Repeated selection of inherited chemical states:
finite-population guarantees and a logarithmic horizon law

19 September 2026

Abstract

Serial transfer is the standard way to demonstrate selection in a population of chemical compartments, but each transfer discards most of the population, and a state that has become rare can be lost outright to sampling rather than to chemistry. We prove finite-population guarantees for an arbitrary prescribed number of such transfers in a fully specified four-species compartment model. The protocol grows a population on a shared precursor, samples exactly M endpoint compartments uniformly and intact, recovers those same compartments, and refills in proportion to their actual retained sizes; a finite marked probability law retains every failure branch and every returned division phase, and nothing is conditioned on success or resampled. For every horizon K and every confidence below one there are finite parameters under which both inherited states remain present at every census while the measured high-to-low compartment-count log odds exceed $Kg - \log 2$, with $g = \frac{3}{5}\log 4 - \frac{19}{500} - \log\frac{51}{49} = 0.75377\ldots$. The mission transfer error is governed by a geometric series whose base is the per-cycle loss of minority share, and the attainable horizon for this event has order $\log M$. We then close part of the gap between the sufficient and necessary constants. A machine-checked exponential-generator inequality is lifted to a batch statement showing that *both* ancestral size totals grow by a factor at least $3/2$, which improves the geometric base from $204/49 = 4.1633$ to $136/49 = 2.7755$; a multiplicative concentration bound for weighted uniform sampling without replacement replaces a Chebyshev tail. Together these move the sufficient leading coefficient from $1/\log(204/49) = 0.7011$ to $1/\log(136/49) = 0.9796$ against a necessary $1/g = 1.3267$, and reduce the population sufficient for a ten-cycle mission from $M = 10^{13}$ to $M = 4 \times 10^9$ at higher confidence. Machine-checked statements, conventional proofs and numerical diagnostics are separated throughout, and the optimal exponential rate remains open.

Contents

1 Introduction

A heritable chemical difference between compartments is selected, in the laboratory sense, when its share of a population increases across repeated rounds of growth and dilution [22, 23, 26]. For compartments whose heritable difference is a chemical composition rather than a template sequence, three separate things must hold for the protocol to work. The composition must change reproductive output while compartments compete for a shared resource [5, 12, 24, 42]; it must be copied into descendants through growth and the random partition of finitely many molecules [33, 37, 39]; and it must survive the interventions between competitions.

The third requirement is the one that finite population size makes sharp, and it is the subject of this paper. Every transfer removes most of the population. Selection makes the disfavoured state rare, and a rare state is lost to the sampling step for reasons that have nothing to do with its chemistry [40, 41]. Consequently a one-cycle or two-cycle theorem does not iterate: the cells that begin cycle $j + 1$ are the actual cells retained at the end of cycle j, with heterogeneous sizes and division phases, and the minority share available as an inductive hypothesis has decayed.

1.1 What is proved

We work with one fully specified stochastic source: a four-species count network with thirteen resident channels inside each compartment, a shared growth precursor, complementary division at doubled size, uniform intact-cell transfer, a precursor-free recovery interval, and proportional refill (Section 2). The probability space is a finite marked law in which every failure branch keeps its mass in an absorbing failure symbol (Section 2.4). The following are the principal statements.

(i) *A mission guarantee at every finite horizon* (Theorem 3.1). For every K, every even M, and admissible N, γ and finite service quotas, the composed source law assigns probability at least

$$1 - K\,E(N, M, T) - \frac{80000}{M} \sum_{j<K} \Big(\frac{204}{49}\Big)^{j} - K(s_B + s_R)$$

to an event on which all K cycles return ready populations, both measured chemical counts are positive at every census, the ancestry size shares obey $q_{b,j} \geq \frac{1}{2}\rho^{j}$ with $\rho = 49/204$, and the measured count log odds satisfy $L_j - L_0 > jg - \log 2$ for every $j \leq K$. For every prescribed K and every confidence below one, finite parameters achieving it exist.

(ii) *A logarithmic horizon law* (Corollary 5.2 and Theorem 5.5). The sufficient condition above is an explicit logarithmic upper bound on K, and positivity of two integer counts summing to M forces $Kg < \log(2(M-1))$. Hence the supremum attainable integer horizon *for this gain event*, when molecular scale and service allowances may increase with M, is $\Theta(\log M)$.

(iii) *An improved exponential base* (Theorem 4.5 and Theorem 4.7). A machine-checked scalar generator inequality is lifted to the batch: on the certified event, *each* ancestral size total is at least $\frac{3}{2}$ times its value at the start of the batch, except with probability at most $e^{-19N/500000}$, a term already present in the error. Since the batch quadruples the total size, the per-cycle share loss improves from $\frac{1}{4} \cdot \frac{1-\varepsilon}{1+\varepsilon}$ to $\frac{3}{8} \cdot \frac{1-\varepsilon}{1+\varepsilon}$, and the geometric base drops from $R = 204/49$ to $R_* = 136/49$.

(iv) *An exponential transfer tail* (Proposition 4.1). For weighted uniform sampling without replacement the second-moment tail $1/(\varepsilon^2\mu)$ is replaced by $e^{-\varepsilon^2\mu/2} + e^{-\varepsilon^2\mu/(2+\varepsilon)}$, using Hoeffding's comparison with sampling *with* replacement. A sharp finite-population variance identity (Proposition 4.3) is also recorded.

(v) *Quantitative consequences* (Section 6). The sufficient leading coefficient improves from 0.701115 to 0.979591 against the necessary 1.326662. A ten-cycle mission that required $M = 10^{13}$ compartments at confidence 0.994 is closed with $M = 4\times10^{9}$ at confidence 0.997, and 0.999 with finer service quotas; the terminal composition guarantee is unchanged, and at least 36 862 minority compartments remain.

(vi) *Limits* (Sections 7 and 9). An exact hypergeometric identity for the conditional loss of a rare ancestry, complete material and duration accounts, and an explicit statement of what the model does and does not calibrate.

1.2 Three kinds of evidence

We keep three categories apart, and label every nontrivial assertion.

Machine-checked.
The finite marked laws, the conditional composition induction, the weighted-mean and share-recurrence inequalities, the closed geometric sum, the chemical sizing envelope, the necessary count ceiling, the exact conditional loss law, the material accounts, the ten-cycle

witness of Theorem 6.1, the terminal composition and integer minority consequences, and the target gain table are verified in Lean 4 [14, 36] against a frozen Mathlib. The scalar generator inequality behind Theorem 4.5 is also verified. Appendix B maps each claim to its declaration.

Conventional.
The lifting of the scalar inequality to the batch population statement (Theorem 4.5), the exponential transfer tail (Proposition 4.1), the finite-population variance identity (Proposition 4.3), the refined mission theorem (Theorem 4.7) and its witness, the $\Theta(\log M)$ formulation, and the reading of the finite marked laws as an ongoing population process are ordinary mathematics, written out here in full.

Numerical.
Decimal expansions, the two figures, the eight-cell enumeration of Section 7.2, the reduced deterministic rate comparison of Section 9, and the parameter grids of Appendix B evaluate or probe proved formulas. They establish nothing and simulate no source trajectory. All rational comparisons printed below are re-derived by an exact replay script with 2083 assertions (Appendix B).

1.3 Relation to prior work

Compositional heredity has been analysed since Segré et al. [33] and Szathmáry and Demeter [35], with sharp doubts about its evolvability [37, 38]; Nowak and Ohtsuki [29] study selection in prelife dynamics; Matsubara et al. [25] give analytical conditions for Darwinian evolution in compartmentalized autocatalytic networks across several growth-and-division regimes, including Moran-type populations [27]. Experimental RNA networks exhibit trade-offs between reproduction, variation and compositional persistence [6], and the surrounding programme is framed by the theory of autocatalytic sets [9, 17, 20]. The dependence of experimental evolution on bottleneck size is classical in population genetics [40, 41], as is the observation that asynchronously dividing lineages can differ in material without differing in number [31], which is why our conclusion is stated in compartment counts. The present contribution is orthogonal to these: it is a quantitative finite-count theorem for one fully specified source, with all failure mass retained, an unconditional mission confidence, and an explicit separation between chemical fidelity and loss of representation at a finite transfer. We make no priority claim on chemically driven compartment competition, on complementary division, or on the serial-transfer protocol, and we do not claim a theorem for arbitrary autocatalytic networks. The single-batch and two-cycle predecessors of the present result are the companion manuscripts [1, 3]; the resident network is taken unchanged from [2] and its materially balanced completion from [4].

1.4 What is not claimed

The resident exchange channels are maintained externally; four resident species do not constitute a closed thermochemically balanced system, and finite gross exchange allowances count exchanged molecules rather than modelling autonomous depletion or free energy [30, 32]. Growth and division are effective constitutive rules. The coordinate m is an integer model size with no calibrated volume or molar interpretation (Section 8.2). The theorem concerns a controlled protocol with externally imposed transfer, recovery and refill; autonomous environmental cycling is a different statement. Nothing here establishes indefinite coexistence, fixation, innovation, open-ended evolution, laboratory feasibility, or optimality of any constant. The witness parameters are astronomically large because every certificate is conservative.

Table 1: The literal resident source. The fourth resident species is named D; the scale constant D_0 of (5) is unrelated to it. Channels 7–10 and 13 are maintained exchange channels, part of the declared source rather than autonomous closed-system chemistry.

Channel	Count increment (A, B, Z, D)	Density rate	Reading
1, 2	$(-1, 1, 1, 0)$; $(1, -1, -1, 0)$	x_A; $x_B x_Z$	$A \rightleftharpoons B + Z$
3, 4	$(0, 0, -1, 1)$; $(0, 0, 1, -1)$	$16x_Z$; x_D	$Z \rightleftharpoons D$
5, 6	$(0, 0, 2, -1)$; $(0, 0, -2, 1)$	x_D; $2x_Z(x_Z - h)$	$D \rightleftharpoons 2Z$
7, 8	$(1, 0, 0, 0)$; $(-1, 0, 0, 0)$	6; x_A	A exchange
9, 10	$(0, 1, 0, 0)$; $(0, -1, 0, 0)$	27; x_B	B exchange
11, 12	$(2, -1, 0, 0)$; $(-2, 1, 0, 0)$	ex_B; $ex_A(x_A - h)$	$B \rightleftharpoons 2A$
13	$(0, 0, 0, -1)$	$x_D/10000$	D sink

2 The chemical source and the operating protocol

2.1 Resident channels

A compartment, called a cell, carries an integer size m together with internal counts $n = (n_A, n_B, n_Z, n_D) \in \mathbb{N}^4$, and its concentration vector is $x = n/m$. Write $h = 1/m$ and $e = 10^{-5}$. The resident channels have jump vectors and density rates of Table 1; the count propensity of a channel is m times its density rate, and the falling-factorial terms ensure that propensities vanish on the boundary of the count lattice.

The density-limit drift of these channels has two stationary compositions whose Z coordinates lie in the certified rational intervals

$$z_{\mathrm{L}} \in [0.99579401232,\ 0.99579401233], \qquad z_{\mathrm{H}} \in [2.97636724376,\ 2.97636724377], \tag{1}$$

with the remaining coordinates given by the rational formulas of Appendix A. Their existence and their local stability estimates are part of the verified source. The two states differ only in their resident composition; no fitness parameter is assigned, and the only feature that influences reproduction is the resident concentration of the growth-linked species Z.

