molecules per daughter for the Semenov protocol. Neither result should be read as a claim about vesicle reproduction, selection or evolution; what the results supply to those questions is a precise, checkable notion of copying fidelity and three worked instances of proving it.

Immediate return versus repair. The two copying criteria in this paper differ in a way that matters for lower bounds. Immediate return requires admission at birth and is subject to the partition obstruction of Lemma 4.4, which no chemistry can evade under fair independent allocation. Return after recovery permits the chemistry to repair allocation noise, and the constructed source’s 874-molecule bound and the Semenov result both exploit this in different ways: the former through the certified recovery built into the birth-to-division interval, the latter through an explicit 500-second recovery. A single-model comparison of the two criteria, with the same chemistry and the same regions, would sharpen the picture and is not carried out here.

Mathematical limits. The constants of the redundancy theorem are existence bounds, many orders of magnitude above anything physically meaningful, and the sufficient prefactor degrades as $\gamma \downarrow 0$. The constructed source’s regions are boxes at a fixed volume and store concentration states, not compositions (Proposition 3.5); its theorem covers two modules and does not extend to more without new local certificates, since the controlled-neighbour argument admits one other module. The Semenov certificate holds at nominal parameters and in narrow regions; Section 5.4 shows that its regions do not survive 0.1% changes in several rate groups, so a laboratory statement would require recentred certificates at measured parameters. All three results concern fair complementary allocation; unequal or correlated division laws, unwanted side reactions and shared-resource fluctuations are outside their scope, and the allocation

obstruction of Section 4.4 shows that some division laws destroy copying regardless of the chemistry.

Method. The methodological point common to the three results is the separation between proposing a certificate and validating it. Numerical solvers proposed backward vectors on a finite grid and polynomial metrics along a trajectory; exact integer or rational arithmetic then checked the small set of inequalities that the probability theorem actually consumes, and Lean 4 checked the assembly of those inequalities into statements about the literal stochastic laws. A successful simulation, a small residual or a stable Jacobian was never accepted as closure. This division of labour is what allowed a 2.2×10^7 -row certificate and a 38-piece polynomial certificate to become theorems with explicit trust boundaries (Appendix A), and it transfers to other mechanisms for which one wants a fidelity guarantee rather than an estimate.

A Formal verification and reproducibility

The finite-model theorems of Sections 3 and 5, and the certified-generation kernel, necessity bound and iteration lemmas of Section 4, are checked in Lean 4 (toolchain `leanprover/lean4:v4.30.0`) against a frozen Mathlib manifest [3, 12]. Verification treats compiler warnings as errors and rejects any declaration depending on `sorry`. Each root file has a machine-readable receipt recording the toolchain, the SHA-256 hashes of the root and of every transitive dependency, the compile time and, for the requested declarations, the elaborated statement and the axioms it depends on. Every declaration named below depends only on the three standard axioms `Classical.choice`, `Quot.sound` and `propext`, plus, where indicated, native-evaluation axioms introduced by `native_decide` for the exact integer and rational table checks. That trust boundary includes Lean’s compiled evaluator for those checks; it is not a claim that every arithmetic step was reduced by the kernel alone.

