Diffuse RAFs and productive operation: asymptotic singleton dominance in a random polymer model

Author name to be inserted

Affiliation to be inserted

14 September 2026

Abstract

A reflexively autocatalytic and food-generated set (RAF) certifies that a reaction network can regenerate its own catalysts from food. In the capped-Zipf binary polymer model at the critical exponent sequence $a_n = 2 - 2/n$, a companion paper proved that RAFs exist with limiting probability $\theta(q_*) \in (0, 1)$, that the minimum RAF size is subexponential given existence, and that a fed, diluted stochastic mass-action reactor started from food exports a fixed amount of nonfood polymer during a fixed window with probability of order $p_n \asymp 2^{-n}$, the probability that one specified molecule catalyzes one specified reaction. It left open whether the rare productive realizations typically rely on a diffuse collective organization or contain a one-channel RAF. We prove the latter. At a sufficient quadratic volume scale, the probability of productive output without a productive one-channel incidence is $o(p_n)$; hence, conditional on productive output, the minimum RAF size converges to one, with or without conditioning on RAF existence. The proof combines a uniform kinetic exclusion, valid pointwise in every catalytic environment with at most one incidence in a fixed local rectangle, with an exact rare-multiplicity estimate for the heavy-tailed source that retains the dependence between assignments of one catalyst. The kinetic exclusion rests on two observables: a conserved product credit that cancels a mixed food–nonfood channel exactly, and a washout penalty that lets the dilution of an outsider catalyst pay for the material it could amplify. Consequences identify a unique short catalytic nucleus under successful-run conditioning and among environments with any fixed positive reliability, show that deleting that incidence makes the output probability exponentially small, and show that its signed catalytic input supplies more than three quarters of the measured export. Exactly 224 productive candidate incidences give the asymptotic upper prefactor 224 for the success probability. In contrast, the minimum RAF size diverges in probability under conditioning on RAF existence alone, and almost all RAF-containing environments have success probability $o(p_n)$. The core conditional theorem, its kinetic and source inputs, and selected consequences are verified in Lean 4; the remaining corollaries have complete conventional proofs. The theorem is a statement about statistical selection in this model, not a universal impossibility of diffuse autocatalysis.

1 Introduction

Collectively autocatalytic sets were proposed by Kauffman [14] as a way for a random chemistry to sustain itself without any single self-replicating molecule. Their combinatorial formalization is the reflexively autocatalytic and food-generated set, or RAF, of Steel and Hordijk [9, 19, 21]: a nonempty set of reactions that generates every molecule it needs from an ambient food set and whose every reaction is catalyzed by a molecule so generated. In the binary polymer model, where molecules are binary words and reactions are ligations and cleavages, RAFs appear once the mean number of reactions catalyzed per molecule grows linearly in the maximum word length n [12, 16]; heterogeneous, power-law catalysis distributions change both when RAFs appear and how large the smallest ones are [11, 13]. Which RAFs are *minimal* is a separate structural question, and a minimum RAF, an irreducible RAF, and the maximal RAF are in general different objects [20, 21].

A RAF is a statement about closure. Whether a reactor started from food actually produces material, and by what mechanism, is a kinetic question. Kinetic simulations of RAF systems go back to the first studies of the model [6], and Hordijk and Steel [10] simulated the molecular flow of RAF networks, including a constructed two-reaction mutually catalytic example under a replenished food supply. Giri and Jain [8] separated the catalytic strengths needed to maintain a large-molecule state from those needed to reach it from food. The stoichiometric theory of autocatalysis [3, 17, 23] classifies the minimal motifs that can amplify their own species and proves growth criteria in the diluted regime under sufficiently weak degradation, and seed-dependent activation [18] shows why food-only initiation is a substantive restriction. None of these frameworks decides, for a random network placed in a fed and diluted reactor with a fixed washout rate, which catalytic structures the rare successful runs actually contain.

The companion paper [1] fixed one explicit model in which that question can be posed exactly: the reversible binary polymer model with the capped-Zipf catalysis law at the critical sequence $a_n = 2 - 2/n$, a fed, diluted stochastic mass-action reactor with all spontaneous reactions active, and the productive event "total mass stays bounded and at least a fixed amount of nonfood polymer is exported during a fixed window." Three facts were proved there for the same sampled network. The probability $\mathbb{P}(G_n)$ that a RAF exists converges to $\theta(q_*) \in (0, 1)$. Conditional on existence, the minimum RAF size M_n is subexponential: $\log M_n/n \to 0$ in probability. And at a sufficient quadratic volume scale the productive event $\mathcal{F}_n$ has probability $\Theta(p_n)$, where $p_n \sim 9/(\pi^2 X_n)$ is the probability that a specified molecule catalyzes a specified channel and $X_n = 2^{n+1} - 2$ is the number of species. The companion paper noted explicitly that these facts say nothing about the RAFs contained in the rare productive environments: the structural size bound has error $o(1)$, not $o(p_n)$, and the constructive lower bound only shows that *some* productive environments of probability $\Theta(p_n)$ contain a one-channel RAF. Whether most of them do was left open.

Main result. We prove that they do, and that the one-channel RAF is the mechanism. Let $\mathcal{T}_n$ be the event that the sampled catalysis contains a *productive singleton incidence*: a channel $u + v \rightleftharpoons p$ with food inputs, nonfood product, and a catalyst in $F \cup \{p\}$, so that $\{u + v \rightleftharpoons p\}$ is by itself a RAF. Along any integer volume sequence $V(n) \geq 10^{60}(n + 1)^2$,

$$\mathbb{P}(\mathcal{F}_n \cap \mathcal{T}_n^c) = o(p_n), \qquad \text{hence} \qquad \mathbb{P}(M_n = 1 \mid \mathcal{F}_n) \to 1, \quad \mathbb{P}(M_n = 1 \mid \mathcal{F}_n \cap G_n) \to 1 \tag{1.1}$$

(Theorem 3.1). The first statement is the whole content: the probability of success without a one-channel RAF is negligible on the only scale that matters after conditioning on the rare event $\mathcal{F}_n$. Its proof has three parts, each of independent use.

(1) *Pointwise kinetic exclusion* (Proposition 4.1). Fix a huge but n-independent rectangle of short catalysts and short products. If an environment has at most one catalytic incidence in that rectangle and that incidence is not productive, then the success probability of the reactor is at most $8e^{-cV/n}$, uniformly over all kinetic marks and over the arbitrary catalytic assignments outside the rectangle. Two observables do the work: for a channel that consumes one nonfood input, the *product credit* $M - a\,x_p$ has exactly zero local increment in both directions; for a channel with food inputs but an outsider catalyst z, the *washout penalty* $M + 2000\,x_z$ lets the dilution of z absorb everything the channel could amplify.

(2) *Rare local multiplicity* (Proposition 5.1). Two incidences in any fixed rectangle have probability $o(p_n)$. A first-moment bound gives only $O(p_n)$, which is useless after division by $\mathbb{P}(\mathcal{F}_n) = \Theta(p_n)$; the gain comes from the excess-count inequality $\mathbf{1}_{\{N \geq 2\}} \leq N - \mathbf{1}_{\{N \geq 1\}}$ together with the exact row-miss probability and the capped second moment of the Zipf degree. No independence between the assignments of one catalyst is inserted, and no enumeration of the rectangle is needed.

(3) *Averaging* (Section 6). The kinetic bound is averaged over environments with at most one local incidence, the multiplicity bound covers the rest, and the imported lower bound $\mathbb{P}(\mathcal{F}_n) \geq c_0 p_n$ permits the conditioning.

Localization of the mechanism. Theorem 3.2 sharpens (1.1) in four ways. With conditional probability tending to one, the local rectangle contains *exactly one* incidence, it is productive, and it is the unique one-channel RAF; the same holds among environments whose success probability exceeds any fixed $\eta > 0$, a different conditioning that we treat separately. Deleting that single incidence, while keeping every basal reaction and every other catalytic assignment, makes the success probability at most $4e^{-cV/n}$. And along the successful trajectory itself, the signed net food-to-nonfood input through that incidence exceeds three quarters of the measured export. The nucleus is therefore more than a coexisting structural certificate: it is necessary in the counterfactual sense and dominant in the trajectory sense.

Contrast with conditioning on structure. Theorem 3.3 makes the selection effect explicit. Exactly 224 productive candidate incidences exist for every $n \geq 4$, so $\mathbb{P}(\mathcal{T}_n) = 224p_n + o(p_n)$ and $\limsup \mathbb{P}(\mathcal{F}_n)/p_n \leq 224$; all 248 singleton-RAF witnesses are similarly short, so $\mathbb{P}(M_n = 1) = 248p_n + o(p_n)$. Conditioning on RAF existence before the run changes the success probability by the factor $1/\theta(q_*)$, and $\mathbb{P}(M_n = 1 \mid G_n) \to 0$. More strongly, for every fixed m, $\mathbb{P}(M_n \leq m) = O_m(p_n)$: under conditioning on G_n alone the minimum RAF size diverges in probability, and the RAF-containing environments with $M_n > m$, which carry almost all of $\mathbb{P}(G_n)$, have success probability $o(p_n)$. The same sampler thus gives $M_n \to \infty$ given structure and $M_n \to 1$ given operation. Proposition 3.4 identifies what remains unknown about the leading constant: $\mathbb{P}(\mathcal{F}_n)/p_n$ is, up to $o(1)$, the sum of 224 environment-averaged conditional success probabilities, each subject to an exact size bias of the catalyst's degree.

What is and is not claimed. The theorem concerns typicality under the stated source, reactor, volume envelope, and food-only initial state. It does not say that every productive environment contains a singleton, that diffuse collective mechanisms are impossible, that every eligible singleton is sufficient for success, or that other reactions in a successful reactor are inactive. A food catalyst is allowed in a productive singleton; the phrase does not mean that the product must catalyze itself, and RAF catalysis must not be confused with stoichiometric autocatalysis in the sense of [3]. The constants 10^{60}, $4 \cdot 10^8$ and 10^{12} are sufficient, not necessary, and the result is not a calibrated laboratory threshold. Section 8 lists these boundaries and the open questions on the other side of them.