2.2 Growth, division and one complete cycle

Let Q be a shared precursor stock with batch reference Ω. Each cell grows with propensity $\gamma(Q/\Omega)n_Z$, which vanishes when $Q = 0$ or $n_Z = 0$. A growth event consumes one unit of Q and one internal Z, and increases m by one. If the resulting size is $2N$, the same event replaces the parent by two daughters of size N: independently for each resident coordinate i, a count $d_i \sim \mathrm{Binomial}(n_i, 1/2)$ goes to one daughter and $n_i - d_i$ to the other, so the partition weight is $\prod_i \binom{n_i}{d_i} 2^{-n_i}$ and the two daughters together preserve the post-growth resident counts. The daughters are complementary, not independent. The ancestry tag carried through the analysis enters no propensity, no partition draw and no operation of the protocol.

A *ready population* consists of exactly M cells with sizes in $[N, 2N)$ and compositions in the inner neighbourhoods of Appendix A. One cycle, illustrated in Figure 1, consists of:

1. *Batch.* With total size W_0, charge $Q = \Omega = 4W_0$ and run until $Q = W_0$, subject to the deadline $T = 8/\gamma$. Conservation then gives endpoint total size $W^- = 4W_0$. A division triggered by the terminal growth event is processed before transfer.

2. *Transfer.* Choose uniformly one subset of exactly M intact endpoint cells. Remove the other cells and the residual precursor medium.

3. *Recovery.* Hold the selected cells at $Q = 0$ for 5376 model-time units. Resident chemistry continues; growth and division cease, so sizes and ancestral totals are frozen.

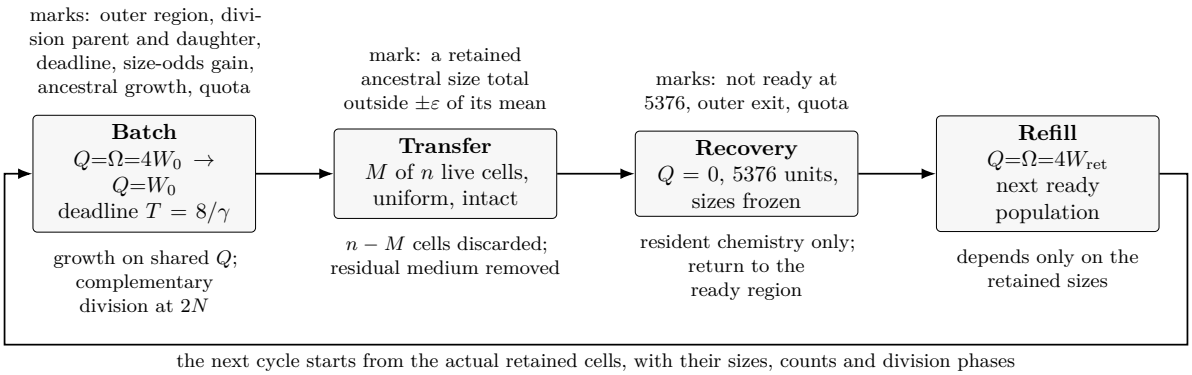

Figure 1: One literal cycle and the analytical marks that define the failure quotient. Marks enter the analysis only: uniform transfer does not read a chemical state, and recovery does not reset division phases.

4. *Refill.* Supply $Q = \Omega = 4W_{\mathrm{ret}}$ in proportion to the actual retained total size. On success this is the next ready population.

We require throughout

$$N \geq N_0 := 1.4 \times 10^{20}, \qquad 0 < \gamma \leq 10^{-11}. \tag{2}$$

Because later cycles begin from cells with arbitrary division phases, the appropriate population cap is $8M$ endpoint cells, not the newborn-only cap $4M$; every live cell has size at least N and the endpoint total is $4W_0 \leq 8NM$.

2.3 Observables

The chemical readout labels a cell high precisely when $n_Z/m > 2$. On ready states this readout agrees with the ancestry tag, with a wide margin: readiness forces $\|x - x_*\|_2 \leq \sqrt{200a} = 1/1600$ (Appendix A), so the largest ready low value of x_Z is at most 0.99641901233 and the smallest ready high value is at least 2.97574224376. Write $C_{\mathrm{H}}, C_{\mathrm{L}}$ for the measured counts, $B_{\mathrm{H}}, B_{\mathrm{L}}$ for the ancestral size totals, $W = B_{\mathrm{H}} + B_{\mathrm{L}}$, and

$$q_b = \frac{B_b}{W}, \qquad S = \log \frac{B_{\mathrm{H}}}{B_{\mathrm{L}}}, \qquad L = \log \frac{C_{\mathrm{H}}}{C_{\mathrm{L}}}. \tag{3}$$

Founders are *balanced newborn*: M is even, each state has $M/2$ cells, and every cell has size exactly N; hence $q_{\mathrm{H},0} = q_{\mathrm{L},0} = \frac{1}{2}$, $S_0 = 0$ and $L_0 = 0$.

2.4 The finite marked law

Every stage is a finite marked jump model, uniformized at a finite clock bound and Poissonized over its duration [18, 28], stopped at its first mark or at its physical endpoint. The output of a stage is either the actual physical population or a single failure symbol $\perp$, which is absorbing. Outer-neighbourhood, division, deadline, size-odds, transfer, recovery and service failures all retain their probability mass in $\perp$. Stages are composed by binding on returned states, so the law of K cycles is a probability distribution on histories of ready states or $\perp$. It is never renormalized on successful trajectories, no failed run is resampled, and no mark licenses an intervention. The conventional reading of these finite laws as an ongoing continuous-time population process is the standard one [7, 15, 21]; this paper adds no identification with an infinite trajectory-space process beyond the existing source construction.

Let J_B, J_R be positive finite batch and per-cell recovery quotas and let q_B, q_R be finite uniform clock bounds for the two source models. A counter that increments on every marked source event and saturates at J leaves the physical marginal unchanged and saturates by time t with probability at most tq/J. The per-cycle service allowances are therefore

$$s_B = \frac{Tq_B}{J_B}, \qquad s_R = \frac{M \cdot 5376\, q_R}{J_R}. \tag{4}$$

Admissible clock bounds exist because the admitted source-state families are finite; the quotas may then be enlarged to make (4) as small as prescribed. This is an existence statement, not an enumeration. Quota capacity bounds gross event allowances; it is neither actual consumption nor free energy, and the virtual holds of uniformization are not physical reaction events.

2.5 The imported chemical input

Put

$$D_0 = 5.12 \times 10^{20}, \qquad u = \frac{N}{D_0} = N\alpha a, \tag{5}$$

with $\alpha = 10^{-12}$ and $a = 1/512000000$ the readiness threshold of Appendix A. The chemical error used throughout is

$$\begin{aligned} E(N, M, T) &= A + e^{-N/2500} + e^{-19N/500000} + M\, R_0(u), \\ A &= Me^{-7u} + 14Me^{-4u} + \frac{MTu}{42}e^{-15u/2} + Me^{-u} + \frac{MTu}{42}e^{-3u/2} \\ &\quad + 56Me^{-N/(35\cdot 10^{12})}, \\ R_0(u) &= e^{-u} + 2e^{-u/2} + e^{-8u} + 16u\, e^{-31u/2}. \end{aligned} \tag{6}$$

Proposition 2.1 (Source-input proposition; machine-checked)**.** *Assume (2), the stationary roots (1), admissible finite clocks and positive quotas. Let a batch start from a ready population with both ancestries present and total size W_0. Then, except on an event of probability at most $E(N, M, T) + s_B + s_R$ under the composed batch-and-recovery law, the following hold at the batch endpoint and after recovery of the selected cells.*

(a) Endpoint. *The batch reaches $Q = W_0$ before the deadline T; the endpoint total size is $W^- = 4W_0$; every endpoint cell has size in $[N, 2N)$ and lies in the outer region.*

(b) Persistence and sandwich. *Both ancestries are present, $B_b^- \geq B_{b,0}$ for $b \in \{\mathrm{H}, \mathrm{L}\}$, and the ancestral Z totals obey*

$$2.97\, B_{\mathrm{H}} \leq Z_{\mathrm{H}} \leq 3\, B_{\mathrm{H}}, \qquad 0.99\, B_{\mathrm{L}} \leq Z_{\mathrm{L}} \leq 1.01\, B_{\mathrm{L}} \tag{7}$$

throughout the batch.

(c) Gain. *The size log odds increase by more than $b_\star = \frac{3}{5}\log 4 - \frac{19}{500} > 0.7937$, the shortfall event having probability at most $e^{-19N/500000}$.*

(d) Return. *Each selected cell returns to the ready region after 5376 units at $Q = 0$, and recovery and refill preserve its size, its tag, and hence both ancestral size totals and counts.*

These conclusions hold for the actual ready input of any cycle, not only for the initial founders, and the estimates are uniform over admitted inputs.

The local exponential chemical barriers, the division concentration bounds using Hoeffding's inequality [16], the size-odds supermartingale and the recovery estimates behind Proposition 2.1 are proved in [1, 3] and verified; Appendix B lists the declarations. Part (b) is the ingredient that Section 4 exploits, and (7) is exactly the certified ancestral rate window.

3 The mission theorem

Fix the transfer tolerance $\varepsilon = 1/50$ and set

$$\rho = \frac{1-\varepsilon}{4(1+\varepsilon)} = \frac{49}{204}, \qquad R = \rho^{-1} = \frac{204}{49}, \qquad g = b_\star - \log\frac{51}{49} = 0.7537712820\ldots > \frac{3}{4}. \tag{8}$$

Let $\mathcal{G}_K$ be the event that all K cycles return ready populations, that both measured counts are positive at every census, and that

$$L_j - L_0 > jg - \log 2, \qquad 1 \le j \le K. \tag{9}$$

Theorem 3.1 (Mission guarantee; machine-checked)**.** *Let $K \in \mathbb{N}$, let $M \ge 2$ be even, and let N, γ satisfy (2). Choose the stationary roots (1), admissible finite clocks, positive quotas, and balanced newborn founders. Under the original composed source law there is an event $\mathcal{H}_K \subseteq \mathcal{G}_K$ with*

$$\mathbb{P}(\mathcal{H}_K) \;\ge\; 1 - K\,E(N,M,T) - \frac{80000}{M}\sum_{j=0}^{K-1} R^j - K(s_B + s_R), \tag{10}$$

and on $\mathcal{H}_K$, for both states and every $0 \le j \le K$,

$$q_{b,j} \ge \tfrac{1}{2}\rho^j, \qquad \frac{C_{b,j}}{M} \ge \tfrac{1}{4}\rho^j. \tag{11}$$

For every prescribed finite K and every $0 < \delta < 1$, finite parameters and quotas exist with $\mathbb{P}(\mathcal{H}_K) \ge 1 - \delta$.

The final assertion does not keep M fixed as K grows. Neither does (9) assert monotone improvement between adjacent censuses: it is a cumulative statement measured from the founders.

