ENTROPY PRODUCTION UNDER HIDDEN-STATE REDUCTION: EXACT SAFETY AND NON-TESTABILITY

SAMUIL PETKOV

Abstract. Restricting hidden-state complexity can raise the least entropy production compatible with partial observations. We determine exactly when a two-hidden-state fit preserves this lower bound for finite irreducible stationary reciprocal continuous-time Markov chains with two individually observed states and one hidden label. We begin with an explicit chain with three hidden states whose observed data admit exact two-hidden-state fits, every one of which has entropy production more than sixty times that of the true chain. For strictly positive consistent stationary masses, flows, excursion counts and first duration moments, the unrestricted infimum is an explicit count functional, and four hidden states approach it while preserving every datum exactly. We classify finite attainment and safety of feasible two-hidden-state fits in every determinant regime; in the positive-determinant regime, safety is equivalent to equality of the conditional cross-excursion means, and explicit bounds quantify the reduction bias. Exact classification does not imply finite-data detectability: rare long excursions can conceal a fixed positive bias on any prescribed finite interval of stationary label observation, even under a common upper rate bound.

hexagon:2611.00657v1

1. Introduction

Dissipation inference from a partially observed trajectory depends on the class of models used to fit it. Minimizing entropy production over a class containing the microscopic model gives a lower bound on its dissipation. A hidden-state budget may exclude that model and raise the fitted minimum. The question is then whether the restricted class preserves the unrestricted dissipation floor. We answer this question for first excursion-duration moments with two visible states. The comparison concerns compatible models; a physically faithful coarse-graining has additional closure requirements [7, Sections III.A–III.B].

The observed stationary label trace distinguishes two visible states and combines all hidden states into one instantaneous, reversal-even label. Its underlying model is a finite irreducible reciprocal continuous-time Markov chain. A hidden excursion runs from a departure into the hidden set to the next visible visit, including returns and excluding visible holding time. The data $d = (P, F, R, L)$ comprise stationary masses, observed flows, excursion occurrence rates and occurrence-weighted first duration moments. We compare the entropy infimum $I_2(d)$ over compatible models with at most two hidden states with $I_\infty(d)$ over all finite hidden dimensions. A nonempty reduced class is *safe* when these infima agree.

Nitzan–Ghosal–Bisker formulate the corresponding first-moment fitting problem [15, Eq. (17), Section III.B.1 and Appendix C]; their simpler-topology experiments leave the general lower-bound validity of such fits open [15, Section III.D and Conclusion]. Skinner–Dunkel optimize general asymmetric two-step counts by hidden-edge rerouting and finite channel reduction [20, p. 3; author-manuscript Appendix, Eqs. (A12)–(A27)], and give a closed-form symmetric negative-determinant optimum [20, author-manuscript Appendix, Eqs. (A28)–(A29)]. The new ingredients here are the explicit count formula for arbitrary asymmetric data together with exact restoration of prescribed first moments, and the resulting safety classification, finite attainment criterion and non-testability result.

Results. Section 3 gives an explicit example: a true chain with three hidden states whose data admit exact two-hidden-state fits, all of which have entropy production above $7/100$ and more than sixty times the true value, which tends to zero as a parameter $\varepsilon \downarrow 0$ (Theorem 3.1). A computer-assisted step in exact arithmetic raises these bounds to $7/5$ and 1200 times the true value (Theorem 3.2). The reduced minimum therefore overestimates the true entropy production, not only the unrestricted fitting floor. The remaining sections explain this phenomenon. Theorem 4.1 and Theorem 9.1 give $I_\infty(d) = I_4(d) = \mathcal{B}(d)$ for every strictly positive consistent datum, where the sharp count functional $\mathcal{B}$ is defined in (9.2). Thus exact first duration moments do not raise the unrestricted count infimum. Four hidden states suffice to approach it at exactly the prescribed data with uniformly bounded rates: rare excursions carry the

Date: 5 September 2026.

2020 *Mathematics Subject Classification.* Primary 82C05; Secondary 60J27, 62M05.