Formal verification. The core estimate $\mathbb{P}(\mathcal{F}_n \cap \mathcal{T}_n^c) = o(p_n)$, the conditional statements in (1.1), the pointwise kinetic exclusion, the source multiplicity estimate, the output-only posteriors, and the pointwise deletion bound are formalized in Lean 4 [5] over Mathlib [22], in a development that imports the companion paper's formal source and reactor developments; Appendix B lists the declarations and receipts. The exact candidate counts, uniqueness, reliable-environment conditioning, current attribution, the divergence of M_n under structural conditioning, the candidate decomposition, and the rate refinement of Proposition 5.3 are conventional proofs given in full; we say so where they occur.

Organization. Section 2 fixes the model and the imported inputs. Section 3 states the theorems. Sections 4–6 prove the core theorem, Section 7 the consequences, and Section 8 discusses scope. Appendix A collects the capacity and stopped-reward estimates; Appendix B is the formal concordance.

2 The model and the imported inputs

Everything below is literally the model of the companion paper [1] and of the Lean developments `PowerLawSmallRAF` (source) and `RandomViability` (reactor); the latter imports the former, so the two halves refer to one configuration weight.

2.1 Catalog and food

Let $n \geq 4$ and let $\mathcal{X}_n$ be the set of nonempty binary words of length at most n; write $\ell(x)$ for length. The food set $F = \{0, 1, 00, 01, 10, 11\}$ consists of the six words of length at most two. A *channel* is an ordered pair $r = (u, v)$ with $\ell(u) + \ell(v) \leq n$, representing both directions of the reversible reaction

$$u + v \;\rightleftharpoons\; p = uv.$$

Different ordered splits of the same product are different channels, even when the chemical stoichiometries coincide. Counting products of length j with their $j - 1$ splits,

$$X_n := |\mathcal{X}_n| = 2^{n+1} - 2, \qquad R_n := |\mathcal{R}_n| = \sum_{j=2}^{n} (j-1)2^j = (n-2)2^{n+1} + 4. \tag{2.1}$$

This convention enters the source probabilities and the constants 224 and 248 below; a model identifying ordered splits must be recounted.

2.2 The capped, shifted Zipf source

Independently for each molecule $z \in \mathcal{X}_n$ draw an integer $K_z \geq 1$ with

$$\mathbb{P}(K_z = j) = \frac{j^{-a_n}}{\zeta(a_n)}, \qquad a_n = 2 - \frac{2}{n}, \qquad D_z := \min(K_z, R_n) - 1, \tag{2.2}$$

and, conditional on $D_z = d$, let the set $C_z \subseteq \mathcal{R}_n$ of channels catalyzed by z be a uniformly random d-subset. The degree law is therefore

$$\omega_{n,d} := \mathbb{P}(D_z = d) = \begin{cases} (d+1)^{-a_n}/\zeta(a_n), & 0 \leq d \leq R_n - 2, \\ \zeta(a_n)^{-1} \sum_{j \geq R_n} j^{-a_n}, & d = R_n - 1. \end{cases} \tag{2.3}$$

Thus $D_z = 0$ has positive probability, the largest degree is $R_n - 1$, and the entire Zipf tail sits at that largest degree. This is neither a truncated-and-renormalized Zipf table nor $\min(K_z, R_n)$ itself. Catalyzing a channel catalyzes both of its directions, and food molecules may catalyze. An *incidence* is a pair (z, r) with $r \in C_z$. Its marginal probability, the same for every pair, and the empty-row probability are

$$p_n = \frac{\mathbb{E}D_z}{R_n}, \qquad q_{0,n} = \mathbb{P}(D_z = 0) = \frac{1}{\zeta(a_n)}. \tag{2.4}$$

For two distinct channels $r \neq s$ in the same row,

$$q_{2,n} := \mathbb{P}(r, s \in C_z) = \frac{\mathbb{E}[D_z(D_z - 1)]}{R_n(R_n - 1)}. \tag{2.5}$$

Rows of different molecules are independent. Within a row the assignments are sampled without replacement given the random degree, and unconditionally they are strongly dependent: nothing below models $\{r \in C_z\}_r$ as independent Bernoulli variables. Every source probability is a finite sum of the configuration weight

$$w_n(C) = \prod_{z \in \mathcal{X}_n} \frac{\omega_{n,|C_z|}}{\binom{R_n}{|C_z|}} \tag{2.6}$$

against an event indicator.

2.3 Kinetic marks and the count process

Independently of the source, draw independent marks S_r, B_r, H_{rz}, each uniform on $\{1, \frac{3}{2}, 2\}$, indexed by channels r and incidences (z, r). Both basal directions of r have coefficient $b_r = \varepsilon S_r B_r$; both catalytic directions of the incidence (z, r), that is $u + v + z \rightleftharpoons p + z$, have coefficient $\gamma_{rz} = 4 S_r H_{rz}$, where

$$\varepsilon = \frac{1}{500000000}, \qquad \varepsilon \leq b_r \leq 4\varepsilon, \qquad 4 \leq \gamma_{rz} \leq 16. \tag{2.7}$$

Basal channels are present whether or not they are catalyzed. Coefficients sharing a channel share the factor S_r; they are not independent rates. For a channel with input multiset α, coefficient k, and integer population $N = (N_x)_x$, the propensity is

$$k V^{1-|\alpha|} \prod_x (N_x)_{\alpha_x}, \qquad (j)_h = j(j-1)\cdots(j-h+1), \tag{2.8}$$

without factorial divisors. A catalyst remains in the input multiset even though its net change is zero: catalysis by the product turns the pathway into $u + v + p \rightleftharpoons 2p$, whose reverse propensity contains $(N_p)_2$; if $u = v = z$ the forward input contains three copies of u. Replacing falling factorials by powers is an upper-bound device, never an identity.

Each food species enters at rate V; every species x leaves at rate N_x. Initially $N_f(0) = V$ for each food species and every nonfood population is zero. With concentrations $x_z = N_z/V$ put

$$L = \sum_z \ell(z) x_z, \qquad M = \sum_z \mathsf{m}(z) x_z, \qquad \mathsf{m}(z) = \ell(z) \mathbf{1}_{\{\ell(z) > 2\}}. \tag{2.9}$$

Internal reactions conserve L; feed and washout give the generator identity $\mathcal{G}L = 10 - L$ with $L(0) = 10$, which does not make the trajectory $L(t)$ constant. Nonexplosion holds because total mass is bounded by its initial value plus food arrivals on finite intervals and, at fixed n and V, a mass sublevel set has finitely many states with finite rates [2]. Paired coefficients preserve internal detailed balance with $q_z = 4^{-\ell(z)}$; this is a remark about the internal pairs, not an equilibrium claim about the fed reactor.

2.4 The productive event

Let $W((s,t])$ be the total nonfood mass $\sum \mathsf{m}(x)$ counted at washout jumps during $(s,t]$. The productive event is

$$\mathcal{F}_{n,V} = \Big\{ \sup_{0\le t\le 100} L(t) \le 11, \qquad E := \frac{W((1,100])}{V} > \frac{1}{10} \Big\}, \tag{2.10}$$

and we write $\mathcal{F}_n = \mathcal{F}_{n,V(n)}$ along the volume sequence of Theorem 3.1. The event involves no RAF predicate, no reaction labels, and no selected product. The compensator of $W((1,100])/V$ is $\int_1^{100} M \,\mathrm{d}t$; the proofs control the count observable itself.

2.5 RAFs, singletons, and productive incidences

For a set S of channels, $\mathrm{cl}_S(F)$ is the least set containing F and closed under both directions of every channel in S, ignoring catalysis during closure. A *RAF* is a nonempty S whose channels all have their endpoints u, v, p in $\mathrm{cl}_S(F)$ and each of which has a catalyst in $\mathrm{cl}_S(F)$. This is the reversible convention of the formal model; accessibility through the full basal catalog does not certify a RAF. Let G_n be the event that some RAF exists and M_n the minimum cardinality of a RAF, with $M_n = \infty$ on G_n^c.

Definition 2.1 (Productive singleton incidence)**.** An incidence (z, r) with $r : u + v \rightleftharpoons p$ is *productive* if

$$u, v \in F, \qquad \ell(p) > 2, \qquad z \in F \cup \{p\}. \tag{2.11}$$

$\mathcal{T}_n$ is the event that at least one productive incidence is present.

Every productive incidence makes $\{r\}$ a RAF (Lemma 7.1), so

$$\mathcal{T}_n \subseteq \{M_n = 1\} \subseteq G_n. \tag{2.12}$$

The adjective names an eligible structural type, not a sufficiency theorem: the paper never claims that every environment containing a productive incidence succeeds, and the catalyst z may be a food molecule supplied by the feed.

2.6 Three probability laws

Let μ_n be the joint law of the source configuration and the marks, an *environment* e, and let $\pi_n(e)$ be the probability of $\mathcal{F}_n$ under the trajectory law in environment e. Then $s_n := \mathbb{P}(\mathcal{F}_n) = \int \pi_n \,\mathrm{d}\mu_n$, and three conditionings occur:

- the *successful-run posterior* $\mu_n^{\mathrm{succ}}(B) = s_n^{-1} \int_B \pi_n \,\mathrm{d}\mu_n$, the law of the environment given that one run succeeded;
- the *reliable-environment law* $\mu_n(\cdot \mid A_{n,\eta})$ with $A_{n,\eta} = \{e : \pi_n(e) \ge \eta\}$, for a fixed $\eta \in (0,1)$;
- the *structural conditioning* $\mu_n(\cdot \mid G_n)$, which depends on the source alone.

These are different distributions and are never substituted for one another. Statements "given $\mathcal{F}_n$" refer to the first, statements "among reliable environments" to the second.

2.7 Imported inputs

The following are proved in the companion paper [1] and its formal developments, for exactly the sampler (2.2) and the kernel (2.8); nothing below reproves them.

Proposition 2.2 (Imported estimates)**.**

(i) *Source moments:* $X_n p_n \to \lambda := 9/\pi^2$, $q_{0,n} \to 6/\pi^2$, *and* $X_n \, \mathbb{E}D_z^2/R_n^2 \to 0$.

(ii) *Structural limit:* $\mathbb{P}(G_n) \to \theta(q_*)$, *where* $\theta(q_*) > 0$ *is the survival probability of an infinite independent reversible closure process at openness* $q_* = 1 - e^{-\lambda}$.