3.1 Weighted shares are the right inductive quantity

Condition on a complete batch endpoint with n cells of sizes $m_i \in [N, 2N]$. For ancestry b put

$$w_{b,i} = \frac{m_i}{2N}\,\mathbf{1}_{\{b_i = b\}} \in [0,1], \qquad X_b = \sum_{i \text{ selected}} w_{b,i}. \tag{12}$$

Lemma 3.2 (Mean of a weighted retained total)**.** *Under the uniform law on M-subsets, $\mu_b = \mathbb{E}X_b = \frac{M}{n}\frac{B_b^-}{2N}$. If moreover $nN \le W^- = 4W_0$ and $B_b^- \ge \lambda B_{b,0}$ for some $\lambda \ge 1$, then*

$$\mu_b \;\ge\; \frac{\lambda}{8}\,M\,q_{b,0}. \tag{13}$$

Proof. Each cell is included with probability M/n, so $\mu_b = (M/n)\sum_i w_{b,i} = (M/n)B_b^-/(2N)$. Every live cell has size at least N, so $nN \le W^-$, that is $M/n \ge MN/(4W_0)$. Hence $\mu_b \ge \frac{MN}{4W_0}\cdot\frac{\lambda B_{b,0}}{2N} = \frac{\lambda}{8}M\frac{B_{b,0}}{W_0}$. □

With the persistence statement $B_b^- \ge B_{b,0}$ of Proposition 2.1(b), Lemma 3.2 applies with $\lambda = 1$ and gives $\mu_b \ge Mq_{b,0}/8$; Section 4 improves λ.

Lemma 3.3 (Variance at most mean)**.** *The inclusion indicators of distinct cells have nonpositive covariance, and for weights in $[0,1]$, $\mathrm{Var}(X_b) \le \mu_b$.*

Proof. Permutation symmetry gives $\mathbb{E}I_i = M/n$ and $\mathbb{E}I_iI_k = M(M-1)/(n(n-1)) \le (M/n)^2$ for $i \ne k$, the inequality being equivalent to $M \le n$. Dropping the nonpositive cross terms, $\mathrm{Var}(X_b) \le \sum_i w_{b,i}^2\,\mathrm{Var}(I_i) \le \sum_i w_{b,i}\mathbb{E}I_i = \mu_b$, using $w^2 \le w$ and $\mathrm{Var}(I) \le \mathbb{E}I$ for an indicator. □

Chebyshev's inequality then bounds the one-type relative failure by $1/(\varepsilon^2\mu_b)$, and a union over the two ancestries, with $\mu_b \geq Mq/8$ when both input shares are at least q, bounds the transfer failure by

$$\frac{16}{\varepsilon^2 Mq}. \tag{14}$$

On the complementary event both retained ancestral totals lie within relative tolerance ε of their means, which carry the common factor M/n; hence

$$q_b^+ \;\geq\; \frac{1-\varepsilon}{1+\varepsilon}\, q_b^- \;\geq\; \frac{1-\varepsilon}{4(1+\varepsilon)}\, q_b = \rho\, q_b, \tag{15}$$

using $B_b^- \geq B_{b,0}$ and $W^- = 4W_0$ for the second inequality. Recovery and refill preserve the selected sizes and ancestral totals (Proposition 2.1(d)), so (15) is the share of the next ready population.

Remark 3.4 (Why shares and not counts). At a ready census $B_b \leq 2NC_b$ and $W \geq NM$, so $C_b/M \geq q_b/2$. This factor of two is an *endpoint conversion*, applied once when a share statement is read as a count statement. The earlier count-floor argument paid it inside the recurrence at every cycle, producing the geometric base $800/49$ and the mission budget $(160000/M)\sum_{j<K}(800/49)^j$. Propagating size shares and converting once is the structural improvement behind (10).

3.2 Composition over the history

Proof of Theorem 3.1. Set $q_j = \frac{1}{2}\rho^j$. Suppose a realized prefix of j cycles is successful, so that both input shares of cycle j are at least q_j. By Proposition 2.1 and (14), the conditional probability that cycle j fails is at most

$$d_j = E(N, M, T) + \frac{16}{\varepsilon^2 M q_j} + s_B + s_R. \tag{16}$$

Write $\mathcal{H}_j$ for the event that the first j cycles all succeed with the stated conclusions. Then $\mathbb{P}(\mathcal{H}_{j+1}) \geq \mathbb{P}(\mathcal{H}_j) - d_j$, because the next conditional failure is charged only on previously successful prefixes and the mass of those prefixes is at most one. No independence between cycles is used. Summing, $\mathbb{P}(\mathcal{H}_K) \geq 1 - \sum_{j<K} d_j$; substituting q_j and $\varepsilon = 1/50$ turns $\sum_j 16/(\varepsilon^2 M q_j)$ into $(80000/M)\sum_{j<K} R^j$, giving (10), and iterating (15) from $q_{b,0} = \frac{1}{2}$ gives the share floor in (11); the count floor follows by Remark 3.4.

For the gain, a good transfer changes the size log odds by at worst $-\log\frac{1+\varepsilon}{1-\varepsilon}$, since both retained totals lie within a factor $1\pm\varepsilon$ of the same multiple of their endpoint totals. Adding the batch gain $b_\star$ of Proposition 2.1(c) and telescoping over j cycles gives $S_j - S_0 > jg$. Newborn founders have no initial mean-size bias, so $\phi_0 = 0$ for the phase $\phi = \log\big((B_{\mathrm{H}}/C_{\mathrm{H}})/(B_{\mathrm{L}}/C_{\mathrm{L}})\big)$; since all sizes lie in $[N, 2N)$ at a ready census, $|\phi_j| \leq \log 2$, and $L = S - \phi$ gives (9). The phase penalty is paid once at the endpoint, not once per cycle. Readiness equates ancestry with chemical readout (Section 2.3), so the conclusion is about measured counts.

Finally, for fixed K and δ: choose a finite even M making the finite rational transfer sum below $\delta/3$, then N making $KE \leq \delta/3$ (Proposition 5.1), then finite quotas making $K(s_B + s_R) < \delta/3$. □

4 Two refinements

Theorem 3.1 is conservative in two independent ways. Its transfer tail is a second moment, and its share recurrence uses only the nondecrease $B_b^- \geq B_{b,0}$ while the batch quadruples the total size. This section improves both. Everything in this section is a conventional proof; the scalar inequality of Lemma 4.4 is machine-checked, and Section 6 evaluates the consequences.

4.1 An exponential tail for weighted uniform sampling

Proposition 4.1 (Multiplicative concentration without replacement)**.** *Let $w_1, \dots, w_n \in [0,1]$ be deterministic, let S be a uniformly random subset of size $M \le n$, and let $X = \sum_{i \in S} w_i$ with $\mu = \mathbb{E}X$. Then for $0 < \varepsilon < 1$,*

$$\mathbb{P}\{X \le (1-\varepsilon)\mu\} \le e^{-\varepsilon^2 \mu/2}, \qquad \mathbb{P}\{X \ge (1+\varepsilon)\mu\} \le e^{-\varepsilon^2 \mu/(2+\varepsilon)}. \tag{17}$$

Proof. Let $Y = \sum_{k=1}^{M} V_k$, where $V_1, \dots, V_M$ are independent and uniform on the multiset $\{w_1, \dots, w_n\}$; this is sampling *with* replacement, and $\mathbb{E}Y = M\bar{w} = \mu$ with $\bar{w} = \frac{1}{n}\sum_i w_i$. By Hoeffding's comparison theorem for sampling without replacement [16, Thm. 4], for every continuous convex f one has $\mathbb{E}f(X) \le \mathbb{E}f(Y)$; applying it to $f(y) = e^{ty}$ for each real t gives $\mathbb{E}e^{tX} \le \mathbb{E}e^{tY}$.

It therefore suffices to bound the moment generating function of a sum of independent $[0,1]$-valued variables. For $V \in [0,1]$ and $t \in \mathbb{R}$, convexity of $v \mapsto e^{tv}$ on $[0,1]$ gives $e^{tV} \le 1 + (e^t - 1)V$, so $\mathbb{E}e^{tV_k} \le 1 + (e^t - 1)\bar{w} \le \exp\big((e^t - 1)\bar{w}\big)$ and $\mathbb{E}e^{tY} \le \exp\big((e^t - 1)\mu\big)$. The standard Chernoff optimization [10, 13] now yields, for $0 < \varepsilon < 1$,

$$\mathbb{P}\{X \ge (1+\varepsilon)\mu\} \le \Big(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\Big)^{\mu} = e^{-\mu[(1+\varepsilon)\log(1+\varepsilon) - \varepsilon]},$$
$$\mathbb{P}\{X \le (1-\varepsilon)\mu\} \le e^{-\mu[(1-\varepsilon)\log(1-\varepsilon) + \varepsilon]},$$

with the second obtained by taking $t < 0$. The elementary inequalities $(1+\varepsilon)\log(1+\varepsilon) - \varepsilon \ge \varepsilon^2/(2+\varepsilon)$ and $(1-\varepsilon)\log(1-\varepsilon) + \varepsilon \ge \varepsilon^2/2$ for $\varepsilon \in (0,1)$ give (17). □

Remark 4.2. Proposition 4.1 is a statement about an *arbitrary* deterministic weight vector, which is what the application needs: the endpoint sizes are heterogeneous and their joint law is not available. The literature on sampling without replacement supplies sharper variance-adapted bounds [8, 34], and the inclusion indicators are negatively associated [19]; (17) is the weakest form that suffices here and the easiest to bind to the existing source.

We write

$$C(\mu) = \min\Big\{\frac{1}{\varepsilon^2 \mu},\ e^{-\varepsilon^2\mu/2} + e^{-\varepsilon^2\mu/(2+\varepsilon)}\Big\} \tag{18}$$

for the resulting one-type two-sided tail at tolerance ε, keeping the Chebyshev branch because it is the better of the two at small μ and is the machine-checked one.

Proposition 4.3 (Exact finite-population variance)**.** *In the setting of Proposition 4.1, for $n > 1$,*

$$\mathrm{Var}(X) = \frac{M(n-M)}{n(n-1)}\Big[\sum_i w_i^2 - \frac{1}{n}\Big(\sum_i w_i\Big)^2\Big] \ \le\ \frac{n-M}{n-1}\,\mu. \tag{19}$$

Proof. $\mathrm{Var}(I_i) = \frac{M}{n}(1 - \frac{M}{n})$ and $\mathrm{Cov}(I_i, I_k) = -\frac{M(n-M)}{n^2(n-1)}$ for $i \ne k$. Hence $\mathrm{Var}(X) = \frac{M}{n}\frac{n-M}{n}\sum_i w_i^2 - \frac{M(n-M)}{n^2(n-1)}\sum_{i \ne k} w_i w_k$, which rearranges to the displayed identity. Dropping the subtracted square and using $w_i^2 \le w_i$ gives $\mathrm{Var}(X) \le \frac{M(n-M)}{n(n-1)}\sum_i w_i = \frac{n-M}{n-1}\mu$. □

Under the source cap $n \le 8M$ the correction factor in (19) is at worst about $7/8$ for large M, so (19) is a constant-factor improvement of Lemma 3.3 rather than a change of rate. It is recorded because it is exact, is cheap to verify, and tightens conditional calculations in which the endpoint is known.

4.2 Both ancestries grow during a batch

The share recurrence (15) loses a factor 4 because the total size quadruples while the minority is only known not to shrink. The following supplies the missing growth.