Key words and phrases. Entropy production, coarse-graining, hidden Markov chains, thermodynamic inference, phase-type distributions, excursion durations, reduction bias.

required moments while their occurrence rates vanish. Theorem 10.1 classifies finite attainment in every determinant regime, and Corollary 10.3 classifies safety of nonempty two-hidden-state classes. For positive determinant, safety is equivalent to equal conditional cross-excursion means (Theorem 4.2); unequal means force the quantitative bias of Theorem 4.3. Rank-one data can be safe without a finite minimizer (Theorem 10.2).

Finally, Proposition 11.1 fixes safe and positively biased unsafe data that share P, F, R, yet makes their stationary label-path laws arbitrarily close on any prescribed finite interval. Four hidden states and a common upper rate bound suffice. Rare long excursions carry the distinguishing moments, a constrained Markov realization of the classical rare-contamination mechanism in mean testing [2].

Scope. All results concern the stated positive-data, two-visible-state class and compatible fits. The entropy σ is stationary path irreversibility; identifying it with total physical dissipation requires thermodynamically resolved transitions obeying local detailed balance. Safety is a statement about infima: it does not say that every individual fit has the same dissipation, identifies the true hidden network, or covers incorrect hidden topologies [11, Section V]. Four hidden states are a sufficient budget, not shown to be minimal, and the gap constant in Theorem 4.3 is not claimed to be optimal. Positive-determinant, rank-one and negative-determinant data are treated separately.

Related work. The entropy tools build on path irreversibility and coarse-grained relative entropy [17–19], trajectory snippets [5, Sections II.C–II.E], and the cross-excursion waiting-time bound [14, Eq. (11)]. Waiting-time inference [21] and sequence-versus-duration comparisons [12, Sections II.B–II.C] use observation protocols distinct from fixed first moments. Ehrich's reconstruction fixes the hidden-state count [6, Section 4]. Phase-type resolvents and realization theory provide classical algebraic ingredients [4, 10, 22]. Fixing moments also differs from preserving memory [23], resolving transitions [24], or controlling hidden pumps [8]. Recent preprints study attainment with the entire stationary path law fixed [1, Section 2 and Theorems A–B] and full excursion-kernel reconstruction under known event anchors and closure [9, Eqs. (4)–(5) and End Matter]; neither observation contract determines the fixed-first-moment infimum studied here.

2. Setting and notation

A microscopic model is a finite irreducible continuous-time chain with finite rates and reciprocal support: a transition is present precisely when its reverse is present. Missing reciprocal edge pairs are allowed. The row generator is Q, stationary probabilities are π, and stationary directed flows are $n_{xy} = \pi_x q_{xy}$. Logarithms are natural, and entropy is measured in units of Boltzmann's constant. Entropy production means

$$\sigma(Q) = \sum_{x<y} (n_{xy} - n_{yx}) \log(n_{xy}/n_{yx}),$$

with an absent pair contributing zero. The labels are unchanged by time reversal. There are two individually visible states, $1, 2$, and one observed label for all hidden states. A hidden excursion includes its whole hidden duration, excludes visible holding times, and includes return excursions. Write R_{ij} for its stationary occurrence rate and $L_{ij} = R_{ij}\mathbb{E}[T \mid i \to H \to j]$. Let $f_{ij}(t)$ be the stationary occurrence-rate density of excursions $i \to H \to j$ of duration t, so that $R_{ij} = \int_0^\infty f_{ij}\, dt$ and $L_{ij} = \int_0^\infty t f_{ij}\, dt$.

The masses are $P_i = \pi_i$ and $P_H = \sum_{h\in H} \pi_h$. For distinct observed labels $a \neq b$, the coarse flow F_{ab} sums stationary microscopic flows from block a to block b. Internal hidden jumps do not change the observed label and are not entries of F. The label map is instantaneous and reversal even, so observation and reversal commute. Mathematically σ is stationary path irreversibility; its identification with total physical dissipation requires thermodynamically resolved transitions obeying local detailed balance [19].

For the positive-determinant results, the fixed data are $d = (P, F, R, L)$. The fixed-data infimum and safety theorems below assume:

- $P_1, P_2, P_H > 0$ and their sum is one;
- $F_{12}, F_{21} > 0$, and every entry of R, L is strictly positive;
- the consistency identities

$$R\mathbf{1} = (F_{1H}, F_{2H})^T, \qquad \mathbf{1}^T R = (F_{H1}, F_{H2}),$$
$$\mathbf{1}^T L \mathbf{1} = P_H, \qquad F_{12} - F_{21} = R_{21} - R_{12}.$$

- $\det R > 0$.