Paper statement	Lean declarations (namespace <code>CompositionalMemory</code>) and status
Theorem 3.1, first division ≥ 0.9901 ; ten cycles ≥ 0.9 ; positive rectangle	<code>ReversibleCertificate.actual_first_division</code> , <code>ReversibleCertificate.ten_cycles</code> , <code>ReversibleCertificate.positive_window</code> . Strict receipt: 174 transitive dependencies, 281 s; 58 axioms (3 standard, 55 native) for the 54 integer table modules and the terminal binomial sums. The formal statement is about the actual monitored count source and its offspring kernel; the parameter and box hypotheses are explicit.
Lemma 3.2 and its lifting to the counter state	<code>finite_deadline_residual_certificate</code> ; <code>FiniteQuotaTransport.killed_quota_deadline</code> , <code>FiniteQuotaTransport.quota_project_generator</code> .
Lemma 3.3 and the quota constant 2×10^{-7}	<code>FiniteEventQuota.quota_exponential_bound</code> , <code>FiniteEventQuota.quota_concrete_constant</code> , <code>FiniteEventQuota.event_quota_deadline</code> ; no native evaluation.
Proposition 3.4	<code>ReversibleDivisionInventory.reversible_complete_pool_supply_budget</code> , <code>...disposal_budget</code> , <code>reversible_complete_inventory_numbers</code> ; <code>ReversibleDimensionalRates</code> for the physical rate table.
Proposition 3.5; the powers 0.9901^{10} , 0.992^{10} ; capacity in moles	Exact rational checks in <code>check_constants.py</code> (distributed with the sources); the ten-cycle bounds ≥ 0.9 are compiled, the sharper powers are elementary.

Paper statement	Lean declarations (namespace <code>CompositionalMemory</code>) and status
Theorem 4.2	<code>general_uniform_word_interval</code> : primitive estimates, choice of $\gamma_0, \kappa_0, c_*, N_0, t_0$, and the finite certified-generation law, for arbitrary k, d , graph and word.
Four-species instance, constants (14), nonexplosion, two-phase physical identity, certified lineage bounds	<code>source_actual_word_scaling</code> . Source certificates in <code>FiniteCopy.QuadraticCertificates</code> , <code>FiniteCopy.LocalGenerator</code> ; second instance and obstruction in <code>TwoSpeciesApplication</code> , <code>TwoSpeciesDenseObstruction</code> , <code>ActivityScope</code> .
Lemma 4.4; first-division kernel; finite projection	<code>sourceFirstParentLaw</code> , <code>sourceImmediateLaw</code> , <code>project_parent_preserves_return</code> , <code>source_immediate_failure_lower</code> , <code>source_immediate_lineage_upper</code> .
Lemma 4.5	<i>Conventional proof</i> (deterministic-time memorylessness and Tonelli). The certified two-phase law and the immediate kernel are both compiled, but no compiled declaration asserts the inclusion between them. Theorem 4.6's upper bound on $\mathbb{P}(A_G)$ and the necessity bounds (17) are compiled; its lower bound uses this lemma.
Proposition 4.7	<code>unequal_allocation_bound</code> (real nonnegative counts); the conservation step and (18) are elementary.
Lemma 2.2, Propositions 2.3 and 4.8, Corollary 4.3	<code>word_lineage_multiplicative_lower</code> , <code>family_fidelity_upper</code> , <code>lifetime_tail_bounds</code> , <code>bounded_degree_row</code> ; abstract finite-kernel statements, specialized in the text.
Theorem 5.1; explicit preparations; decoder	<code>Semenov.both_daughters_bound</code> (uniform joint return $\geq 124/125$), <code>Semenov.measured_ten_divisions</code> , <code>Semenov.good_parent_decoder</code> , <code>Semenov.measured_explicit_first_division</code> , <code>Semenov.measured_explicit_ten_divisions</code> . Root <code>SemenovMeasuredEncoding</code> : 130 dependencies; 125 axioms (3 standard, 122 native) for the 38 polynomial pieces, endpoint geometry and encoder membership. Sibling dependence is retained (<code>SemenovComplementaryAllocation</code>); the union bound, not a product, is used.
Section 5.4	Cube tolerances (23): exact rational evaluation. Sensitivity table: floating-point ODE integration, explicitly not a certificate.

Trajectory-level meaning. For the constructed source and the Semenov protocol the compiled statements are about finite jump models that are, by construction, the actual protocols: the domain monitors, quota and absorbing caps are part of the stated protocol, so no separate nonexplosion or localization argument is needed to interpret them. For the four-species source of Section 4 the compiled statement includes nonexplosion of the unrestricted count process and the identification of the certified two-phase payoff with the finite Poissonized law; Lemma 4.5 is the only step of Section 4 proved conventionally.