Lemma 4.4 (Scalar generator inequality; machine-checked)**.** *Let $N \geq 1000$, let $H, L \geq N$, and let $a_{\mathrm{H}} \in [2.97, 3]$, $a_{\mathrm{L}} \in [0.99, 1.01]$. Put $\eta = \frac{8}{25}$ and*

$$v_{\mathrm{H}} = \frac{N}{1000}\,\eta \log\Big(1 + \frac{1}{H+L}\Big), \qquad v_{\mathrm{L}} = \frac{N}{1000}\Big[\eta \log\Big(1 + \frac{1}{H+L}\Big) - \log\Big(1 + \frac{1}{L}\Big)\Big].$$

Then $a_{\mathrm{H}} H(e^{v_{\mathrm{H}}} - 1) + a_{\mathrm{L}} L(e^{v_{\mathrm{L}}} - 1) \leq 0$. The same inequality holds with the roles of the two coordinates exchanged, that is with the $-\log(1 + 1/H)$ term attached to the first coordinate.

The displayed form is verified in Lean (`SerialTransferSelection.minority_growth_scalar_barrier`). The exchanged form is the identical computation with a strictly larger margin, because it replaces the constraint $3\eta \leq a_{\mathrm{L}}(1 - o(1))$ by $3\eta \leq a_{\mathrm{H}}(1 - o(1))$ and $a_{\mathrm{H}} \geq 2.97 > 1.01 \geq a_{\mathrm{L}}$; the binding case is the one that is checked. The choice $\eta = 8/25$ is the largest convenient rational below the ceiling $0.99 \cdot 0.999/3 = 0.32967$ imposed by the sandwich (7), and the linear drift margin it leaves is $-2.901 \times 10^{-5} N$ against an exponential jump remainder of at most $1.536 \times 10^{-5} N$.

Theorem 4.5 (Ancestral growth during a batch)**.** *Assume the hypotheses of Proposition 2.1 and let the batch start from a ready population with both ancestries present. For each $b \in \{\mathrm{H}, \mathrm{L}\}$, except on an event of probability at most $e^{-19N/500000}$,*

$$B_b^- \;\geq\; \lambda\, B_{b,0}, \qquad \lambda = \frac{3}{2}. \tag{20}$$

On the joint event, which adds at most $e^{-19N/500000}$ to the error of Proposition 2.1, (20) holds for both ancestries simultaneously.

Proof. Fix $b = \mathrm{L}$; the high case is treated at the end. Write $H = B_{\mathrm{H}}$, $L = B_{\mathrm{L}}$, $W = H + L$, and consider the nonnegative observable

$$V = \exp\Big[\frac{N}{1000}(\eta \log W - \log L)\Big], \qquad \eta = \frac{8}{25}. \tag{21}$$

Resident channels change neither H nor L, and complementary division preserves both ancestral size totals, so V is unchanged by every transition except growth. A growth event in a cell of ancestry H increases H and W by one and multiplies V by $e^{v_{\mathrm{H}}}$; a growth event in a cell of ancestry L increases L and W by one and multiplies V by $e^{v_{\mathrm{L}}}$, with $v_{\mathrm{H}}, v_{\mathrm{L}}$ exactly as in Lemma 4.4. The total growth propensity of ancestry b is $\gamma(Q/\Omega) Z_b$, the same coefficient $\gamma(Q/\Omega) \geq 0$ multiplying both. Hence the generator of V at an unflagged state is

$$\mathcal{A}V = \gamma \frac{Q}{\Omega}\, V\Big[Z_{\mathrm{H}}(e^{v_{\mathrm{H}}} - 1) + Z_{\mathrm{L}}(e^{v_{\mathrm{L}}} - 1)\Big].$$

On the unflagged domain the sandwich (7) gives $Z_{\mathrm{H}} = a_{\mathrm{H}} H$ and $Z_{\mathrm{L}} = a_{\mathrm{L}} L$ with $a_{\mathrm{H}} \in [2.97, 3]$ and $a_{\mathrm{L}} \in [0.99, 1.01]$, and both ancestral totals are at least N because each carries at least one cell of size at least N. Lemma 4.4 and $\gamma(Q/\Omega)V \geq 0$ give $\mathcal{A}V \leq 0$ on the whole unflagged domain, and the terminal states contribute zero. Therefore V is a supermartingale under the uniformized stopped law, and $\mathbb{E}V_{\mathrm{end}} \leq V_0$.

At the endpoint $W^- = 4W_0$, so

$$\frac{V_{\mathrm{end}}}{V_0} = \exp\Big[\frac{N}{1000}\Big(\eta \log 4 - \log \frac{L^-}{L_0}\Big)\Big].$$

If (20) fails, that is $L^- < \lambda L_0$, then $V_{\mathrm{end}}/V_0 > \exp\big[\frac{N}{1000}(\eta \log 4 - \log \lambda)\big]$. Markov's inequality applied to the nonnegative supermartingale V/V_0 bounds the probability of that event by $\exp\big[-\frac{N}{1000}(\eta \log 4 - \log \lambda)\big]$. With $\eta = 8/25$ and $\lambda = 3/2$,

$$\eta \log 4 - \log \tfrac{3}{2} = \tfrac{8}{25} \log 4 - \log \tfrac{3}{2} = 0.0381588\ldots \;\geq\; \tfrac{19}{500} = 0.038, \tag{22}$$

the margin being 1.49087×10^{-4}, so the failure probability is at most $e^{-19N/500000}$, as claimed. (The rational comparison (22) is certified by enclosures of $\log 2$ and $\log \frac{3}{2}$ in the replay script.)

For $b = \mathrm{H}$ no new barrier is needed. On the event of Proposition 2.1(b)–(c), $B_{\mathrm{L}}^- \geq B_{\mathrm{L},0}$ and $\log \frac{B_{\mathrm{H}}^-}{B_{\mathrm{L}}^-} - \log \frac{B_{\mathrm{H},0}}{B_{\mathrm{L},0}} > b_\star$, so

$$\frac{B_{\mathrm{H}}^-}{B_{\mathrm{H},0}} \;>\; e^{b_\star} \frac{B_{\mathrm{L}}^-}{B_{\mathrm{L},0}} \;\geq\; e^{b_\star} \;>\; e^{0.7937} \;>\; \frac{3}{2},$$

which is (20) for the high ancestry with no additional failure term. Alternatively, the exchanged form of Lemma 4.4 gives the same conclusion with an independent $e^{-19N/500000}$ allowance. □

Remark 4.6. Theorem 4.5 is what the conservation-only countermodel rules out for free arguments: initial ancestral sizes $(4, 4)$ and endpoint sizes $(4, 28)$ are consistent with $W^- = 4W_0$ and with nondecrease of each ancestry, and attain a minority share of exactly $\frac{1}{4} \cdot \frac{1}{2}$. Conservation alone therefore permits zero minority growth, and no rearrangement of the transfer bookkeeping can improve the base; the improvement must come from the chemistry, and (7) is where it comes from. Sharpening the normalization map instead yields at most a factor 155/154: with the two-sided map $q \mapsto (1-\varepsilon)q/[4(1+\varepsilon) - 2\varepsilon q]$ the exact reciprocal from $q_0 = \frac{1}{2}$ is $q_j^{-1} = \frac{308}{155} R^j + \frac{2}{155}$, and the ratio to $\frac{1}{2}\rho^j$ never exceeds 155/154.

4.3 The refined mission theorem

Theorem 4.7 (Refined mission guarantee; conventional)**.** *Set*

$$\rho_* = \frac{\lambda(1-\varepsilon)}{4(1+\varepsilon)} = \frac{3}{2} \cdot \frac{49}{204} = \frac{49}{136}, \qquad R_* = \rho_*^{-1} = \frac{136}{49} = 2.7755102\ldots \tag{23}$$

Under the hypotheses of Theorem 3.1, with balanced newborn founders, there is an event $\mathcal{H}_K^* \subseteq \mathcal{G}_K$ *with*

$$\mathbb{P}(\mathcal{H}_K^*) \;\geq\; 1 - K\big[E(N, M, T) + e^{-19N/500000}\big] - \sum_{j=0}^{K-1} \tau_j - K(s_B + s_R), \tag{24}$$

where

$$\tau_j = C\Big(\tfrac{3}{32} M \rho_*^j\Big) + C\Big(\tfrac{3}{32} M\Big) \tag{25}$$

with C *as in* (18)*, and on* $\mathcal{H}_K^*$*, for every* $0 \leq j \leq K$*,*

$$q_{\mathrm{L},j} \geq \tfrac{1}{2}\rho_*^j, \qquad q_{\mathrm{H},j} \geq \tfrac{1}{2}, \qquad \frac{C_{\mathrm{L},j}}{M} \geq \tfrac{1}{4}\rho_*^j, \qquad \frac{C_{\mathrm{H},j}}{M} \geq \tfrac{1}{4}. \tag{26}$$

Proof. Let $\mathcal{H}_j^*$ be the event that the first j cycles succeed, that (26) holds up to index j, and that $S_i - S_0 > ig$ for $i \leq j$. Condition on a realized successful prefix. Two facts are available at the start of cycle j: the low share is at least $q_{\mathrm{L},j} = \frac{1}{2}\rho_*^j$, and the high share is at least $\frac{1}{2}$, because $S_j > S_0 = 0$ for balanced newborn founders means $B_{\mathrm{H}} > B_{\mathrm{L}}$. (This is why the theorem is stated for balanced founders; for general founders one carries $q_{\mathrm{H},j} \geq \frac{1}{2}\rho_*^j$ instead, at the cost of a second slow term in τ_j.)

Apply Proposition 2.1 and Theorem 4.5 to cycle j. On their joint event, whose complement has conditional probability at most $E(N, M, T) + e^{-19N/500000}$, the batch reaches its endpoint with $W^- = 4W_0$ and $B_b^- \geq \frac{3}{2}B_{b,0}$ for both b. Lemma 3.2 with $\lambda = \frac{3}{2}$ then gives

$$\mu_{\mathrm{L}} \;\geq\; \tfrac{3}{16}Mq_{\mathrm{L},j} \;=\; \tfrac{3}{32}M\rho_*^j, \qquad \mu_{\mathrm{H}} \;\geq\; \tfrac{3}{16}Mq_{\mathrm{H},j} \;\geq\; \tfrac{3}{32}M.$$

By Proposition 4.1 and Lemma 3.3, the probability that either retained ancestral total leaves its relative tolerance ε is at most $C(\mu_{\mathrm{L}}) + C(\mu_{\mathrm{H}}) \leq \tau_j$, since C is nonincreasing. On the complementary event both retained totals lie within a factor $1 \pm \varepsilon$ of the same multiple M/n of their endpoint totals, whence for both ancestries

$$q_b^+ \;\geq\; \frac{1-\varepsilon}{1+\varepsilon}\cdot\frac{B_b^-}{W^-} \;\geq\; \frac{1-\varepsilon}{1+\varepsilon}\cdot\frac{\frac{3}{2}B_{b,0}}{4W_0} \;=\; \rho_*\, q_{b,j},$$

and the size log odds change by at worst $-\log\frac{1+\varepsilon}{1-\varepsilon}$, so the batch gain $b_\star$ yields $S_{j+1} - S_0 > (j+1)g$. Since the high share is recovered from the gain rather than from the recurrence, this gives $q_{\mathrm{L},j+1} \geq \frac{1}{2}\rho_*^{j+1}$ and $q_{\mathrm{H},j+1} \geq \frac{1}{2}$. Recovery and refill preserve ancestral totals and counts.