(iii) *Uniform witness: let* H_n *be the source event that all six food rows are empty and* 0011 *catalyzes* $00 + 11 \rightleftharpoons 0011$*, all other rows unrestricted. Then* $\mu_n(H_n) = q_{0,n}^6 p_n$ *and, uniformly over marks and over* $V \geq 2n/d$ *with* $d = (6 \cdot 10^{20})^{-1}$,

$$\pi_{n,V}(e) \geq 1 - 24e^{-cV/n} \quad (e \in H_n), \qquad c = \frac{d^2}{4 \cdot 10^7} = \frac{1}{1.44 \cdot 10^{49}}. \tag{2.13}$$

(iv) *Upper bound without short incidences: if no molecule of length at most* $k_0 = 4 \cdot 10^8$ *catalyzes any channel with product length at most* $k_0 + 2$*, then for every mark realization and every* $V \geq 2n/d$,

$$\pi_{n,V}(e) \leq 4e^{-cV/n}. \tag{2.14}$$

(v) *Along every integer sequence* $V(n) \geq 10^{60}(n+1)^2$, $s_n = \Theta(p_n)$ *and* $\mathbb{P}(\mathcal{F}_n \cap G_n) = \Theta(p_n)$*; in fact* $s_n \geq q_{0,n}^6 p_n (1 - 24e^{-cV(n)/n})$.

Item (iv) is the companion paper's capacity theorem: without any short incidence the positive catalytic nonfood input has drift at most $85184/k_0$ on $L \leq 11$ (Appendix A), and a stopped-reward estimate turns this into (2.14). Item (v) is its consequence together with (iii). The constant $\theta(q_*)$ is the exact limit of the formal development, not a fitted number.

3 Main results

Throughout, $V(n)$ denotes an integer sequence with

$$V(n) \geq 10^{60}(n+1)^2, \tag{3.1}$$

and all probabilities refer to the joint law of source, marks, and trajectory at volume $V(n)$. Time, feed, thresholds, and coefficient bounds are fixed as n grows; only the catalog, its source law, and the volume vary. Write

$$r_n := \mathbb{P}(\mathcal{F}_n \cap \mathcal{T}_n^c), \qquad s_n := \mathbb{P}(\mathcal{F}_n).$$

For an integer $K \geq 4$ let N_K be the number of incidences (z, r) with $\ell(z) \leq K$ and $\ell(p) \leq K+2$ (the *K-rectangle*), and let $U_{n,K}$ be the event that $N_K = 1$ and this unique local incidence is productive. Two cutoffs are used:

$$k_0 = 4 \cdot 10^8, \qquad K_* = 10^{12}. \tag{3.2}$$

Both rectangles are finite and independent of n; neither is ever enumerated. "Unique local incidence" always refers to the K_*-rectangle, never to global sparsity: outside it the environment may contain arbitrarily many assignments, including heavy-tailed rows of degree up to $R_n - 1$.

Theorem 3.1 (Singleton dominance). *In the model of Section 2, along every volume sequence* (3.1),

$$r_n = \mathbb{P}(\mathcal{F}_n \cap \mathcal{T}_n^c) = o(p_n). \tag{3.3}$$

Consequently

$$\mathbb{P}(\mathcal{T}_n \mid \mathcal{F}_n) \to 1, \qquad \mathbb{P}(M_n = 1 \mid \mathcal{F}_n) \to 1, \qquad \mathbb{P}(G_n \mid \mathcal{F}_n) \to 1, \qquad \mathbb{P}(M_n = 1 \mid \mathcal{F}_n \cap G_n) \to 1. \tag{3.4}$$

The last statement is the answer to the companion paper's question in its original form: among productive environments that contain a RAF, the minimum RAF has size one with probability tending to one. Every RAF is nonempty, so the existence of a singleton RAF is exactly the event $M_n = 1$; no separate minimization is involved.

Theorem 3.2 (Localization of the mechanism). *Along every volume sequence* (3.1)*:*

- (a) (Unique nucleus.) *For every fixed $K \geq 4$, $\mathbb{P}(U_{n,K} \mid \mathcal{F}_n) \to 1$. In particular, with conditional probability tending to one there is exactly one singleton RAF and it has exactly one catalytic witness in the K-rectangle.*
- (b) (Reliable environments.) *For every fixed $\eta \in (0,1)$ and $K \geq 4$, $\mu_n(\mathcal{T}_n \mid A_{n,\eta}) \to 1$ and $\mu_n(U_{n,K} \mid A_{n,\eta}) \to 1$. Moreover $\mu_n(A_{n,\eta}) = \Theta(p_n)$ with $\liminf \mu_n(A_{n,\eta})/p_n \geq (6/\pi^2)^6$ and $\limsup \mu_n(A_{n,\eta})/p_n \leq 224$.*
- (c) (Catalytic deletion.) *On U_{n,K_*}, erase the unique local incidence, removing both of its catalytic directions and keeping its basal channel, every other assignment, all marks, the feed, and the initial state. The erased environment has success probability*

  $$\pi_n^{\mathrm{del}}(e) \leq 4e^{-cV(n)/n} \qquad (e \in U_{n,K_*}). \tag{3.5}$$

  This holds with probability tending to one under the successful-run posterior and under every reliable-environment law; under the latter the drop in success probability is at least $\eta - 4e^{-cV(n)/n}$.
- (d) (Input attribution.) *Let $J_{\mathrm{loc}}(100)$ be the signed nonfood input through the unique local incidence on $[0, 100]$, normalized by $V(n)$ and counting reverse firings negatively. Then, with probability tending to one given $\mathcal{F}_n$,*

  $$J_{\mathrm{loc}}(100) > \tfrac{3}{4}E > \tfrac{3}{40}, \qquad E = W((1,100])/V(n). \tag{3.6}$$

Part (a) is uniqueness of the one-channel RAF, not of all RAFs; larger RAFs, including larger irreducible ones, are not excluded. Part (c) compares two reactors with different sources and asserts no pathwise monotonicity under catalytic deletion; observing one successful run also does not imply that its environment had reliability near one. Part (d) is a statement about net transfer from food into nonfood on the original trajectory, not about the number of firings or the inactivity of other channels.

Theorem 3.3 (Contrast with conditioning on structure). *Along every volume sequence* (3.1)*:*

- (a) (Finite prefactors.) *For every $n \geq 4$ there are exactly 224 productive candidate incidences and exactly 248 incidences witnessing a singleton RAF, and*

  $$\begin{gathered}\mathbb{P}(\mathcal{T}_n) = 224p_n + o(p_n), \qquad \mathbb{P}(M_n = 1) = 248p_n + o(p_n), \\ \left(\frac{6}{\pi^2}\right)^6 \leq \liminf_n \frac{s_n}{p_n} \leq \limsup_n \frac{s_n}{p_n} \leq 224.\end{gathered} \tag{3.7}$$

(b) (Change of conditioning.)

$$\frac{\mathbb{P}(\mathcal{F}_n \mid G_n)}{\mathbb{P}(\mathcal{F}_n)} \longrightarrow \frac{1}{\theta(q_*)}, \qquad \mathbb{P}(M_n = 1 \mid G_n) \sim \frac{248\, p_n}{\theta(q_*)} \longrightarrow 0. \tag{3.8}$$

(c) (Divergence under structural conditioning.) *For every fixed integer $m \geq 1$, with $K_m = 2^{m+1}$,*

$$\begin{gathered}\mathbb{P}(M_n \leq m) \leq 36\, X_{K_m}\, p_n = O_m(p_n), \qquad \mathbb{P}(M_n \leq m \mid G_n) \to 0, \\ \mathbb{P}(\mathcal{F}_n \mid G_n \cap \{M_n > m\}) = o(p_n).\end{gathered} \tag{3.9}$$

So $M_n \to \infty$ in probability under $\mu_n(\cdot \mid G_n)$, while $M_n \to 1$ under the successful-run posterior; the RAF-containing environments with $M_n > m$ carry asymptotically all of $\mathbb{P}(G_n)$ and have success probability of lower order than the overall $\Theta(p_n)$. Both statements are compatible with the companion paper's $\log M_n/n \to 0$ given G_n: the ordinary minimum diverges subexponentially. The equality $s_n \sim 224 p_n$ is *not* asserted; it would require a sufficiency theorem for every candidate in its typical surroundings.

Proposition 3.4 (Candidate decomposition)**.** *Index the 224 productive candidate incidences by $i \in \mathcal{I}$, let $I_{i,n}$ be the event that candidate i is present, and let $\alpha_i(n) = \mathbb{P}(\mathcal{F}_n \mid I_{i,n}) \in [0,1]$ be its environment-averaged conditional success probability, with the full law of all other assignments and marks retained. Then*

$$\frac{s_n}{p_n} = \sum_{i \in \mathcal{I}} \alpha_i(n) + o(1), \qquad \mathbb{P}(I_{i,n} \mid \mathcal{F}_n) = \frac{p_n \alpha_i(n)}{s_n}. \tag{3.10}$$

Moreover, conditioning on the presence of one incidence biases its catalyst's degree:

$$\mathbb{P}(D_z = d \mid r \in C_z) = \frac{d\, \omega_{n,d}}{\mathbb{E} D_z}, \qquad \mathbb{E}[D_z \mid r \in C_z] = \frac{\mathbb{E} D_z^2}{\mathbb{E} D_z}. \tag{3.11}$$

The proposition locates the open problem behind an exact success constant: if the 224 numbers $\alpha_i(n)$ converge, s_n/p_n converges to their sum. The counts alone do not force $\alpha_i(n) \to 1$, nor do the 192 food-catalyzed and 32 product-catalyzed candidates imply posterior proportions 6/7 and 1/7; and by (3.11) the catalyst of the selected incidence has, on average, many more global assignments than a typical molecule, so no reduction of $\alpha_i(n)$ to an isolated motif is available without a kinetic estimate for that biased background.