Status of the receipts at the time of writing. The two roots of Sections 3 and 5 were re-audited for this manuscript: every one of their 174 and 130 recorded dependency hashes matches the current source tree, so the compiled statements refer to the sources as distributed. The entry point for Section 4, `Publication.lean`, was first strictly verified on 10 September 2026. One of its dependency modules was later overwritten in the shared repository by a same-named module of the constructed-source development; the original was recovered byte-exactly (its hash matches

the 10 September receipt), reinstated as `ScalingLocalGenerator.lean`, and its two importers reprinted. `Publication.lean` was then re-verified on 13 September 2026 on the current tree: strict pass, 316 transitive dependencies, all recorded dependency hashes matching the distributed sources, and the four declarations `general_uniform_word_interval`, `source_actual_word_scaling`, `source_immediate_failure_lower` and `unequal_allocation_bound` depending only on the three standard axioms, with no native evaluation.

Reproduction. All commands run from the repository root through the proof bridge, which fixes the working directory to the frozen Mathlib project and does not update dependencies:

```
python scripts/verify_proof.py proofs/CompositionalMemory/ReversibleCertificate.lean
python scripts/verify_proof.py proofs/CompositionalMemory/SemenovMeasuredEncoding.lean
python scripts/verify_proof.py proofs/CompositionalMemory/Publication.lean
```

each with `-timeout 600` and, optionally, `-declaration <name>` to export an authenticated declaration interface. The last file is the entry point for Section 4. The numerical certificates were proposed by scripts in the research workspaces (a layered backward solve on the finite grid for Section 3; a Chebyshev cover of the recovery trajectory with exact rational rounding for Section 5); their outputs are checked, not trusted, by the Lean roots. The manuscript sources include `check_constants.py`, which re-derives every constant printed in the text in exact arithmetic and generates the data of Figure 2, and `terminal_regions.json` with the exact Semenov return regions.

B Certificates for the four-species module

This appendix records the exact data behind Section 4.3: the stationary states, the quadratic forms, the local dissipation and generator estimates, and how the constants (14) arise. Coordinates are ordered (A, B, z, H) and every terminating decimal denotes its exact rational value.

Stationary states without numerical roots. Setting (13) to zero and eliminating A, B, H in favour of z gives

$$\begin{aligned} B(z) &= \frac{60}{z+2}, & H(z) &= \frac{16z+2z^2}{2+d_0}, \\ K(z) &= \frac{2(1+2d_0)z^2-16(1-d_0)z}{2+d_0}, & A(z) &= zB(z)+K(z), \end{aligned}$$

and stationarity reduces to the scalar equation $27 - (1+e)B(z) + K(z) + eA(z)^2 = 0$. Its left side has opposite signs at the two rational endpoints of each of $I_0 = [0.99579401232, 0.99579401233]$ and $I_1 = [2.97636724376, 2.97636724377]$, and all denominators are positive there, so the intermediate value theorem supplies a stationary state $a_j = (A(z_j), B(z_j), z_j, H(z_j))$ with $z_j \in I_j$. Rational enclosures used in the estimates are $a_0 \in [12, 14] \times [20.0280, 20.0282] \times I_0 \times [8, 10]$ and $a_1 \in [20, 22] \times [12.0569, 12.0571] \times I_1 \times [32, 34]$. The sign of $10z - B$ is negative on the whole label-0 birth region and positive on the whole label-1 region, with margins that survive the energy radius $\|y\|_\infty \leq 1/400$ below.