Charging each conditional failure only on previously successful prefixes, as in the proof of Theorem 3.1, gives $\mathbb{P}(\mathcal{H}_{j+1}^*) \geq \mathbb{P}(\mathcal{H}_j^*) - [E + e^{-19N/500000} + \tau_j + s_B + s_R]$, and summation gives (24). The count floors follow by Remark 3.4, and $L_j - L_0 > jg - \log 2$ exactly as before. □

Remark 4.8 (What each refinement buys). The two refinements are independent and act on different quantities. Theorem 4.5 changes the geometric *base* of the transfer budget, from $R = 4.1633$ to $R_* = 2.7755$; this is the only one of the two that affects the horizon exponent. Proposition 4.1 changes the *tail*, from $1/(\varepsilon^2\mu)$ to an exponential in μ; this is worth a large constant factor but leaves the base untouched. Section 6 reports the four combinations. Neither refinement touches the third bottleneck, the molecular scale N, which is controlled by the imported chemical certificate and is discussed in Section 9.

5 The logarithmic horizon law

5.1 Sufficient conditions

The transfer budget of Theorem 3.1 has the closed form

$$D_K(M) = \frac{80000}{M}\cdot\frac{R^K - 1}{R - 1}, \tag{27}$$

and for every $\delta_t > 0$ and $M > 0$ the following are equivalent:

$$D_K(M) \leq \delta_t \iff M \geq \frac{80000}{\delta_t(R-1)}(R^K - 1) \iff K \leq \frac{\log(1 + \delta_t(R-1)M/80000)}{\log R}. \tag{28}$$

Proposition 5.1 (Molecular scale; machine-checked)**.** *For $M \geq 1$, $T \geq 0$ and $u = N/D_0$, put $C_T = 142 + 4T/21$. Then $E(N, M, T) \leq C_T M e^{-u/4}$, and for $K \geq 1$ and $\delta_c > 0$ it suffices to choose*

$$N \;\geq\; \max\{N_0,\; 4D_0 \log(C_T M K/\delta_c)\} \tag{29}$$

to ensure $KE(N, M, T) \leq \delta_c$.

Proof. Termwise comparison in (6) gives $E \leq [M(76 + 16u + Tu/21) + 2]e^{-u/2}$. Split $e^{-u/2} = e^{-u/4}e^{-u/4}$ and use $e^{-u/4} \leq 1$ together with $ue^{-u/4} \leq 4$, which follows from $1 + u/4 \leq e^{u/4}$; absorbing the polynomial into one factor gives $(78 + 64 + 4T/21)Me^{-u/4}$. Taking logarithms gives (29). □

The envelope of Proposition 5.1 is convenient for existence and asymptotics and wasteful for numerical design: at the witnesses below, termwise rational evaluation is smaller by many orders of magnitude, and (29) should not be used to size a witness. Note also that $N = O(\log(MK))$ carries the enormous constant $4D_0 = 2.048 \times 10^{21}$ and the hard floor N_0: favourable asymptotic order does not imply a small practical molecular scale.

Corollary 5.2 (Source-level logarithmic existence; machine-checked)**.** *Fix* $0 < \delta < 1$, $\gamma = 10^{-11}$ *and even* $M \geq 2$. *Every integer* $K \geq 1$ *with*

$$K \;\leq\; \frac{\log(1 + \delta(R-1)M/240000)}{\log R} \tag{30}$$

has actual source instances with confidence at least $1 - \delta$*, after choosing* N *and finite quotas sufficiently large.*

Proof. Allocate $\delta/3$ each to transfer and chemistry and each per-cycle service allowance below $\delta/[6(K+1)]$; (28) and (29) make the total failure strictly below δ. The compiled source-existence theorem binds the roots, founders, clocks and quotas, so these are nonempty source instances. □

For the refined theorem, a clean sufficient rule follows by keeping only the exponential branch of C and its slowest term.

Corollary 5.3 (Refined sufficient population; conventional)**.** *Let* $\delta_t > 0$ *and* $K \geq 1$. *If*

$$M \;\geq\; \frac{32}{3} \cdot \frac{2+\varepsilon}{\varepsilon^2} R_*^{K-1} \log \frac{4K}{\delta_t} \;=\; \frac{161600}{3} R_*^{K-1} \log \frac{4K}{\delta_t} \qquad (\varepsilon = \tfrac{1}{50}), \tag{31}$$

then $\sum_{j<K} \tau_j \leq \delta_t$. *Consequently the attainable horizon under* (24) *satisfies* $K \geq \frac{\log M}{\log R_*} - O(\log\log M)$.

Proof. Each of the $2K$ terms composing $\sum_{j<K} \tau_j$ is at most $e^{-\varepsilon^2\mu/(2+\varepsilon)}$ for some $\mu \geq \frac{3}{32} M \rho_*^{K-1}$, because $e^{-\varepsilon^2\mu/2} \leq e^{-\varepsilon^2\mu/(2+\varepsilon)}$ and C is nonincreasing. Hence $\sum_{j<K} \tau_j \leq 4K \exp\!\big(-\frac{3\varepsilon^2 M}{32(2+\varepsilon)} \rho_*^{K-1}\big)$, which is at most δ_t exactly when (31) holds. □

5.2 The necessary ceiling

Proposition 5.4 (Count-odds ceiling; machine-checked)**.** *Let* $C_{\mathrm{H}}, C_{\mathrm{L}}$ *be positive integers with* $C_{\mathrm{H}} + C_{\mathrm{L}} = M$. *Then* $L \leq \log(M-1)$. *Consequently, any balanced-founder census satisfying* (9) *at* $j = K$ *with both counts positive obeys*

$$Kg < \log(2(M-1)), \qquad \textit{equivalently} \qquad M > 1 + \tfrac{1}{2} e^{gK}. \tag{32}$$

More generally, keeping a reserve of $r \geq 1$ *minority compartments through the final census requires*

$$Kg < \log \frac{2(M-r)}{r}, \qquad r < M. \tag{33}$$

Proof. $C_{\mathrm{H}} \leq M - 1$ and $C_{\mathrm{L}} \geq 1$ give $L \leq \log(M-1)$; combining with $Kg - \log 2 < L_K$ and $L_0 = 0$ gives (32). For (33), $C_{\mathrm{L}} \geq r$ forces $C_{\mathrm{H}} \leq M - r$ and hence $L_K \leq \log \frac{M-r}{r}$. □

This is a deterministic restriction on the *gain event*. It does not prove that demographic extinction occurs when the ceiling is exceeded: a history can instead fail to achieve the required gain while both ancestries remain present. Neither does it say anything about whether a lost state could later be recreated as a chemical phenotype.

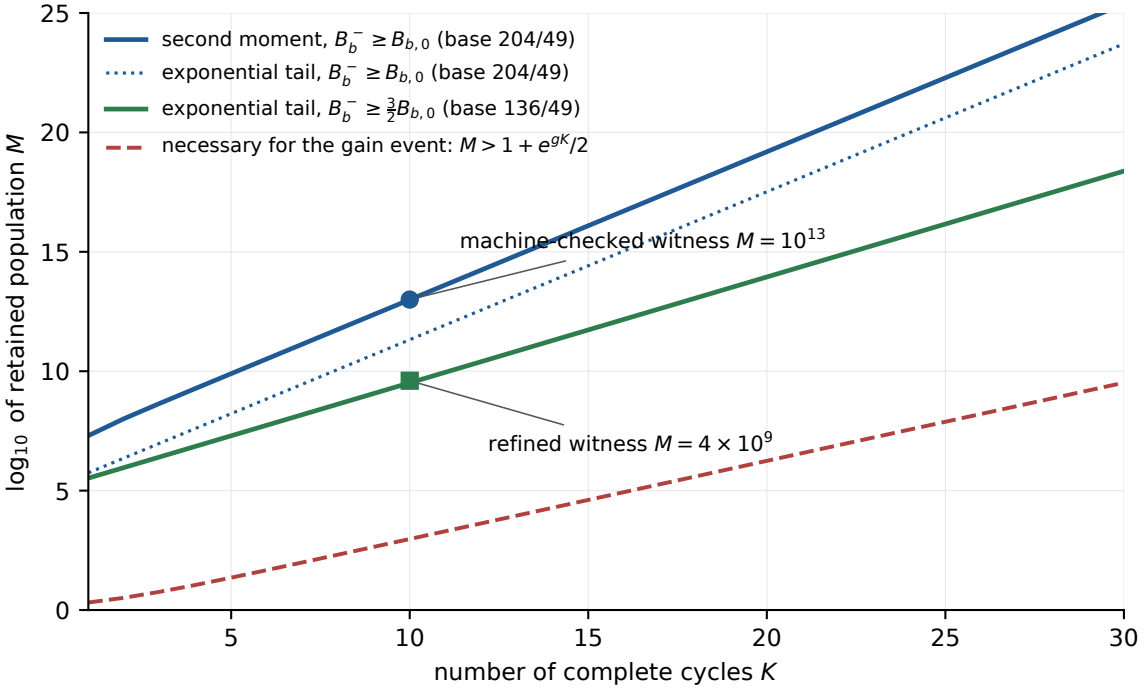

Figure 2: Retained population sufficient for a K-cycle mission at transfer allocation $\delta_t = 0.004$, and the necessary ceiling. The solid blue curve is Theorem 3.1, the green curve Theorem 4.7, and the red dashed curve Proposition 5.4. Separating the two refinements shows that the exponential tail (dotted) is a large constant factor at fixed base, while the minority-growth theorem changes the slope. The gap between the green and red lines is the open part of the rate problem; it is not a confidence interval. The curves evaluate the formulas of Sections 4 and 5 and are not simulations.

Theorem 5.5 (Logarithmic order; conventional)**.** *Fix the confidence $1 - \delta$ and allow the molecular scale and the service quotas to increase with M. Then the supremum attainable integer horizon for the event $\mathcal{G}_K$ satisfies*

$$\frac{\log M}{\log R_*} - O(\log\log M) \;\leq\; K_{\mathrm{att}}(M) \;\leq\; \frac{\log M}{g} + O(1), \qquad M \to \infty \text{ through even integers}, \tag{34}$$

so $K_{\mathrm{att}}(M) = \Theta(\log M)$. The leading coefficients are $1/\log R_* = 0.979591\ldots$ from Corollary 5.3, $1/\log R = 0.701115\ldots$ from the machine-checked Corollary 5.2, and $1/g = 1.326662\ldots$ from Proposition 5.4.

Proof. The upper bound is (32) divided by $g > 0$. The lower bound is Corollary 5.3 together with Proposition 5.1 and the service allocation of Corollary 5.2; taking logarithms of (31) gives $K \geq 1 + \frac{\log(3\delta_t M/161600) - \log\log(4K/\delta_t)}{\log R_*}$, and the log log correction is absorbed. Integer rounding changes the sufficient threshold by less than one. □

Neither coefficient in (34) is established as optimal. The ceiling is an event obstruction, not a matching stochastic converse: proving that the horizon really is $\log M/\log R_*$, or really is $\log M/g$, would require a survival converse on the same source law. Figure 2 shows the four sufficient curves and the necessary one.

Table 2: Mission failure budgets at $K = 10$. Decimals are rounded summaries of exact rational bounds. The refined design is 2500 times smaller in M and in NM, at higher confidence; its dominant term is now the service allowance, which can be made arbitrarily small by enlarging the finite quotas.