4 Uniform kinetic exclusion with at most one local incidence

Proposition 4.1 (Pointwise kinetic exclusion)**.** *Let e be an environment with $N_{K_*} \leq 1$ and $e \notin \mathcal{T}_n$, with arbitrary marks and arbitrary assignments outside the K_*-rectangle. For every integer V with $V \geq 2n/d$ and $n/V \leq 1/200000$,*

$$\pi_{n,V}(e) \leq b_n(V) := 4e^{-cV/n} + 2e^{-d_m^2 V/[4n(96011\cdot 101 + d_m)]} + 2e^{-d_o^2 V/[4n(96011\cdot 101 + d_o)]} \leq 8e^{-cV/n}, \tag{4.1}$$

with $d_m = 1/200$ and $d_o = 1/200000$. Both volume conditions hold under (3.1).

This section gives the mechanism and the numerical budgets. The catalog estimates and the stochastic normalization are in Appendix A; all bounds concern the original count process (2.8) and remain valid with repeated molecules in the input multiset.

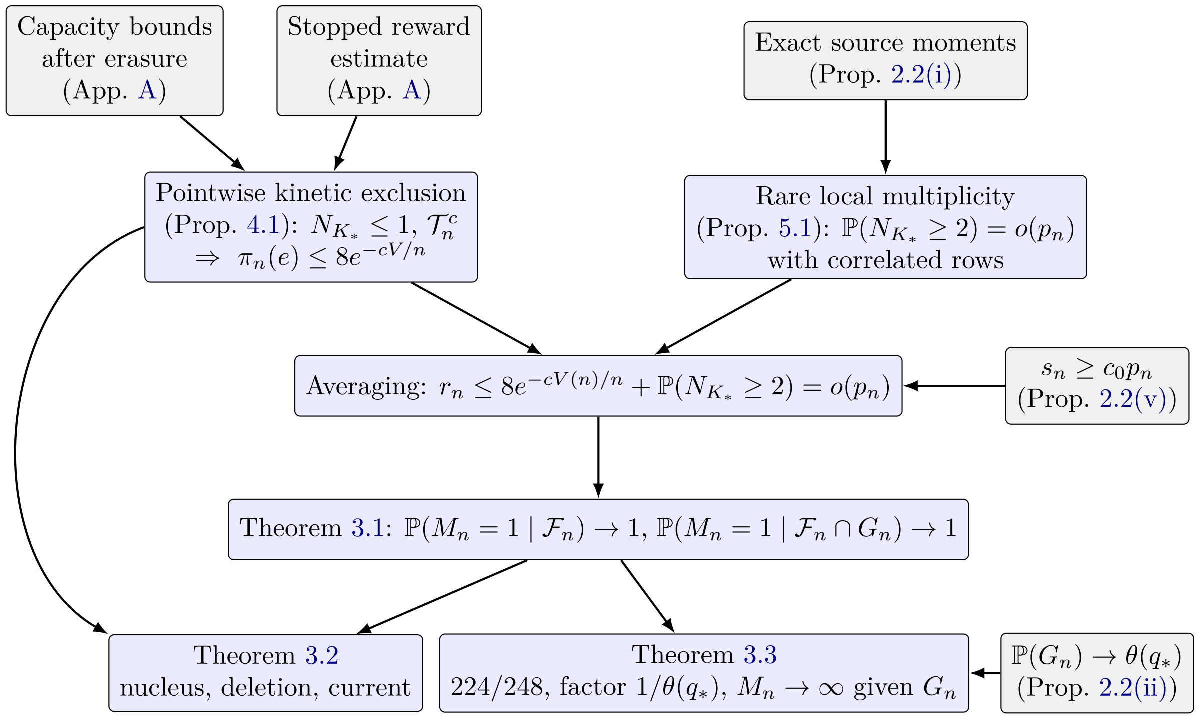

Figure 1: Proof map. Shaded boxes are imported from the companion development; blue boxes are proved here. The pointwise exclusion is also used directly for the deletion and current statements.

4.1 Capacity bounds and the case without short incidences

On $L \le 11$ every concentration monomial is bounded, and the mass constraint $\sum_{\ell(z)>K} x_z \le L/K$ converts bounded catalytic coefficients into a capacity bound. Three envelopes, valid on $L \le 11$ for a source with no incidence in the K-rectangle (Appendix A.1), are used repeatedly:

$$b = 1936\varepsilon, \qquad t_K = \frac{85184}{K}, \qquad \ell_K = 1012\varepsilon + \frac{44528}{K}, \qquad \beta_K = \frac{550}{3}\varepsilon + \frac{29040}{K} : \tag{4.2}$$

b bounds the positive basal nonfood-input drift, t_K the positive catalytic nonfood-input drift, ℓ_K the drift of loss of a fixed target coordinate of length at most $K/2$, and β_K the drift of creation of a fixed nonfood target of length at most $K+2$. All basal pathways remain present; nonlocal catalytic assignments are arbitrary.

If e has no incidence in the k_0-rectangle, (2.14) gives $\pi_n(e) \le 4e^{-cV/n}$, the first term of (4.1). Otherwise fix an incidence (z, r) in the k_0-rectangle. Since $N_{K_*} \le 1$, it is the unique incidence of the larger K_*-rectangle, and erasing it leaves a source with no incidence in the K_*-rectangle, to which the envelopes (4.2) apply with $K = K_*$. Its product has length at most $k_0 + 2 < K_*/2$ and its catalyst length at most k_0; these stronger lengths are what the coordinate estimates require.

For $r : u + v \rightleftharpoons p$ put $a = \mathsf{m}(p) - \mathsf{m}(u) - \mathsf{m}(v)$, the nonfood mass gained by one forward firing. By additivity of length there are exactly four cases (Table 1, Figure 2); the fourth is the productive case excluded by $e \notin \mathcal{T}_n$.

4.2 Neutral and mixed channels: a conserved product credit

Let r be mixed and set $P = M - a\,x_p$. A forward firing of either local direction of r changes M by a/V and x_p by $1/V$; a reverse firing changes both by the opposite amounts. Hence the local

Case	Endpoints and catalyst	a	Controlling observable
Neutral	all endpoints food, or both inputs nonfood	0	M; local increment zero
Mixed	one food input, one nonfood input	1 or 2	$P = M - a\,x_p$; local increment zero in both directions
Outsider	food inputs, nonfood product, $z \notin F \cup \{p\}$	3 or 4	$P = M + 2000\,x_z$; washout of z absorbs the local input
Productive	food inputs, nonfood product, $z \in F \cup \{p\}$	3 or 4	a one-channel RAF; excluded by $\mathcal{T}_n^c$

Table 1: The exhaustive local classification of the unique incidence (z, r), $r : u+v \rightleftharpoons p$. Coincidences of catalyst with an endpoint and equal inputs $u = v$ are included in each case.

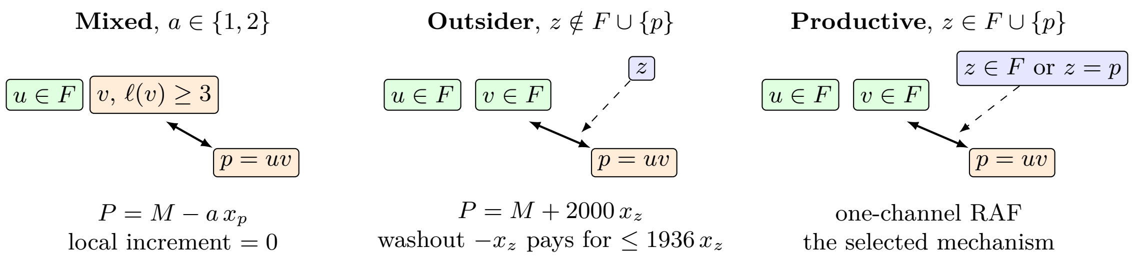

Figure 2: Three of the four local cases (the neutral case has zero nonfood gain and needs no credit). Green: food; orange: nonfood; blue: catalyst. In the mixed case the product credit cancels the channel exactly; in the outsider case the catalyst is neither fed nor produced by the channel, so its washout controls the amplification it could provide.

catalytic pathway has *exactly* zero increment of P in both directions. Only net stoichiometry enters, so the identity survives catalyst coincidences and repeated inputs. Since the nonfood input has length at least three and the food input has length $a \leq 2$, the product satisfies $\ell(p) \geq 3 + a$, whence $\mathsf{m}(p) - a \geq \frac{3}{5}\mathsf{m}(p)$ and

$$P \geq \tfrac{3}{5} M \geq 0, \tag{4.3}$$

molecule by molecule; the same comparison holds for the modified mass carried out by each washout jump. The local channel may shuttle a lot of material, but its throughput cannot raise P.

Let $I_P^+(t)$ be the cumulative positive internal P-input on $[0, t]$, normalized by V. Since $P(0) = 0$, food inflow carries zero P-weight, and $P(100) \geq 0$, the balance of P together with (4.3) forces, on $\mathcal{F}_n$,

$$\tfrac{1}{2} I_P^+(100) \;\geq\; \tfrac{1}{2} \cdot \tfrac{3}{5} E \;>\; \tfrac{3}{100}. \tag{4.4}$$

On the other side, a positive P-increment from any channel other than the two erased local directions is at most its positive ordinary nonfood input plus a times its positive loss of the coordinate x_p. Erasure removes only the local pathways, so with $K = K_*$ the envelopes (4.2) bound the drift of $\frac{1}{2} I_P^+$ by

$$\frac{b + t_{K_*} + 2\ell_{K_*}}{2} = \frac{101178}{25 \cdot 10^9} = 4.04712 \cdot 10^{-6} < \frac{1}{200000}. \tag{4.5}$$

The halving keeps every jump of the reward at most $4/V$ (an internal nonfood increment is at most $4/V$, a coordinate increment at most $2/V$). The stopped-reward estimate (A.4) with $d = d_m = 1/200$, allowance $1/100$, and horizon $T = 101$ shows that, outside an event of probability at most $2e^{-d_m^2 V/[4n(96011\cdot 101+d_m)]}$,

$$\tfrac{1}{2} I_P^+(100) \le \frac{100}{200000} + \frac{1}{100} = 0.0105 < \frac{3}{100},$$

contradicting (4.4). This is the middle term of (4.1). The neutral case is the same argument with $a = 0$ and $P = M$, where (4.3) is an equality and the export comparison is even stronger.