Quadratic forms. With $E_j(y) = y^T P_j y$,

$$P_0 = \begin{pmatrix} 1.115801 & -0.613708 & 2.346047 & 3.444885 \\ -0.613708 & 1.111499 & -2.339313 & -3.434384 \\ 2.346047 & -2.339313 & 6.767757 & 10.178806 \\ 3.444885 & -3.434384 & 10.178806 & 15.517433 \end{pmatrix},$$

$$P_1 = \begin{pmatrix} 0.846084 & -1.159490 & 2.352853 & 3.510058 \\ -1.159490 & 4.333988 & -6.781643 & -10.233028 \\ 2.352853 & -6.781643 & 11.398763 & 17.222276 \\ 3.510058 & -10.233028 & 17.222276 & 26.082111 \end{pmatrix}.$$

Exact rational LDL^T decompositions of $P_j - I/200$ and of $U_j I - P_j$ with $(U_0, U_1) = (24, 42)$ give

$$\frac{1}{200}|y|_2^2 \leq E_0(y) \leq 24|y|_2^2, \quad \frac{1}{200}|y|_2^2 \leq E_1(y) \leq 42|y|_2^2.$$

Local dissipation. For the Jacobian $J_j = Df(a_j)$, rational substitution of the enclosures bounds each of the sixteen entries of $P_j J_j + J_j^T P_j + I$ in absolute value by $1/100$, so the linearized dissipation satisfies $2y^T P_j J_j y \leq -0.84|y|_2^2$. The exact nonlinear remainder of (13) at displacement y is

$$R(y) = (-2ey_A^2 + y_B y_z, ey_A^2 - y_B y_z, -y_B y_z - 4y_z^2, 2y_z^2),$$

and expanding $2y^T P_j R(y)$ and bounding each cubic monomial by its coefficient times $|y|_2^2/400$ gives $2y^T P_j R(y) \leq |y|_2^2/4$ for $\|y\|_\infty \leq 1/400$. Hence

$$2y^T P_j f(a_j + y) \leq -\frac{59}{100}|y|_2^2 \quad (\|y\|_\infty \leq 1/400),$$

which is (H2) of Assumption 4.1 with λ determined by this coefficient and the coercivity constants; (H1) and (H3) follow from the displayed matrices, the coefficient table and the coordinate bound $\|u\|_\infty \leq 35$ on the outer box. The finite-volume drift correction $v^{-1}(2eA, -eA, 4z, -2z)$ from the two falling-factorial channels is retained through D_0 .

Literal generator estimate. With $\alpha = 10^{-12}$ and $V_i = \exp(\alpha N E_i)$, substituting the fourteen jumps of one module (thirteen resident channels and the consuming membrane channel), the exchange jumps, and the shared-growth dilution jumps of the other modules into $\sum_r \lambda_r [\exp(\alpha N \Delta E_r) - 1]$, and using $|e^u - 1 - u| \leq u^2$ for $|u| \leq 1$, gives on the outer tube

$$\mathcal{L}V_i \leq \alpha V_i \left(-\frac{N}{4}|u_i - a_i|_2^2 + 2 \times 10^8 + 1.2 \times 10^9 N(\gamma + \kappa)^2 \right).$$

The resident part contributes $-N|y|^2/2 + 2 \times 10^8$ after transferring the current-volume exponent to birth scaling by $e^{kt} - 1 \leq k(e^t - 1)$ for $k = N/v \in [0, 1]$; the membrane jump $v = x + e_z$, with $L = -2y^T P_i v$, $Q = v^T P_i v$, $|L| \leq \min\{14400|y|, 36\}$ and $0 \leq Q \leq 211722$, contributes $+N|y|^2/4$ plus the $N(\gamma + \kappa)^2$ forcing after Young's inequality absorbs the term linear in $\gamma N|y|$. On the outer tube, coercivity gives $|u_i - a_i|_2^2 \leq 200 \cdot 16A_0 = 1/160000$.