Contribution	Theorem 6.1 ($M = 10^{13}$)	Theorem 6.2 ($M = 4 \times 10^9$)
Transfer	0.003956352374	0.000967176369
Chemistry (binary exponential envelope)	0.000010842024	0.000000004337
Service allowances	0.002	0.002
Total	0.005967194397	0.002967180706
Exported confidence	≥ 0.994	≥ 0.997
Total with $s_B, s_R < 10^{-6}$	—	0.000987180706

6 Certified designs

6.1 Two ten-cycle witnesses

Theorem 6.1 (Machine-checked ten-cycle instance)**.** *Let*

$$K = 10, \quad M = 10^{13}, \quad N = 6.5536 \times 10^{22}, \quad \gamma = 10^{-11}, \quad T = 8 \times 10^{11}, \tag{35}$$

so that $u = 128$ and (2) holds, and choose finite clocks for the actual source with quotas giving $s_B, s_R < 10^{-4}$. Then Theorem 3.1 applies with $\mathbb{P}(\mathcal{H}_{10}) \geq 0.994$. On that same event, at the tenth retained census,

$$\frac{C_{\mathrm{H}}}{M} > 0.998935, \qquad C_{\mathrm{L}} \geq 1\,598\,083. \tag{36}$$

Proof. Every exponential in (6) except the recovery term $2Me^{-64}$ is bounded by 2^{-128} at $u = 128$, so $KE \leq K[M(74 + 16u + Tu/21) + 2]2^{-128} + 2KM2^{-64}$. The three contributions are tabulated in Table 2; the exact rational sum is below 0.006. For the composition, (9) at $j = 10$ gives $C_{\mathrm{H}}/C_{\mathrm{L}} > \exp(10g - \log 2) = 2^{11}e^{-19/50}(49/51)^{10} > 938.76$, proved by a finite Taylor bound with explicit remainder, and rational rearrangement gives the fraction. For the minority, (11) gives $C_{\mathrm{L}} \geq \frac{10^{13}}{4}(49/204)^{10} = 1\,598\,082.2754\ldots$, and C_{L} is an integer. Both conclusions hold on the same source event, not on two separately conditioned events. □

Theorem 6.2 (Refined ten-cycle instance; conventional)**.** *Let*

$$K = 10, \quad M = 4 \times 10^9, \quad N = 6.5536 \times 10^{22}, \quad \gamma = 10^{-11}, \quad T = 8 \times 10^{11}, \tag{37}$$

with quotas giving $s_B, s_R < 10^{-4}$. Then Theorem 4.7 applies with $\mathbb{P}(\mathcal{H}^*_{10}) \geq 0.997$, and with quotas giving $s_B, s_R < 10^{-6}$, with $\mathbb{P}(\mathcal{H}^*_{10}) \geq 0.999$. On that event $C_{\mathrm{H}}/M > 0.998935$ as before, and $C_{\mathrm{L}} \geq 36\,862$.

Proof. The exponential chemical envelope is as in Theorem 6.1 with one additional pure term $e^{-19N/500000} \leq 2^{-128}$ per cycle for Theorem 4.5, giving $KE' \leq 4.34 \times 10^{-9}$. The transfer sum $\sum_{j<10} \tau_j$ is dominated by its last term, $\tau_9 = 9.6718 \times 10^{-4}$, the previous term being 1.26×10^{-9}; the total is 9.67177×10^{-4}. The composition bound is unchanged because the gain statement is unchanged, and the minority floor is $\frac{4\times10^9}{4}(49/136)^{10} = 36\,861.39\ldots$, whose ceiling is 36 862. All entries are exact rational or enclosed-logarithm comparisons in the replay script. □

Table 3 separates the two effects at $K = 10$. The exponential tail alone is worth a factor 47; the minority-growth theorem alone a factor 97; together a factor 3038. The simple sufficient rule (31) gives $M \geq 4.85 \times 10^9$ at $K = 10$, within 50% of the exact threshold 3.26×10^9, and the round witness $M = 4 \times 10^9$ of Theorem 6.2 sits between them.

Table 3: Population sufficient for a ten-cycle mission at transfer allocation $\delta_t = 0.004$, separating the two refinements. Each entry is the smallest even M for which the corresponding transfer budget is at most δ_t.

Share recurrence	Transfer tail	Base	Sufficient M
$B_b^- \geq B_{b,0}$ (Thm. 3.1)	Chebyshev	204/49	9.891×10^{12}
$B_b^- \geq B_{b,0}$	exponential (Prop. 4.1)	204/49	2.087×10^{11}
$B_b^- \geq \frac{3}{2} B_{b,0}$ (Thm. 4.5)	Chebyshev	136/49	1.019×10^{11}
$B_b^- \geq \frac{3}{2} B_{b,0}$	exponential (Thm. 4.7)	136/49	3.256×10^{9}
Necessary for the gain event, $M > 1 + \frac{1}{2} e^{10g}$			940

6.2 Inverse target rules

Proposition 6.3 (Composition target; machine-checked)**.** *For a desired high fraction $0 < f_* < 1$, the gain requirement*

$$Kg \;\geq\; \log \frac{2f_*}{1 - f_*} \tag{38}$$

implies $C_{\mathrm{H}}/M > f_*$ on the certified event. The following conservative cycle counts have compiled gain bounds, and each is minimal for (38):

Target high fraction	90%	95%	99%	99.9%
Sufficient cycles for the gain	4	5	8	11

A composition target and a confidence target are different design inputs; each column above still needs its own population, chemistry and service budgets. A complete design sequence is therefore: choose f_* and read K from Proposition 6.3; choose δ and split it into transfer, chemical and service allocations; read M from (31) and round up to an even integer; read N from Proposition 5.1 or, better, from a termwise rational evaluation; choose quotas satisfying (4); and read the duration and material bills from Section 8.

Corollary 6.4 (Diversity reserve; conventional)**.** *Guaranteeing at least $r \geq 1$ minority compartments at the final census is ensured by the refined share certificate whenever*

$$M \;\geq\; 4r\, R_*^K \tag{39}$$

(with R in place of R_* under Theorem 3.1), and requires (33). Combine (39) with (31), the chemical sizing rule and the service allocation before rounding M to an even integer.

Proof. Immediate from $C_{\mathrm{L},K}/M \geq \frac{1}{4}\rho_*^K$ in (26) and from Proposition 5.4. □

For the worked example, request high fraction 99.89% with confidence at least 99% and a reserve of 10^4 minority compartments. Then $K = 10$ suffices for the composition by Proposition 6.3; $M = 4 \times 10^9$ satisfies (31) and exceeds $4 \cdot 10^4 \cdot R_*^{10} = 1.08 \times 10^9$ from (39); Theorem 6.2 exceeds the confidence target with room to spare; the scheduled duration is $10(T + 5376) = 8\,000\,000\,053\,760$ model-time units. This is a complete chain from target composition, confidence and diversity to K, M, N, quotas and duration, without asserting a calibrated physical timescale.

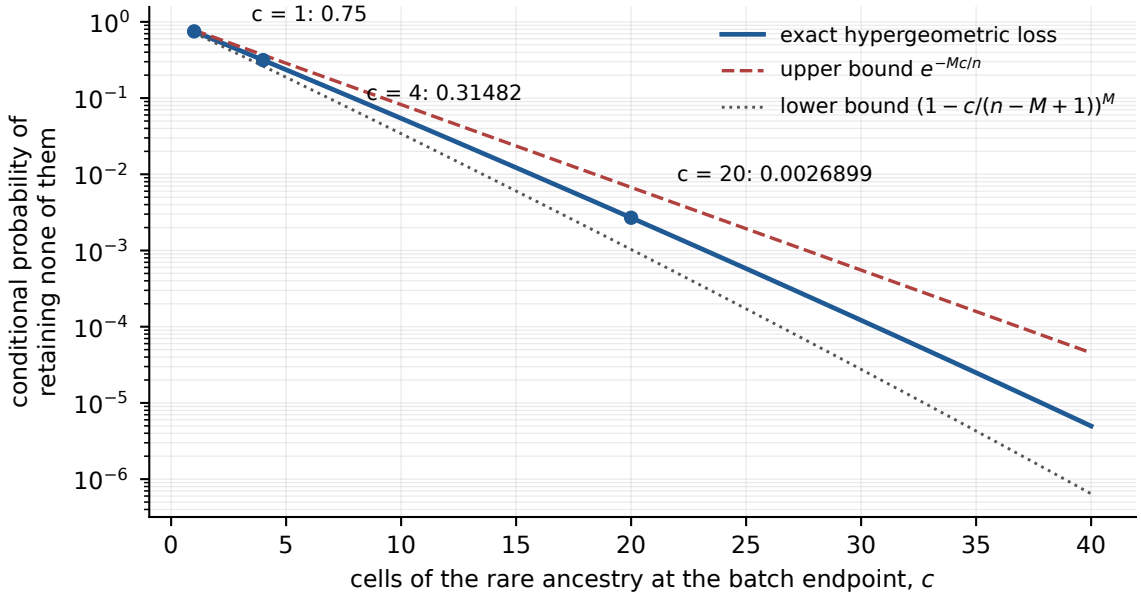

Figure 3: Exact conditional loss at $n = 400$, $M = 100$, with the bounds (41). The figure conditions on the stated endpoint; it assigns no probability to reaching that endpoint under the chemical source.

7 Rarity and the limits of the guarantee

7.1 Exact conditional loss

Proposition 7.1 (Conditional loss law; machine-checked)**.** *If the batch endpoint contains n cells of which c carry one ancestry, and $M \le n$, then uniform intact-cell transfer gives exactly*

$$\mathbb{P}(C_b^+ = 0 \mid \textit{endpoint}) = \frac{\binom{n-c}{M}}{\binom{n}{M}} = \prod_{j<M} \frac{n-c-j}{n-j}, \tag{40}$$

which is zero when $n - c < M$. For $M \le n - c$, $0 \le c \le n$ and $n > 0$,

$$\left(1 - \frac{c}{n - M + 1}\right)^M \;\le\; \mathbb{P}(C_b^+ = 0 \mid \textit{endpoint}) \;\le\; e^{-Mc/n}. \tag{41}$$

The lower bound is strictly positive in the admitted regime, and this is the point: no improvement of the chemical certificate and no increase of N can prevent uniform transfer from missing a sufficiently rare ancestry. The formula covers empty and full samples; expressions with vanishing denominators must not be extended into those cases.

For $n = 400$ and $M = 100$, minority endpoint counts $c = 1, 4, 20$ give losses $\frac{3}{4}$, $\frac{13231647}{42029596} = 0.3148174\ldots$ and $0.0026899\ldots$. Yet at $c = 20$ the expected retained minority count is only five, and the one-type 2% relative-accuracy Chebyshev bound $1/(0.02^2 \cdot 5) = 500$ is vacuous. Retaining at least one cell, preserving its weighted abundance accurately, and sustaining a declared mission gain are three distinct requirements, and an engineering discussion should report all three. A simple conditional sufficient survival rule is $Mc/n \ge \log(1/\eta)$, which makes the upper bound in (41) at most η; it is not a substitute for the mission theorem, because an unconditional guarantee also needs control of the endpoint distribution.

7.2 Counts and shares do not close the transfer law

An exact enumeration explains why the proof carries weighted size totals rather than counts. Take an eight-cell endpoint with four cells of each ancestry and $M = 4$, and enumerate all

Table 4: Mission accounting under the literal returned-state convention, for $K \geq 1$ cycles from newborn founders. Quota capacity bounds gross event allowances; it is not free energy and not actual consumption. The exact intact-cell material balance applies separately to every resident coordinate.