4.3 An outsider catalyst: its washout pays for the amplification

Now $u, v \in F$, $\ell(p) > 2$, and the catalyst z is a nonfood molecule different from p. The local pathway has zero net increment of x_z (the catalyst is returned), and its positive nonfood-input drift is at most $1936\, x_z$: the coefficient is at most 16, the gain is $a \le 4$, and $4x_u x_v \le L^2 \le 121$. Erasing the incidence does not change the drift of x_z, and the nonfood target-creation envelope gives

$$\mathcal{G} x_z \le \beta_{K_*} - x_z, \tag{4.6}$$

where $-x_z$ is washout and the feed contributes nothing to a nonfood coordinate. Take $P = M + 2000\, x_z$; since $2000 > 1936$, the washout of z dominates the local input. To keep the reward jump uniformly small, use the signed reward

$$R_o(t) = \frac{B_+(t) + C_+(t) + 2000\big(x_z(t) - x_z(0)\big)}{1001}, \tag{4.7}$$

where B_+ and C_+ are the cumulative positive internal nonfood inputs from basal and from catalytic channels, normalized by V. Its drift is at most $F_*/1001$, where

$$F_* = 1936\varepsilon + \frac{85184}{K_*} + 2000\beta_{K_*} = \frac{37282993}{46875000000}, \tag{4.8}$$

because the remaining coefficient of x_z is $1936 - 2000 \le 0$. Every jump of (4.7) has absolute value at most $(4 + 2000 \cdot 2)/(1001V) = 4/V$.

Initially $M = x_z = 0$. Pathwise nonfood mass balance gives $E \le B_+(100) + C_+(100)$ on $\mathcal{F}_n$, hence $E \le 1001 R_o(100)$ because $x_z(100) \ge 0$. Outside the reward-noise event of allowance $1/100000$, which by (A.4) with $d = d_o$ has probability at most $2e^{-d_o^2 V/[4n(96011\cdot 101+d_o)]}$,

$$E \le 100F_* + \frac{1001}{100000} = \frac{83950361}{937500000} = 0.089547051733\ldots < \frac{1}{10}, \tag{4.9}$$

a contradiction with the count event (2.10). This is the last term of (4.1), and Proposition 4.1 is proved.

Remark 4.2. Neither a deterministic fluid approximation [4, 15] nor monotonicity of success under added catalysis is used. The finite-copy correction is handled at the level of the falling-factorial propensities and the stopped trajectory law, and the stopping at the mass exit is legitimate because success itself requires a bounded-mass path. The two observables are designed for this reactor: the product credit removes a channel that merely converts existing nonfood material while incorporating a small food fragment, and the washout penalty uses the dilution of a catalyst that is neither fed nor produced by the channel. Neither is a general theorem about arbitrary reaction networks.

5 Rare local multiplicity under correlated rows

Proposition 5.1. *For every fixed integer K, $\mathbb{P}(N_K \geq 2) = o(p_n)$.*

Proof. Consider a rectangle of Z catalyst rows and J channel columns and let N be the number of its incidences. By (2.3) and uniform sampling without replacement, the probability that a given row hits at least one of the J columns is exactly

$$h_J = 1 - \sum_{d=0}^{R_n-1} \omega_{n,d} \frac{\binom{R_n-J}{d}}{\binom{R_n}{d}}, \tag{5.1}$$

with an impossible binomial coefficient read as zero. Independence of rows gives $\mathbb{P}(N=0) = (1-h_J)^Z$, and linearity of expectation gives $\mathbb{E}N = ZJp_n$. The pointwise inequality $\mathbf{1}_{\{N\geq 2\}} \leq N - \mathbf{1}_{\{N\geq 1\}}$ yields

$$\mathbb{P}(N \geq 2) \leq ZJp_n - \big[1-(1-h_J)^Z\big]. \tag{5.2}$$

No independence within a row has been inserted. Two elementary without-replacement bounds hold for $R_n \geq J$:

$$h_J \geq Jp_n - \Big(\frac{J}{R_n - J + 1}\Big)^2 \mathbb{E}D_z^2, \qquad h_J \leq \frac{J\,\mathbb{E}D_z}{R_n - J + 1}. \tag{5.3}$$

The lower bound subtracts pairwise overlaps from the sum of the J single-column probabilities; conditional on $D_z = d$, the probability that two given columns are both hit is $d(d-1)/(R_n(R_n-1)) \leq (d/(R_n-J+1))^2$, and there are $\binom{J}{2} \leq J^2$ pairs. The upper bound is the single-column union bound with an enlarged denominator. Also $Zh_J - [1-(1-h_J)^Z] \leq \binom{Z}{2}h_J^2 \leq Z^2 h_J^2$. Substituting in (5.2),

$$\mathbb{P}(N \geq 2) \leq Z\Big(\frac{J}{R_n-J+1}\Big)^2 \mathbb{E}D_z^2 + Z^2\Big(\frac{J\,\mathbb{E}D_z}{R_n-J+1}\Big)^2. \tag{5.4}$$

For fixed Z, J multiply by X_n. The first term tends to zero by Proposition 2.2(i) and $R_n/(R_n - J + 1) \to 1$; the second is $O(X_n p_n^2) = o(1)$. Dividing by $X_n p_n \to \lambda > 0$ proves the claim for that rectangle. For N_K the row and column cardinalities are at most X_K and R_{K+2} and stabilize once $n \geq K+2$, so the conclusion is uniform. □

The cancellation in (5.2) is the essential gain: a first-moment bound alone gives $\mathbb{P}(N_K \geq 2) \leq \mathbb{E}N_K = O(p_n)$, which is of the same order as the success probability. A heavy-tailed row may contain a huge number of assignments; the argument controls that possibility with the actual capped second moment rather than by replacing the row law.

Remark 5.2 (The same estimate through pair probabilities). For the reader who prefers a second-moment count: for a rectangle of Z rows and J columns, by (2.5) and row independence,

$$\mathbb{E}\binom{N}{2} = Z\binom{J}{2}q_{2,n} + \binom{Z}{2}J^2p_n^2, \tag{5.5}$$

and $\mathbf{1}_{\{N\geq 2\}} \leq \binom{N}{2}$. Since $q_{2,n} \leq \mathbb{E}D_z^2/(R_n(R_n-1))$, Proposition 2.2(i) again gives $o(p_n)$. Formula (5.5) also shows precisely why same-row pairs cannot be replaced by p_n^2: the proof succeeds because $q_{2,n}$ is negligible relative to p_n, not because assignments become independent.

Proposition 5.3 (Rate)**.** *For every fixed K, $\mathbb{P}(N_K \geq 2)/p_n = O_K(n^{-1})$. Consequently the conditional failure probabilities $\mathbb{P}(\mathcal{T}_n^c \mid \mathcal{F}_n)$ and $\mathbb{P}(U_{n,K}^c \mid \mathcal{F}_n)$ are $O_K(n^{-1})$.*

Proof. Write $a = a_n \geq 3/2$, $R = R_n$, $Y = \min(K_z, R)$. Integral comparison gives $\mathbb{P}(K_z \geq t) \leq t^{-a} + t^{1-a}/(a-1) \leq 3t^{1-a}$ for integers $t \geq 1$, using $\zeta(a) \geq 1$ and $a - 1 \geq 1/2$. The tail-sum identity for the bounded variable Y gives

$$\mathbb{E}D_z^2 \leq \mathbb{E}Y^2 = \sum_{t=1}^{R}(2t-1)\mathbb{P}(K_z \geq t) \leq 6\sum_{t=1}^{R} t^{2-a} \leq 6R^{3-a} = 6R \cdot R^{2/n}.$$

For $n \geq 4$, $R_n \leq n2^{n+1} \leq 2^{2n}$, so $R^{2/n} \leq 16$ and $\mathbb{E}D_z^2 \leq 96R_n$; hence $q_{2,n} \leq 192/R_n$. Since $R_n p_n = \mathbb{E}D_z \sim \lambda n$, dividing (5.5) by p_n gives a first term $O_{Z,J}(1/n)$ and a second term $O_{Z,J}(p_n)$. The conditional statements follow from the proofs in Sections 6 and 7, whose kinetic exceptions are $O(e^{-n})$. $\square$

The hidden constants depend on the rectangle and are astronomically large for $K = K_*$. The rate neither establishes numerical convergence at accessible n nor permits a cutoff $K(n)$ growing with n.

6 Averaging and the proof of Theorem 3.1

Split the finite environment average defining r_n according to whether $N_{K_*} \leq 1$. On $\{N_{K_*} \leq 1\} \cap \mathcal{T}_n^c$ Proposition 4.1 applies uniformly in the marks; on the complement bound $\pi_n \leq 1$. Thus

$$r_n \leq b_n(V(n)) + \mathbb{P}(N_{K_*} \geq 2). \tag{6.1}$$

The second term is $o(p_n)$ by Proposition 5.1. For the first, (3.1) gives $cV(n)/n \geq 3n$ for every $n \geq 4$, while Proposition 2.2(i) gives $p_n \geq e^{-2n}$ eventually, so

$$\frac{b_n(V(n))}{p_n} \leq 8e^{-3n}e^{2n} = 8e^{-n} \longrightarrow 0.$$

This proves (3.3). Since $s_n \geq c_0 p_n$ eventually (Proposition 2.2(v)), $\mathbb{P}(\mathcal{T}_n^c \mid \mathcal{F}_n) = r_n/s_n \to 0$, and the inclusions (2.12) give the first three limits in (3.4). Finally $\mathbb{P}(\mathcal{F}_n \cap G_n) = s_n - \mathbb{P}(\mathcal{F}_n \cap G_n^c) \geq s_n - r_n$ is eventually positive and of order p_n, and $\mathcal{F}_n \cap G_n \cap \{M_n \neq 1\} \subseteq \mathcal{F}_n \cap \mathcal{T}_n^c$; dividing proves the last limit. $\square$

7 Proofs of the consequences

7.1 Singleton RAFs and their exact count

Lemma 7.1. *A single channel $r : u + v \rightleftharpoons p$ forms a RAF $\{r\}$ if and only if $u, v \in F$ and some catalyst of r lies in $F \cup \{p\}$. For every $n \geq 4$ there are exactly 224 productive candidate incidences (of which 192 have a food catalyst and 32 are catalyzed by their own product) and exactly 248 incidences witnessing a singleton RAF; all of them lie in the 4-rectangle.*

Proof. Some direction of r must be enabled from F: otherwise $\mathrm{cl}_{\{r\}}(F) = F$, and the RAF condition would put $u, v \in F$, which enables the forward direction. If the reverse direction is enabled, $p \in F$, and by additivity of length $u, v \in F$. Otherwise the forward direction is enabled, again $u, v \in F$. In both cases $\mathrm{cl}_{\{r\}}(F) = F \cup \{p\}$, so the RAF catalyst condition is exactly the one stated. Conversely, the stated conditions place the endpoints and a catalyst in $F \cup \{p\} = \mathrm{cl}_{\{r\}}(F)$.