Constants. For $N \geq N_{\min} = 1.4 \times 10^{20}$ and $\gamma, \kappa \leq 10^{-11}$, the constant $2 \times 10^8 + 1.2 \times 10^9 N(\gamma + \kappa)^2$ is at most $NA_0/672$, which turns the drift bound into the affine inequality $\mathcal{L}V_i \leq -kV_i + 2ke^{\alpha NA_0/2}$ with $k = N\alpha A_0/672$; the recovery time $t_0 = 2688$ makes $kt_0 = 4\alpha NA_0$, so that the probability of being active at t_0 with energy above A_0 is at most $3e^{-cN}$ with $c = \alpha A_0/2 = 1/(1.024 \times 10^{21})$. Tube exits over a phase of length t cost at most $(1+t)e^{-cN}$ per module, the early and late division clocks cost $e^{-N/125}$ and $e^{-N/500}$, and the partition step

with δ such that $504\delta^2 < A_0$ costs $8e^{-N\delta^2/35}$; each of $\delta^2/35$, $1/500$ and $1/125$ is at least c . The first phase therefore contributes the coefficient $1 + t_0 + 3 + 1 = 2693$ and the second phase $1 + t_1 + 1 + 8 = 10 + t_1$ with $t_1 = 11/(5\gamma)$, whose sum is $C(\gamma) = 2703 + 11/(5\gamma)$, per module, giving (14). The birth count bounds follow from $\|u\|_\infty \leq 35$ at birth ($140N$ per module) and positivity with a lower reference coordinate ($N/2$). Rounding the centre coordinates down to the lattice changes the energy by at most $168/N^2$, which is below $4A_0$ at the floor, so every word has an integer newborn.

Nonexplosion weight. For one module let $\Lambda = (e + \frac{3}{2})A + (2e + 1)B + z + \frac{3}{2}H$ in concentration units. Multiplying the channel increments $(e + \frac{1}{2}, -e - \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, e + \frac{3}{2}, -e - \frac{3}{2}, 2e + 1, -2e - 1, 2, -2, -\frac{3}{2}, 0)$ by the propensities of Table 3 and dropping the negative terms leaves $(36 + 60e) + 8z \leq (36 + 60e)(\Lambda + 1)$ per unit volume, and the membrane channel changes the mass weight V of Section 4.3 by zero. This is the bound $\mathcal{L}V \leq 33V$ used there after absorbing the constant, and it holds without any tube hypothesis.

C Semenov certificate data

Return regions. The exact rational centres z_ℓ and matrices P_ℓ of (21) are distributed with the manuscript sources in `terminal_regions.json`, extracted from the terminal pieces of the checked polynomial certificates. Their approximate values in molar units are

	S	C	P	E	D	U	V	I
z_0	4.98×10^{-2}	2.31×10^{-7}	1.22×10^{-8}	1.21×10^{-6}	1.00×10^{-1}	2.99×10^{-5}	2.95×10^{-7}	2.13×10^{-3}
z_1	1.32×10^{-2}	1.47×10^{-2}	9.91×10^{-3}	9.92×10^{-3}	4.75×10^{-2}	2.63×10^{-2}	2.62×10^{-2}	8.92×10^{-7}

with thiol totals $C + P + E \approx 1.4488 \mu\text{M}$ and 34.5098 mM . The low state is thioester-rich with the inhibitor near its feed level; the high state has consumed most of the thioester and the inhibitor and holds thiol and mixed disulfides at millimolar levels. The certified terminal radii, that is the largest coordinate deviations compatible with $D \leq 1$, are below $4.703 \mu\text{M}$ (low) and $1.626 \mu\text{M}$ (high); the semi-axes of the low ellipsoid range from about 38 nM to $1.1 \mu\text{M}$. The explicit integer preparations are $\lfloor \Omega z_\ell \rfloor$ componentwise; for the low state this is (1195666503776200872, 5549468225961, 293102171777, 28928608368403, 2399274129818...) in molecule counts, and exact rational energy checks certify membership.