Quantity	Upper allowance	At (37)
Precursor charged to the K operating batches	$(8K-4)NM$	1.9923×10^{34}
Growth consumption	$(6K-3)NM$	1.4942×10^{34}
Removed residual precursor	$(2K-1)NM$	4.9807×10^{33}
Discarded cellular size	$(7K-4)NM$	1.7302×10^{34}
Terminal ready-to-run refill stock	$8NM$	2.0972×10^{33}
Precursor including terminal refill	$(8K+4)NM$	2.2020×10^{34}
Duration, excluding handling	$K(T+5376)$	8.0000×10^{12}
Each resident exchange allowance	$K(J_B + MJ_R)$	2.5038×10^{55}

$\binom{8}{4} = 70$ subsets. With all sizes equal, the two-type relative failure at tolerance $\varepsilon = 0.02$ is 17/35; with within-type sizes $\{1, 1, 1.9, 1.9\}$ in each ancestry it is 27/35, even though the type counts and the type size shares are identical. At tolerance $\varepsilon = 0.5$ the values are 1/35 and 13/35. Counts and shares alone therefore do not determine the conditional transition law, and the full weighted endpoint must be retained. This is a conditional subset computation: no probability of reaching such an endpoint under the chemical source is inferred from it.

8 Accounts, units and interpretation

8.1 Material and duration

Table 4 collects the literal balances. The terminal refill is included because the law returns a refilled ready state; it is not counted again among the K operating batches. Gross event allowances, actual consumption, stored material and free energy are four different accounts. The explicit clock bounds $\bar{q}_R = 34000N$ and $\bar{q}_B = 280000NM$ of [3], with $J_B = \lceil 10^4 T \bar{q}_B \rceil + 1$ and $J_R = \lceil 10^4 M \cdot 5376\, \bar{q}_R \rceil + 1$, give $s_B, s_R < 10^{-4}$ as required; at (37) these are $\bar{q}_R = 2.2282 \times 10^{27}$, $\bar{q}_B = 7.3400 \times 10^{37}$, $J_B = 5.8720 \times 10^{53} + 1$ and $J_R = 4.7916 \times 10^{44} + 1$.

Relative to Theorem 6.1, the refined witness reduces M and NM by a factor 2500; relative to the original count-floor ten-cycle witness ($M = 10^{20}$, $N = 2.62144 \times 10^{23}$), the factors are 2.5×10^{10} and 10^{11}. These compare sufficient certificates for a common task. They are not estimates of experimentally necessary resources.

8.2 Units

The coordinate m is an integer model size. It is not automatically a count of resident molecules, a count of solvent molecules, or a volume, and M counts cells, so it must never be converted using the Avogadro constant. Near a stationary composition the resident counts are approximately Nx_i^* and their sum approximately $N \sum_i x_i^*$.

A transparent conditional calibration introduces a reference concentration c_* with physical concentrations $c_i = c_* x_i$. If the model size coordinate is proportional to physical volume, a newborn reference volume would be $V_N = N/(N_A c_*)$ with $N_A = 6.02214076 \times 10^{23}\,\mathrm{mol}^{-1}$ exactly, by the SI definition of the mole [11]. At $N = 6.5536 \times 10^{22}$ this is $N/N_A \simeq 0.108825$ mol of reference count units, that is 0.108825 litres at $c_* = 1$ M or 108.8 litres at $c_* = 1$ mM. These are conditional dimensional conversions, not calibrated or physically endorsed implementations; actual species concentrations carry the dimensionless stationary coordinates as further factors, and a large reference concentration introduces its own inconsistencies.

The comparison runs the other way as well. Under the same illustrative convention, 1 pL at 1 mM corresponds to a reference scale near 6.0×10^8 and 1 fL at 1 mM to about 6.0×10^5, far below the source floor $N_0 = 1.4 \times 10^{20}$, which is itself only 2.3×10^{-4} mol of reference units. Renaming model-time units does not accelerate any chemistry: a physical calibration must map unimolecular, bimolecular, influx, growth and exchange rates together and preserve their ratios, and in particular the requirement that growth be slow relative to chemical relaxation survives any rescaling of time alone.

For completeness, since the point occasionally arises in discussion: the product $NM = 2.62144 \times 10^{32}$ reference size units at (37), or 6.5536×10^{35} at (35), is some ten to fourteen orders of magnitude *below* the number of water molecules in Earth's oceans, which is about 4.5×10^{46}. The real limitation is not a comparison of that kind; it is that a certificate with billions of retained compartments and an enormous per-compartment copy scale has no demonstrated laboratory implementation.

9 Scope, open problems and routes

9.1 What the rate result does and does not settle

Theorem 5.5 establishes logarithmic order, not an optimal leading constant. After the refinements of Section 4, the sufficient base is $R_* = 2.7755$ and the event ceiling requires a base at least $e^g = 2.1250$; the ratio of leading coefficients has fallen from 1.892 to 1.354. Closing the remainder cannot be done by sharpening sampling tails alone, because the sampling tail no longer appears in the base. It requires either a larger certified λ in Theorem 4.5, or a matching survival converse on the same source law.

Two diagnostics bound what to expect. First, the ceiling $\eta \leq 0.32967$ in Lemma 4.4 comes directly from the certified sandwich (7): the minority grows at relative rate a_{L} while the population grows at the weighted average, at most a_{H}, so the best λ obtainable from this barrier is $4^{a_{\mathrm{L}}/a_{\mathrm{H}}}$ up to the exponential-jump remainder, and the corresponding base is $4^{1-a_{\mathrm{L}}/a_{\mathrm{H}}} \cdot \frac{1+\varepsilon}{1-\varepsilon}$. Second, a reduced deterministic constant-rate comparison at the exact stationary roots, integrating $H = H_0 e^{a_{\mathrm{H}} s}$ and $L = L_0 e^{a_{\mathrm{L}} s}$ under the fourfold stopping rule, gives a dilution base approaching $4^{1-z_{\mathrm{L}}/z_{\mathrm{H}}} = 2.515536575\ldots$, not $e^g = 2.125$. That diagnostic concerns a reduced model and establishes nothing about the source; but it does indicate that e^g is a ceiling coming from the gain event and is unlikely to be the true sufficient base. The honest reading is that the sharp constant, if it exists, probably lies between 2.5155 and 2.7755, and that the ceiling 2.125 is not the target.

What a sharp result would additionally require is quantitative shrinking neighbourhood tracking with the finite-γ dilution correction, a uniform long-horizon comparison through the minority regime where the ancestral total is small, and a converse controlling minority survival below the corresponding population scale. A pointwise drift bound, a deterministic limiting calculation, or a scalar barrier that is not docked to the source does not supply these.

9.2 The molecular scale is the binding practical obstruction

Improving M cannot bring the molecular scale into a small-compartment regime: $u = N/D_0$ with $D_0 = 5.12 \times 10^{20}$, and even an infinitely sharp transfer estimate leaves the imported floor $N_0 = 1.4 \times 10^{20}$. A productive campaign on this axis would decompose the source drift, jump, partition, deadline and recovery estimates to identify which inequality imposes N_0 and which numerical constants dominate; compare those constants with the drift and noise evaluated near each stationary composition in dimensionally consistent coordinates; search for better quadratic forms and stopping regions and certify positive definiteness, drift and jump bounds by rational or interval arithmetic; test whether state-dependent bounds avoid paying the worst

global constant everywhere; and rebind the improved return theorem to the mission proof. Every such optimization should report the old and new effective exponent, admissible region, count floor, growth-rate restriction and probability budget. Rescaling a quadratic form while compensating its threshold is not a physical improvement if the certified region and exponent are unchanged.

9.3 Robustness before implementation

The protocol is idealized in six places, and the useful next step is a tolerance theorem rather than an assertion that a real protocol is approximately the same. A clean algebraic interface for a perturbed batch specifies a total amplification in $[A_-, A_+]$, an ancestral growth lower bound λ, and a proved batch gain b_-; the transfer geometry then gives the share factor $(1-\varepsilon)\lambda/[(1+\varepsilon)A_+]$ and the net gain $b_- - \log\frac{1+\varepsilon}{1-\varepsilon}$. Establishing those source inputs under each perturbation is the substantive work: a bounded endpoint error and terminal overshoot; a nonzero interval of recovery times; a deterministic or probabilistic refill tolerance; state- or size-dependent sampling bias; deviations in partition variance, daughter sizes or segregation bias; and a calibrated misclassification rate in the readout. A single source-bound tolerance interval would be a useful result even while a complete experimental implementation remains open.

A Ready neighbourhoods and stationary compositions

For $z \in \{z_{\mathrm{L}}, z_{\mathrm{H}}\}$ define the stationary composition $x_*(z) = (A, B, z, D)$ by

$$B = \frac{60}{z+2}, \qquad k(z) = \frac{2(1+2d)z^2 - 16(1-d)z}{2+d}, \qquad A = zB + k(z), \qquad D = \frac{16z + 2z^2}{2+d}, \tag{42}$$

where $d = 10^{-4}$. The residual equation is $27 - (1+e)B + k(z) + eA^2 = 0$ with $e = 10^{-5}$. The roots in the intervals (1) are certified by rational endpoint signs and continuity, and substitution gives stationarity of the resident drift. These formulas are also what converts a dimensionless composition into measurable concentrations once a concentration scale has been chosen.

For $y = x - x_*(z)$ write $\mathcal{E}_b(y) = 10^{-6}\sum_{i\le j} a_{ij}^{(b)} y_i y_j$ with the coefficients below. Off-diagonal entries are polynomial coefficients, not matrix entries to be doubled. Both forms satisfy $\|y\|_2^2/200 \le \mathcal{E}_b(y) \le 42\|y\|_2^2$.

(i,j)	$a_{ij}^{(\mathrm{L})}$	$a_{ij}^{(\mathrm{H})}$
(A,A)	1115801	846084
(A,B)	−1227416	−2318980
(A,Z)	4692094	4705706
(A,D)	6889770	7020116
(B,B)	1111499	4333988
(B,Z)	−4678626	−13563286
(B,D)	−6868768	−20466056
(Z,Z)	6767757	11398763
(Z,D)	20357612	34444552
(D,D)	15517433	26082111

Set $a = 1/512000000$. A cell is ready when $N \le m < 2N$ and $\mathcal{E}_b(n/m - x_*(z_b)) \le a$; readiness therefore implies $\|x - x_*\|_2 \le \sqrt{200a} = 1/1600$, which is the separation used in Section 2.3. The batch uses outer stopping threshold $8a$, predivision threshold $2a$ and daughter threshold $4a$; recovery returns to the ready threshold a. The chemical estimates use $\alpha = 10^{-12}$, so $u = N\alpha a = N/D_0$. These are definitions of the fixed imported source regions, not hypotheses

chosen after observing a history; altering any threshold is a new source-certificate task, not a numerical adjustment to an existing theorem.

B Formal evidence and reproducibility

The implementation uses a frozen Lean 4.30.0 and Mathlib environment [14, 36]; strict verification treats warnings as errors. The publication root is `proofs/SerialTransferSelection/PublicationResolution.lean`, whose receipt records 202 dependency modules, exit code 0, and no diagnostics. All proofs use only the standard axioms `Classical.choice`, `Quot.sound` and `propext`; no `sorry` is admitted.

Table 5: Claim-to-declaration map. Names after a source directory are module or declaration identifiers; the namespace for serial-transfer declarations is `SerialTransferSelection`.