There are $2 \cdot 4 + 4 \cdot 2 = 16$ ordered food–food channels with a product of length three and $4 \cdot 4 = 16$ dimer–dimer channels with a product of length four; each has 7 admissible catalysts, six food words and the product, giving $32 \cdot 7 = 224$ productive candidates. The remaining singleton channels are the four ordered monomer–monomer channels, whose product is a dimer in F; each has six food catalysts, giving 24 more, and 248 in total. $\square$

7.2 Proof of Theorem 3.3

Part (a). For either candidate family, of size $b \in \{224, 248\}$, let $Y \le b$ count the candidates present. Then $\mathbb{E}Y = bp_n$ exactly, and

$$0 \le \mathbb{E}Y - \mathbb{P}(Y \ge 1) = \mathbb{E}[(Y-1)_+] \le (b-1)\mathbb{P}(Y \ge 2) \le (b-1)\mathbb{P}(N_4 \ge 2) = o(p_n)$$

by Proposition 5.1 with $K = 4$. This proves the two asymptotics in (3.7). Next, $s_n \le \mathbb{P}(\mathcal{T}_n) + r_n$ gives $\limsup s_n/p_n \le 224$, and Proposition 2.2(v) with $q_{0,n} \to 6/\pi^2$ gives $\liminf s_n/p_n \ge (6/\pi^2)^6$.

Part (b). By Bayes' identity $\mathbb{P}(\mathcal{F}_n \mid G_n)/\mathbb{P}(\mathcal{F}_n) = \mathbb{P}(G_n \mid \mathcal{F}_n)/\mathbb{P}(G_n)$, whose numerator tends to one by Theorem 3.1 and whose denominator tends to $\theta(q_*)$ by Proposition 2.2(ii). For the second limit use $\{M_n = 1\} \subseteq G_n$, $\mathbb{P}(M_n = 1) = 248p_n + o(p_n)$, and $\mathbb{P}(G_n) \to \theta(q_*) > 0$.

Part (c). First a length bound for small closures.

Lemma 7.2. *If S contains at most m channels, every word of $\mathrm{cl}_S(F)$ has length at most $K_m = 2^{m+1}$.*

Proof. Generate $\mathrm{cl}_S(F)$ by adjoining, one at a time, a product or a pair of fragments whenever the corresponding direction of a channel of S is enabled. The initial maximum length is two. A cleavage never increases the maximum. A ligation increases it by at most a factor two, and a ligation that increases it adjoins a product not previously present; each channel has one product, so at most m such increases occur. The final maximum is at most $2 \cdot 2^m$. $\square$

Now let S be a RAF with $|S| \le m$. Some channel of S is enabled from F before any nonfood word is adjoined: if nothing is ever adjoined, the closure is F and every channel of S has all endpoints in F; otherwise the first adjoined word comes from a channel enabled at F. If that direction is cleavage, its product is in F and so are its shorter inputs. In every case S contains a channel with both inputs in F and product of length at most four; there are $4+16+16 = 36$ such channels. That channel has a catalyst in $\mathrm{cl}_S(F)$, hence of length at most K_m by Lemma 7.2. There are at most $36X_{K_m}$ incidences of this form, each of marginal probability p_n, so the union bound gives $\mathbb{P}(M_n \le m) \le 36X_{K_m}p_n$ without enumerating RAFs or assuming independence. Dividing by $\mathbb{P}(G_n) \to \theta(q_*) > 0$ gives the second statement. For the third, $\mathbb{P}(G_n \cap \{M_n > m\}) = \mathbb{P}(G_n) - \mathbb{P}(M_n \le m) \to \theta(q_*)$, while the numerator $\mathbb{P}(\mathcal{F}_n \cap G_n \cap \{M_n > m\}) \le \mathbb{P}(\mathcal{F}_n \cap \{M_n \ne 1\}) \le r_n = o(p_n)$ by (2.12). $\square$

7.3 Proof of Proposition 3.4

Let $Y = \sum_{i \in \mathcal{I}} \mathbf{1}_{I_{i,n}} \le 224$. Then $\sum_i \mathbb{P}(\mathcal{F}_n \cap I_{i,n}) = \mathbb{E}[\mathbf{1}_{\mathcal{F}_n} Y]$, whose excess over $\mathbb{P}(\mathcal{F}_n \cap \mathcal{T}_n) = \mathbb{E}[\mathbf{1}_{\mathcal{F}_n}\mathbf{1}_{\{Y \ge 1\}}]$ is $\mathbb{E}[\mathbf{1}_{\mathcal{F}_n}(Y-1)_+] \in [0, 223\,\mathbb{P}(N_4 \ge 2)] = [0, o(p_n)]$. Adding $r_n = o(p_n)$ gives $s_n = \sum_i \mathbb{P}(\mathcal{F}_n \cap I_{i,n}) + o(p_n)$, and $\mathbb{P}(\mathcal{F}_n \cap I_{i,n}) = p_n\alpha_i(n)$ by definition. The posterior identity is Bayes' rule. For (3.11), $\mathbb{P}(r \in C_z \mid D_z = d) = d/R_n$, so $\mathbb{P}(D_z = d \mid r \in C_z) = (d/R_n)\omega_{n,d}/p_n = d\,\omega_{n,d}/\mathbb{E}D_z$, and the conditional mean follows. $\square$

7.4 Proof of Theorem 3.2

Part (a). Every productive candidate lies in the 4-rectangle, hence in the K-rectangle for $K \ge 4$, so

$$U_{n,K}^c \subseteq \mathcal{T}_n^c \cup \{N_K \ge 2\}.$$

Intersect with $\mathcal{F}_n$, apply (3.3) and Proposition 5.1, and divide by $s_n \geq c_0 p_n$. By Lemma 7.1 every singleton-RAF witness also lies in the K-rectangle, so on $U_{n,K}$ the unique local incidence is the only singleton witness and its channel is the only singleton RAF.

Part (b). By (2.13), $H_n \subseteq A_{n,\eta}$ once $24e^{-cV(n)/n} \leq 1-\eta$, so $\mu_n(A_{n,\eta}) \geq q_{0,n}^6 p_n$, which gives the lower prefactor. Next, since $\pi_n \geq \eta$ on $A_{n,\eta}$,

$$\eta\, \mu_n(\mathcal{T}_n^c \cap A_{n,\eta}) \leq \mathbb{E}_{\mu_n}[\pi_n \mathbf{1}_{\mathcal{T}_n^c}] = r_n = o(p_n), \tag{7.1}$$

and likewise $\mu_n(U_{n,K}^c \cap A_{n,\eta}) \leq \mu_n(\mathcal{T}_n^c \cap A_{n,\eta}) + \mu_n(N_K \geq 2) = o(p_n)$; dividing by $\mu_n(A_{n,\eta}) \geq q_{0,n}^6 p_n$ gives both conditional limits. Finally $\mu_n(A_{n,\eta}) \leq \mu_n(\mathcal{T}_n) + \mu_n(\mathcal{T}_n^c \cap A_{n,\eta}) \leq 224 p_n + o(p_n)$. The fixed threshold matters: the bound divides by η.

Part (c). On U_{n,K_*} every incidence of the K_*-rectangle is the erased one, so the erased source has no incidence in the K_*-rectangle and a fortiori none in the k_0-rectangle. The erased reactor has the same rate caps and the same food-only initial state, so (2.14) applies to it and gives (3.5). The two typicality statements are Parts (a) and (b), and on $A_{n,\eta}$ the original success probability is at least η.

Part (d). Fix $e \in U_{n,K_*}$ with its unique local incidence (z,r). Let $C_{\text{other},+}$ be the positive nonfood-input reward of all catalytic pathways other than the two local directions of (z,r), on the *original* reactor. Its drift at any population equals the positive catalytic nonfood-input drift of the erased source at the same population, because the reward is simply zero on the two local channels; this is an identity between instantaneous rate sums and not a transfer between trajectory laws. Hence its drift is at most $t_{K_*} = 85184/K_*$ on $L \leq 11$, while B_+ has drift at most b. Both rewards have jumps in $[0, 4/V]$ and satisfy the variance bound (A.3) under the original total rate. Apply (A.4) to each with $d = 1/200$, allowance $1/100$, $T = 101$; a union bound gives an exceptional probability at most $4e^{-cV/n}$, uniform over environments in U_{n,K_*}. On the complementary bounded-mass paths

$$B_+(100) + C_{\text{other},+}(100) \leq 100\Big(1936\varepsilon + \frac{85184}{10^{12}}\Big) + \frac{2}{100} = 0.0203957184 < \frac{1}{40}. \tag{7.2}$$

Pathwise nonfood mass balance with $M(0) = 0$ reads

$$M(100) + \frac{W((0,100])}{V} = J_{\text{loc}}(100) + J_{\text{other}}(100) + J_{\text{basal}}(100),$$

where the three signed inputs are the local, other-catalytic, and basal contributions. Since $M(100) \geq 0$, $W((0,100]) \geq W((1,100])$, and signed inputs are at most their positive parts,

$$J_{\text{loc}}(100) \geq E - B_+(100) - C_{\text{other},+}(100) > E - \tfrac{1}{40} > \tfrac{3}{4}E > \tfrac{3}{40}$$

on $\mathcal{F}_n$ outside the exceptional event. The incidence is determined by the environment, so no union over candidate incidences is needed. Averaging over U_{n,K_*} and adding $\mathbb{P}(\mathcal{F}_n \cap U_{n,K_*}^c) = o(p_n)$ gives a total exception of $o(p_n)$; divide by $s_n \geq c_0 p_n$. □

8 Discussion and scope

Selection by an operational event. The same sampler gives three different answers to "what does a typical environment contain", depending on the conditioning (Table 2). RAF existence has a positive limit $\theta(q_*)$, and under that conditioning the minimum RAF is large: $M_n \to \infty$ in probability, although subexponentially. Singleton RAFs have probability $248p_n + o(p_n)$, exponentially small. Productive output has probability $\Theta(p_n)$, and conditioning on it selects, with probability tending to

Question	Answer proved here
What does a typical successful realization contain?	A productive singleton incidence, unique in the K_*-rectangle, with probability $\to 1$ (Theorems 3.1, 3.2(a)).
Is the singleton only a coexisting structural certificate?	No: deleting it makes success exponentially unlikely, and it supplies more than $\frac{3}{4}$ of the measured export (Theorem 3.2(c),(d)).
Do ordinary RAF-containing environments contain a singleton?	Almost never: $\mathbb{P}(M_n = 1 \mid G_n) \to 0$ and $M_n \to \infty$ in probability given G_n (Theorem 3.3).
Must every successful realization contain a singleton, or any RAF?	Not proved; the theorem is about typicality.
Does every eligible singleton guarantee success?	Not proved; "productive" names a structural type.
Is the rest of the chemistry inactive in a successful run?	Not claimed; localization concerns a fixed rectangle and one net-input observable.
Must a one-channel RAF copy its catalyst?	No; 192 of the 224 candidates have a food catalyst.