Decoder separation. With terminal radius r_ℓ in molar units, the exact scalar checks are $100(\text{thiol}_0 + 3r_0) < S_0 - r_0$ for the low region and $1/5 < 100(\text{thiol}_1 - 3r_1)$ for the high region, which give the ratio rule $100(n_C + n_P + n_E) > \sum_i n_i$ on the high region and its negation on the low region, in addition to the fixed threshold $n_C + n_P + n_E > 1.2 \times 10^{17}$.

Polynomial cover. The recovery interval $[0, 500]$ is covered by 16 (low) and 22 (high) pieces. On each piece the centre curve and the metric are degree-16 polynomials with dyadic rational coefficients (rounded from Chebyshev interpolants to a 2^{-70} grid with exact endpoint corrections), so that adjacent pieces share exact endpoint values; the joins are continuous but not differentiable, and the finite Markov semigroup is composed across joins rather than a global C^1 argument. On every piece the following are checked as rational coefficient inequalities: positive definiteness of $P(t)$ by congruence and Gershgorin bounds; a row-norm bound 10^{-3} on the matrix residual $\dot{P} + J^T P + P J + I$; a bound on the force residual $|P(f(z) - \dot{z})|$ (6×10^{-11} low, 5×10^{-11} high), charged through Young's inequality; a metric upper bound (4537 low, 5148 high); a channel jump-energy bound (2084 low, 1290 high); and a noise coefficient (0.61 low, 0.713 high). The curvature spends 0.48 of the unit restoring form and Young's inequality 0.01, leaving at least 0.5 for the negative drift of D . A Bernstein-coefficient certificate then

bounds the degree-12 polynomial in $\sqrt{D}$ that expresses $(\partial_t + \mathcal{L})D^6$ by 10^{-9} on $D \leq 1$, and three-piece Bernstein certificates give the derivative bounds 4 and 12 of $g(D) = \phi(D^6)$ used in the allocation step.

Inventory and allocation constants. $m_0 = \Omega \cdot 0.15231/2 = 1.82772 \times 10^{18}$, $\lambda = \Omega \cdot 0.15231 \cdot 0.002 = 7.31088 \times 10^{15} \text{ s}^{-1}$, $K = 1.2 \times 10^{19}$, so the inventory drift $4\lambda/K^2$ over 500 s and the refill variance term $64m_0/K^2$ together contribute about 10^{-18} to (22). The dominant terms are $\phi(r^6)$ and $(4 + 48r)m_1$ for the high state. Twice the exact per-daughter totals are about 2.54×10^{-5} (low) and 7.90×10^{-3} (high), both below $2/250$.

Physical scale. $V = 39.85294 \mu\text{L}$; the nominal flow at $s_v = 0.002 \text{ s}^{-1}$ is $4.78 \mu\text{L}/\text{min}$; the feed quota corresponds to $19.93 \mu\text{mol}$ of reactive feed per daughter; refilling and recovering both daughters consumes a mean liquid feed volume of $3V \approx 119.6 \mu\text{L}$ per cycle. The active molecule count of a parent is at most $0.2\Omega = 4.8 \times 10^{18}$, which is why this result and the 874-molecule constructed source are different claims.

References

- [1] Andrea Agazzi, Amir Dembo, and Jean-Pierre Eckmann. Large deviations theory for Markov jump models of chemical reaction networks. *The Annals of Applied Probability*, 28(3):1821–1855, 2018. doi: 10.1214/17-AAP1344.
- [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] 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.
- [4] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977. doi: 10.1021/j100540a008.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.

- [10] 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.
- [11] Angad Yuvraj Singh and Sanjay Jain. Multistable protocells can aid the evolution of prebiotic autocatalytic sets, 2023. arXiv:2310.19744.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] Thomas Wilhelm. The smallest chemical reaction system with bistability. *BMC Systems Biology*, 3:90, 2009. doi: 10.1186/1752-0509-3-90.