Claim	Principal checked source or declaration
Literal source, stationary roots (1), ready regions	`FiniteCopy/Source.lean`, `SourceAdapter.lean`, `ReadyRegions.lean`
Proposition 2.1 (chemical return, gain, persistence, recovery)	`CycleProbability.sourceCycleLaw_probability`, `RecoveryRefill`, `BatchEndpointGain`, `BatchOddsProbability`, `BatchAncestralPersistence`
Lemmas 3.2, 3.3; transfer bound (14)	`ShareTransfer`, `TransferMoments`, `TransferVariance`, `TransferConcentration`
Share recurrence (15) and cycle law	`ShareCycleGeometry`, `ShareCycleProbability`
Theorem 3.1 (history law, all-prefix mission)	`ManyCycleLaw`, `ManyCycleShare`, `ManyCycleMeasured`, `StrongShareInstance`
Closed sum (27), inverse horizon (28), Proposition 5.1	`HorizonScaling.share_transfer_closed`, `transfer_inverse_horizon`, `chemical_sizing`
Corollary 5.2	`SourceHorizonDesign.logarithmic_source_horizon`
Proposition 5.4	`MinorityBoundary.finite_population_horizon`, `positive_count_log_odds_ceiling`
Theorem 6.1	`PublicationResolution.publication_tenCycle_instance`
Terminal composition (36)	`PublicationResolution.publication_terminal_interpretation`, `CompactWitness`
Proposition 6.3	`EnrichmentDesign.target_gain_table`, `inverse_target_fraction`
Proposition 7.1	`TransferLoss.uniformTransfer_loss`, `TransferLossBounds`
Lemma 4.4 (displayed form)	`MinorityGrowthBarrier.minority_growth_scalar_barrier`
Table 4	`ManyCycleAccounts`, `CycleMaterial`, `DiscardAccounting`
Normalization cap 155/154 (Remark after Thm. 4.5)	`NormalizationRefinement`

Status of the refinements. Theorem 4.5, Proposition 4.1, Proposition 4.3, Theorem 4.7, Theorem 6.2, Corollaries 5.3 and 6.4, and Theorem 5.5 are conventional proofs. The scalar inequality of Lemma 4.4 in the displayed orientation is machine-checked; its lifting to the batch population statement in Theorem 4.5 reuses, in written form, exactly the barrier and uniformization machinery that is machine-checked for the size-odds statement of Proposition 2.1(c), and it is not itself compiled. The exchanged orientation of Lemma 4.4 is not compiled and is not needed: the high-ancestry case of Theorem 4.5 is derived from the compiled gain and persistence statements. Readers who wish to remain entirely within the machine-checked perimeter should use Theorem 3.1, Corollary 5.2 and Theorem 6.1, whose conclusions are unaffected by anything in Section 4.

Diagnostics for the unverified steps. The replay script evaluates the scalar barrier of Lemma 4.4 at 4000 pseudorandom points spanning $N \in [10^3, 10^{23}]$, $H/N, L/N \in [1, 10^{12}]$ and

the full rate windows, and finds the expression at most $-3.048 \times 10^{-5} N$ throughout; it verifies the two Chernoff exponent inequalities of Proposition 4.1 on a grid of 999 tolerances; it checks the exact variance identity of Proposition 4.3 against the eight-cell enumeration of Section 7.2; and it checks that the exponential tail (17) dominates the exact subset law in that enumeration at three tolerances. These are diagnostics, not proofs.

Reproduction. From the repository root, the strict verification of the publication root and its five exported interfaces is

```
.\.venv\Scripts\python.exe scripts/verify_proof.py
  proofs/SerialTransferSelection/PublicationResolution.lean -timeout 900
  -declaration SerialTransferSelection.publication_tenCycle_instance
  -declaration SerialTransferSelection.publication_terminal_interpretation
  -declaration SerialTransferSelection.logarithmic_source_horizon
  -declaration SerialTransferSelection.chemical_sizing
  -declaration SerialTransferSelection.target_gain_table
  -output verification/publication.json
```

and the scalar barrier is verified by the same command with `MinorityGrowthBarrier.lean` and `-declaration SerialTransferSelection.minority_growth_scalar_barrier`. In this source package, `check_paper.py` re-derives every constant printed above (2083 assertions) and writes `check_paper_output.json`; `make_figures.py` regenerates both figures; `build.ps1` runs both and then compiles this document. The repository location will be supplied with the author information.

On automated campaign diagnostics. Freshness and route diagnostics produced by the repository's campaign tooling are administrative records. They are not kernel evidence, they do not certify or refute any statement above, and no authority closure is asserted here.

References

[1] Anonymous. Finite-batch selection between inherited chemical states under a shared limiting resource. Companion manuscript, 2026.

[2] Anonymous. Reliable copying of chemical states: finite-molecule guarantees and molecular redundancy. Companion manuscript, 2026.

[3] Anonymous. Inherited chemical-state selection across serial transfers: a finite stochastic construction with a machine-checked proof. Companion manuscript, 2026.

[4] Anonymous. Thermochemically consistent realization of a stochastic autocatalytic exporter: chemical completion, inherited operating certificates, cycle affinities and a necessary initiation scale. Companion manuscript, 2026.

[5] Katarzyna Adamala and Jack W. Szostak. Competition between model protocells driven by an encapsulated catalyst. *Nature Chemistry*, 5:495–501, 2013. doi: 10.1038/nchem.1650.

[6] Sandeep Ameta, Simon Arsène, Sandrine Foulon, Baptiste Saudemont, Bryce E. Clifton, Andrew D. Griffiths, and Philippe Nghe. Darwinian properties and their trade-offs in autocatalytic RNA reaction networks. *Nature Communications*, 12:842, 2021. doi: 10.1038/s41467-021-21000-1.

[7] 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.

[8] Rémi Bardenet and Odalric-Ambrym Maillard. Concentration inequalities for sampling without replacement. *Bernoulli*, 21(3):1361–1385, 2015. doi: 10.3150/14-BEJ605.

[9] Alex Blokhuis, David Lacoste, and Philippe Nghe. Universal motifs and the diversity of autocatalytic systems. *Proceedings of the National Academy of Sciences*, 117(41): 25230–25236, 2020. doi: 10.1073/pnas.2013527117.

[10] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford, 2013. doi: 10.1093/acprof:oso/9780199535255.001.0001.

[11] Bureau International des Poids et Mesures. The International System of Units (SI). BIPM, Sèvres, 2019. The Avogadro constant is fixed at $6.02214076 \times 10^{23}\,\mathrm{mol}^{-1}$.

[12] Irene A. Chen, Richard W. Roberts, and Jack W. Szostak. The emergence of competition between model protocells. *Science*, 305(5689):1474–1476, 2004. doi: 10.1126/science.1100757.

[13] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 1952. doi: 10.1214/aoms/1177729330.

[14] 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.

[15] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977. doi: 10.1021/j100540a008.

[16] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

[17] Wim Hordijk and Mike Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227(4):451–461, 2004. doi: 10.1016/j.jtbi.2003.11.020.

[18] Arne Jensen. Markoff chains as an aid in the study of Markoff processes. *Skandinavisk Aktuarietidskrift*, 36:87–91, 1953.

[19] Kumar Joag-Dev and Frank Proschan. Negative association of random variables, with applications. *The Annals of Statistics*, 11(1):286–295, 1983. doi: 10.1214/aos/1176346079.

[20] Stuart A. Kauffman. Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119 (1):1–24, 1986. doi: 10.1016/S0022-5193(86)80047-9.

[21] 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.

[22] Richard E. Lenski, Michael R. Rose, Suzanne C. Simpson, and Scott C. Tadler. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *The American Naturalist*, 138(6):1315–1341, 1991. doi: 10.1086/285289.

[23] Tracey A. Lincoln and Gerald F. Joyce. Self-sustained replication of an RNA enzyme. *Science*, 323(5918):1229–1232, 2009. doi: 10.1126/science.1167856.

[24] Heng Lu, Alex Blokhuis, Rebecca Turk-MacLeod, Jayaprakash Karuppusamy, Andrea Franconi, Gabrielle Woronoff, Cyrille Jeancolas, Afshin Abrishamkar, Estelle Loire, Fabien Ferrage, Philippe Pelupessy, Ludovic Jullien, Eörs Szathmáry, Philippe Nghe, and Andrew D. Griffiths. Small-molecule autocatalysis drives compartment growth, competition and reproduction. *Nature Chemistry*, 16:70–78, 2024. doi: 10.1038/s41557-023-01276-0.

[25] 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 (first version 2022).

[26] D. R. Mills, R. L. Peterson, and S. Spiegelman. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proceedings of the National Academy of Sciences*, 58(1):217–224, 1967. doi: 10.1073/pnas.58.1.217.

[27] Patrick A. P. Moran. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54(1):60–71, 1958. doi: 10.1017/S0305004100033193.

[28] James R. Norris. *Markov Chains*. Cambridge University Press, Cambridge, 1997. doi: 10.1017/CBO9780511810633.

[29] Martin A. Nowak and Hisashi Ohtsuki. Prevolutionary dynamics and the origin of evolution. *Proceedings of the National Academy of Sciences*, 105(39):14924–14927, 2008. doi: 10.1073/pnas.0806714105.

[30] Matteo Polettini and Massimiliano Esposito. Irreversible thermodynamics of open chemical networks. I. Emergent cycles and broken conservation laws. *The Journal of Chemical Physics*, 141(2):024117, 2014. doi: 10.1063/1.4886396.

[31] E. O. Powell. Growth rate and generation time of bacteria, with special reference to continuous culture. *Journal of General Microbiology*, 15:492–511, 1956. doi: 10.1099/00221287-15-3-492.

[32] Riccardo Rao and Massimiliano Esposito. Nonequilibrium thermodynamics of chemical reaction networks: wisdom from stochastic thermodynamics. *Physical Review X*, 6(4):041064, 2016. doi: 10.1103/PhysRevX.6.041064.

[33] 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.

[34] Robert J. Serfling. Probability inequalities for the sum in sampling without replacement. *The Annals of Statistics*, 2(1):39–48, 1974. doi: 10.1214/aos/1176342611.

[35] Eörs Szathmáry and László Demeter. Group selection of early replicators and the origin of life. *Journal of Theoretical Biology*, 128(4):463–486, 1987. doi: 10.1016/S0022-5193(87)80191-1.

[36] 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.

[37] Vera Vasas, Eörs Szathmáry, and Mauro Santos. Lack of evolvability in self-sustaining autocatalytic networks constraints metabolism-first scenarios for the origin of life. *Proceedings of the National Academy of Sciences*, 107(4):1470–1475, 2010. doi: 10.1073/pnas.0912628107.

[38] 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.

[39] 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.

[40] Lindi M. Wahl and Philip J. Gerrish. The probability that beneficial mutations are lost in populations with periodic bottlenecks. *Evolution*, 55(12):2606–2610, 2001. doi: 10.1111/j.0014-3820.2001.tb00772.x.

[41] Lindi M. Wahl, Philip J. Gerrish, and Ivan Saika-Voivod. Evaluating the impact of population bottlenecks in experimental evolution. *Genetics*, 162(2):961–971, 2002. doi: 10.1093/genetics/162.2.961.

[42] David Zwicker, Rabea Seyboldt, Christoph A. Weber, Anthony A. Hyman, and Frank Jülicher. Growth and division of active droplets provides a model for protocells. *Nature Physics*, 13:408–413, 2017. doi: 10.1038/nphys3984.