Table 2: What the theorems decide and what they leave open.

one, an environment containing one of the 224 short productive incidences, exactly one incidence in a huge fixed neighbourhood, and nothing else short. The selection is by the operational event, not by an alteration of the chemistry or of the RAF definition, and the strength of the conclusion is exactly that the exceptional set is $o(p_n)$ rather than $o(1)$.

The mechanism. The kinetic exclusion is a resource and startup constraint made quantitative by two observables. A mixed channel can incorporate food while cycling existing nonfood material, but it cannot create the credited mass needed to initialize that material; the product credit $M - a\,x_p$ records this as an exact cancellation. An outsider catalyst can accelerate food conversion, but in a locally sparse environment it is neither fed nor made by the channel it accelerates, and its washout removes it faster than it can pay for itself; the penalty $M + 2000\,x_z$ records this with the margin $2000 - 1936$. A productive singleton escapes both obstructions because its catalyst is either supplied by the feed or made by the same food-enabled channel. A large catalytic degree by itself does not defeat the mass-tail bound: the envelopes (4.2) allow every nonlocal assignment. This is why the result needs no sparsity of the global network and no bound on the number of catalysts a molecule can have.

A transferable pattern. The structure of the argument is not specific to the observables. Let μ_n be environment laws, $\pi_n \in [0, 1]$ success probabilities, T_n candidate events, and N_n nonnegative integer local counts, and suppose for some scale $p_n > 0$: (1) $\int \pi_n \, d\mu_n \geq c_0 p_n$ eventually; (2) $\pi_n \leq b_n = o(p_n)$ on $T_n^c \cap \{N_n \leq 1\}$; (3) $\mu_n(N_n \geq 2) = o(p_n)$. Then $\mu_n^{\mathrm{succ}}(T_n^c) \leq (b_n + \mu_n(N_n \geq 2))/(c_0 p_n) \to 0$, and if also $\mu_n(A_{n,\eta}) \geq c_\eta p_n$ the reliable-environment conclusion follows as in (7.1). In the present paper (1) is the constructive witness, (2) the two observables, and (3) the excess-count estimate; each is replaced separately if the source or reactor changes. An independent-row source with common marginal $p_n \to 0$ and same-row pair probability $q_{2,n} = o(p_n)$ satisfies (3) automatically by (5.5), but that alone supplies neither (1) nor (2).

The screen interpretation and its limits. The 224 candidate incidences form an asymptotically sensitive screen among successful realizations: the conditional false-negative probability tends to zero, at rate $O(1/n)$ by Proposition 5.3. The theorems do not give a low false-positive rate, a sufficiency test for any candidate, or the leading success constant; Proposition 3.4 states exactly what is missing. The volume envelope (3.1) is a sufficient condition of the proof method, chosen so that $e^{-cV(n)/n} = o(p_n)$ with ample slack; it is not a lower bound on physically necessary reactor volume, and the huge cutoffs k_0, K_* prevent any reading of the result as a calibrated laboratory threshold.

Open questions. Exceptional productive systems with large minimum RAFs, or with no RAF at all, are compatible with every theorem here. Establishing a productive 16-channel catalytic cycle at $n = 4$, a productive acyclic organization without any RAF, a uniform theorem for arbitrary finite catalytic support, a smaller sufficient volume, or the limits of the $\alpha_i(n)$ each requires an argument beyond this paper. None alters the conditional assertion proved: in the stated source and volume regime, and after conditioning on productive output, the minimum RAF has size one and the one-channel RAF is the mechanism.

A Capacity, target, and stopped-reward bounds

Write $Q = \sum_z x_z$; then $Q \le L$ and $\sum_{\ell(z)>K} x_z \le L/K$. For every count propensity $(N_z)_h/V^h \le x_z^h$, so upper drift bounds follow from nonnegative concentration monomials without any continuum approximation.

A.1 Catalog counting and coordinate bounds

The map from a channel to its product word and left-input length is injective. A fixed target q is produced by at most one ligation for each ordered pair of reactants, and by the cleavage of any fixed parent at most twice (both fragments may equal q). The unit-coefficient positive target-creation sum is therefore at most

$$Q^2 + 2Q; \tag{A.1}$$

the multiplicity two is real and cannot be removed by treating the fragments as a set. Losses of a target are bounded by counting its occurrences in the two input positions and the product position, which gives the basal loss bound $4\varepsilon(2L^2 + L) \le 1012\varepsilon$ at $L \le 11$.

For a nonfood target the basal creation bound sharpens. Let $A = \sum_{\ell(z)=1} x_z$. Monomer–monomer pairs cannot make a nonfood word, so forward creation is at most $Q^2 - A^2$; since $L \ge 2Q - A$ and $(Q - 2A)^2 \ge 0$,

$$3(Q^2 - A^2) \le (2Q - A)^2 \le L^2.$$

A parent cleaving to a nonfood target has length at least four, so the two-occurrence bound makes reverse creation at most $L/2$. Multiplying by 4ε,

$$4\varepsilon\Big(\frac{L^2}{3} + \frac{L}{2}\Big) \le \frac{550}{3}\varepsilon. \tag{A.2}$$

This sharpening is what makes the outsider budget (4.9) close; the crude $Q^2 + 2Q$ would not.

For catalytic bounds in a source with no incidence in the K-rectangle, split channels by whether the product length is at most $K + 2$. A short-product channel has catalyst length greater than K, so the mass-tail bound applies to the catalyst sum; a long product created by cleavage of a specified short target has a parent of small concentration; a long product consuming a target of length at most $K/2$ in its forward direction has another input of length greater than $K/2$. Applying the two positional multiplicities, (A.1), and the coefficient cap 16:

Quantity, after local erasure	Envelope at mass L	At $L \le 11$
Positive catalytic nonfood input	$64L^3/K$	$85184/K$
Loss of a target with $\ell(q) \le K/2$	$(16L^2 + 32L^3)/K$	$44528/K$
Creation of a target with $2 < \ell(q) \le K+2$	$(16L^3 + 64L^2)/K$	$29040/K$

For the first row write $Q_K = \sum_{\ell(z)>K} x_z$ and $F_0 = \sum_{f\in F} \ell(f)x_f$. The nonfood gain of a ligation is at most the food weight among its two inputs. For each food input position, absence of a local incidence forces either the catalyst or the other input to have length greater than K; summing over the two positions and the two alternatives bounds the unit-coefficient food-input sum by $4Q_K F_0 Q \le 4L^3/K$, and the cap 16 gives $64L^3/K$. The second row separates product loss with a long catalyst ($16L^2/K$) from target consumption with a long catalyst or a long partner ($32L^3/K$). The third separates long-catalyst forward production ($16L^3/K$) from cleavage contributions, whose two target occurrences and long-parent-or-long-catalyst partition give $64L^2/K$. These envelopes may count overlapping contributions more than once; none uses a bound on catalytic degree. Positive basal nonfood input is at most $16\varepsilon L^2 \le 1936\varepsilon$. In (4.6) the washout supplies $-x_z$, the feed contributes zero, and the local incidence has zero net z-increment; in (4.5) the target p has length at most $k_0 + 2 < K_*/2$ as the loss bound requires.

A.2 The stopped reward estimate and the physical time window

The following estimate is proved for the actual jump kernel in the formal development (`physical_marked_reward_two_sided_interval_tail`). Let $\rho(N, j)$ be a reward attached to the jump of type j from population N with $|\rho| \le n/V$, and suppose that whenever $L \le 11$,

$$\sum_j \lambda_j(N)\rho(N,j)^2 \le \frac{384000 + 11n}{V}. \tag{A.3}$$

If $d > 0$, $T \ge 0$, and $d + n/V \le h$, then the probability that the absolute compensated reward exceeds h anywhere on the physical interval $[0, T]$, stopped at the first exit from $L \le 11$, is at most

$$2\exp\Big(-\frac{d^2V}{4n(96011T + d)}\Big). \tag{A.4}$$

The statement covers the unfinished holding interval at the observation time, not only jump epochs; n/V is the interval-boundary allowance. The proof compensates each marked jump by its conditional rate, uses the exponential waiting-time law and $e^u - 1 - u \le u^2$ for $|u| \le 1$ to bound the logarithmic transform by the accumulated quadratic rate, stops the resulting nonnegative exponential supermartingale at the first crossing, and uses both signs; passing through finite jump prefixes and the censored last waiting time, and identifying the countable-prefix formulation with the physical event through nonexplosion, gives (A.4). General martingale tail bounds [7] and Markov-chain approximation results [4] provide context; the constants and the interval adapter are checked for this kernel. Censoring at the mass exit is essential, and it is legitimate because the output event itself supplies the bounded-mass path.

All rewards used in Sections 4.2, 4.3, and 7 have absolute jump at most $4/V \le n/V$. On $L \le 11$ the total physical rate, with basal coefficients at most one and catalytic coefficients at most 16, is at most $24000V$, so $\sum_j \lambda_j\rho_j^2 \le 24000V(4/V)^2 = 384000/V$, verifying (A.3). For the signed outsider reward the coordinate term includes washout, and its jump estimate covers external channels as well. For the masked reward $C_{\mathrm{other},+}$ of Theorem 3.2(d), setting two rewards to zero leaves the variance argument valid under the original law.

The proofs use $T = 101$ to cover the horizon 100 and its censored interval representation. The volume envelope gives $n/V \le 1/200000$, the scale used in (2.14), and the margins $d_m + n/V \le 1/100$ and $d_o + n/V \le 1/100000$. Since

$$c \le \frac{d_m^2}{4(96011 \cdot 101 + d_m)}, \qquad c \le \frac{d_o^2}{4(96011 \cdot 101 + d_o)},$$

the three terms of (4.1) are bounded by $8e^{-cV/n}$, and the current-attribution exception by $4e^{-cV/n}$.

B Formal concordance and reproducibility

The Lean development is a single project pinned to `leanprover/lean4:v4.30.0` with Mathlib at commit `c5ea00351c28e24afc9f0f84379aa41082b1188f`. The reactor namespace is `RandomViability`; it imports the source namespace `PowerLawSmallRAF`, and both use the configuration weight `sourcePowerLawConfigWeight`, which is (2.6). The productive event is `physicalOutputEvent` (mass at most $11V$ through time 100 and windowed export reward above 1/10 on $(1, 100]$), the joint probability `outputJointProbability` (source average, uniform mark average, and trajectory law), the incidence probability `productiveSourceIncidence` $= p_n$, the noise rate `markedNoiseRate` $= c$, the cutoffs `collectiveUpperCutoff` $= k_0$ and `singleIncidenceCutoff` $= K_*$, and the volume envelope `collectiveMinimalVolume` $= 10^{60}(n+1)^2$. The productive-incidence predicate is `ProductiveSingletonIncidence`, the RAF event `anySourceRAF`, and the singleton event `hasSingletonRAF`.

Two strict verification receipts accompany the paper. Each records exit code zero, warnings treated as errors, no `sorry`, the axioms `Classical.choice`, `Quot.sound`, and `propext` only, and the SHA-256 of every compiled source in the import closure (407 and 408 modules).

Paper statement	Lean declaration (file in `proofs/RandomViability/` unless stated)
Prop. 4.1, pointwise exclusion	`physical_output_upper_at_most_one_local` (`LocalOutputExclusion.lean`); hypotheses: no productive incidence, at most one incidence in the K_*-rectangle, food-only initial state, the rate caps, and the volume margins
Prop. 5.1, source multiplicity	`two_local_incidence_mass_div_incidence_tendsto_zero` (`LocalMultiplicity.lean`), for every fixed cutoff
Thm. 3.1, (3.3)	`output_without_productive_singleton_relative_tendsto_zero` (`SingletonDominance.lean`); root receipt on `Remark34.lean`, source SHA-256 `2fec12f3...`, 80.4 s
Lemma 7.1, first sentence ($\Leftarrow$)	`productive_incidence_singleton_isRAF` (`ProductiveSingletonRAF.lean`)
Thm. 3.1, $\mathbb{P}(M_n = 1 \mid \mathcal{F}_n \cap G_n) \to 1$	`remark34_no_singleton_given_productive_raf_tendsto_zero` with `singleton_raf_minimum_card` (`Remark34.lean`); the denominator is output jointly with RAF existence
Thm. 3.1, output-only posteriors	`no_RAF_given_output_tendsto_zero`, `no_singleton_RAF_given_output_tendsto_zero` (`PublicationFollowups.lean`); receipt source SHA-256 `c370c68c...`, 217.5 s
Thm. 3.2(c), pointwise bound (3.5)	`deleted_unique_incidence_output_upper` (`PublicationFollowups.lean`), for the erased source `eraseLocalIncidence`
Prop. 2.2(v)	`output_joint_probability_theta` (`OutputJointLimit.lean`)
Prop. 2.2(ii)	`PowerLawSmallRAF.source_RAF_probability_tendsto` (`proofs/PowerLawSmallRAF/Main.lean`), limit `(staticSurvival sourceCriticalOpenness).toReal`, positive by `sourceCriticalSurvival_pos`
Appendix A.2, (A.4)	`physical_marked_reward_two_sided_interval_tail` (`MarkedRewardInterval.lean`)

Conventional proofs. Lemma 7.1 (the counts), Theorem 3.2(a), (b), the typicality parts of (c), and (d), Theorem 3.3 in full, Proposition 3.4, and Proposition 5.3 are proved in the text and have no separate terminal Lean export. The current-attribution proof verifies the reward hypotheses in the original law explicitly rather than transferring a concentration estimate between reactors. The Lean statements are event-level: M_n is not a formal random variable, and the minimum-cardinality reading is the elementary interpretation of `singleton_raf_minimum_card` on the finite configuration space.

Reproduction. The exact-arithmetic script `check_paper.py` distributed with the source replays every number printed in this paper: the catalog counts 30, 68, 36, 224, 248, 192, 32; the envelopes (4.2); the budgets (4.5), (4.9), (7.2); the noise constants and the inequalities $c \le d_m^2/[4(96011 \cdot 101 + d_m)]$, $cV(n)/n \ge 3n$; and the bounds in Proposition 5.3. It performs no simulation. No numerical simulation is used as evidence anywhere in this paper. Proof verification uses the repository bridge, which sets the Mathlib working directory, treats warnings as errors, and writes the receipts named above; no dependency update is required.

References

[1] Author name to be inserted. Structural autocatalysis can be common while productive operation is rare: an exact separation theorem in a capped-Zipf binary polymer model. Companion manuscript, 2026; arXiv preprint in preparation, 2026.

[2] David F. Anderson and Thomas G. Kurtz. Continuous time Markov chain models for chemical reaction networks. In Heinz Koeppl, Gianluca Setti, Mario di Bernardo, and Douglas Densmore, editors, *Design and Analysis of Biomolecular Circuits*, pages 3–42. Springer, New York, 2011. doi: 10.1007/978-1-4419-6766-4_1.

[3] 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.

[4] R. W. R. Darling and J. R. Norris. Differential equation approximations for Markov chains. *Probability Surveys*, 5:37–79, 2008. doi: 10.1214/07-PS121.

[5] Leonardo de Moura and Sebastian Ullrich. The Lean 4 theorem prover and programming language. In *Automated Deduction – CADE 28*, volume 12699 of *Lecture Notes in Computer Science*, pages 625–635. Springer, 2021. doi: 10.1007/978-3-030-79876-5_37.

[6] J. Doyne Farmer, Stuart A. Kauffman, and Norman H. Packard. Autocatalytic replication of polymers. *Physica D: Nonlinear Phenomena*, 22(1–3):50–67, 1986. doi: 10.1016/0167-2789(86)90233-2.

[7] David A. Freedman. On tail probabilities for martingales. *The Annals of Probability*, 3(1): 100–118, 1975. doi: 10.1214/aop/1176996452.

[8] Varun Giri and Sanjay Jain. The origin of large molecules in primordial autocatalytic reaction networks. *PLoS ONE*, 7(1):e29546, 2012. doi: 10.1371/journal.pone.0029546.

[9] 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.

[10] Wim Hordijk and Mike Steel. Autocatalytic sets extended: dynamics, inhibition, and a generalization. *Journal of Systems Chemistry*, 3:5, 2012. doi: 10.1186/1759-2208-3-5.

[11] Wim Hordijk and Mike Steel. Autocatalytic sets in polymer networks with variable catalysis distributions. *Journal of Mathematical Chemistry*, 54(10):1997–2021, 2016. doi: 10.1007/s10910-016-0666-z.

[12] Wim Hordijk, Stuart A. Kauffman, and Mike Steel. Required levels of catalysis for emergence of autocatalytic sets in models of chemical reaction systems. *International Journal of Molecular Sciences*, 12(5):3085–3101, 2011. doi: 10.3390/ijms12053085.

[13] Wim Hordijk, Leonard Hasenclever, Jie Gao, Dilyana Mincheva, and Jotun Hein. An investigation into irreducible autocatalytic sets and power law distributed catalysis. *Natural Computing*, 13(3):287–296, 2014. doi: 10.1007/s11047-014-9429-6.

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

[15] 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.

[16] Elchanan Mossel and Mike Steel. Random biochemical networks: the probability of self-sustaining autocatalysis. *Journal of Theoretical Biology*, 233(3):327–336, 2005. doi: 10.1016/j.jtbi.2004.10.011.

[17] Praneet Nandan, Philippe Nghe, and Jérémie Unterberger. Autocatalytic cores in the diluted regime: classification and properties. *Journal of Mathematical Biology*, 92:36, 2026. doi: 10.1007/s00285-026-02357-7.

[18] Zhen Peng, Jeff Linderoth, and David A. Baum. The hierarchical organization of autocatalytic reaction networks and its relevance to the origin of life. *PLoS Computational Biology*, 18(9): e1010498, 2022. doi: 10.1371/journal.pcbi.1010498.

[19] Mike Steel. The emergence of a self-catalysing structure in abstract origin-of-life studies. *Applied Mathematics Letters*, 13(3):91–95, 2000. doi: 10.1016/S0893-9659(99)00191-3.

[20] Mike Steel, Wim Hordijk, and Joshua Smith. Minimal autocatalytic networks. *Journal of Theoretical Biology*, 332:96–107, 2013. doi: 10.1016/j.jtbi.2013.04.032.

[21] Mike Steel, Wim Hordijk, and Joana C. Xavier. Autocatalytic networks in biology: structural theory and algorithms. *Journal of the Royal Society Interface*, 16(151):20180808, 2019. doi: 10.1098/rsif.2018.0808.

[22] 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.

[23] Jérémie Unterberger and Philippe Nghe. Stoechiometric and dynamical autocatalysis for diluted chemical reaction networks. *Journal of Mathematical Biology*, 85:26, 2022. doi: 10.1007/s00285-022-01798-0.