All rates use a fixed time unit. P, L are dimensionless, while F, R, σ have inverse-time units. Put

$$B(d) = (F_{12} - F_{21}) \log(F_{12}/F_{21}) + (R_{12} - R_{21}) \log(R_{12}/R_{21}). \tag{2.1}$$

Let $\mathcal{C}_k(d)$ be all compatible models with at most k hidden states, $I_k(d) = \inf_{Q \in \mathcal{C}_k(d)} \sigma(Q)$, and let I_∞ denote the infimum over any finite hidden dimension. This is a union over finite models, not an infinite-state chain. Infima over empty classes are not used as safety certificates.

3. A five-state example: a feasible reduced fit that overestimates

Consider a true chain with two visible states and only three hidden states. Its observed data admit exact fits with two hidden states, so the reduced class is feasible, yet every such fit has entropy production far above that of the true chain, and the relative overestimate diverges as a parameter $\varepsilon \downarrow 0$. The theorems of the later sections explain this behaviour in general: the unrestricted floor is reached with four hidden states, while for unequal conditional cross-excursion means two hidden states cannot reach it.

3.1. The true family and exact reduced fits.

In the order $(1, 2, h_1, h_2, h_3)$ set

$$\pi = (11/40, 11/40, 1/10, 3/10, 1/20), \qquad N_\varepsilon = \begin{pmatrix} 0 & 1 & 3 & 1 & 3\varepsilon/4 \\ 1+\varepsilon/2 & 0 & 1 & 3 & \varepsilon/4 \\ 3 & 1 & 0 & 1 & 0 \\ 1 & 3 & 1 & 0 & 0 \\ \varepsilon/4 & 3\varepsilon/4 & 0 & 0 & 0 \end{pmatrix}, \quad 0 < \varepsilon \le 10^{-3}. \tag{3.1}$$

Rows and columns of the flow table have the same sums. Define off-diagonal rates by $q_{xy} = N_{xy}/\pi_x$ and choose the negative row sum on the diagonal. The graph is connected and reciprocal, so the model is admissible. Direct edge calculation gives

$$\sigma(Q_\varepsilon) = \varepsilon \log 3 + \frac{\varepsilon}{2} \log(1 + \varepsilon/2) \le \frac{8001}{4000000} < \frac{1}{400}. \tag{3.2}$$

The small direct-edge term has coefficient $\varepsilon/2$.

Put

$$R_0 = \frac{1}{3}\begin{pmatrix} 7 & 5 \\ 5 & 7 \end{pmatrix}, \quad S = \frac{1}{16}\begin{pmatrix} 3 & 9 \\ 1 & 3 \end{pmatrix}, \quad L = \frac{1}{2880}\begin{pmatrix} 251 & 337 \\ 265 & 443 \end{pmatrix}. \tag{3.3}$$

The observations are $R_\varepsilon = R_0 + \varepsilon S$, this fixed L, visible masses $11/40, 11/40$, and direct flows $1, 1 + \varepsilon/2$. To verify R, L, the fast hidden pair has entrance and exit flow matrix $B_f = \left(\begin{smallmatrix} 3 & 1 \\ 1 & 3 \end{smallmatrix}\right)$, killed flow matrix $J_f = \left(\begin{smallmatrix} 5 & -1 \\ -1 & 5 \end{smallmatrix}\right)$, and hidden masses $D_f = \mathrm{diag}(1/10, 3/10)$. Its matrices are

$$B_f J_f^{-1} B_f = R_0, \qquad B_f J_f^{-1} D_f J_f^{-1} B_f = \frac{1}{90}\begin{pmatrix} 7 & 8 \\ 8 & 13 \end{pmatrix}.$$

The slow state's occurrence and first-moment contributions are $\varepsilon S, S/20$, because its mean residence time is $1/(20\varepsilon)$.

An exact two-hidden-state fit exists throughout the interval. Define

$$X = \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix}, \quad Y = \begin{pmatrix} 55/84 & 47/116 \\ 29/84 & 69/116 \end{pmatrix}, \quad D_h = \mathrm{diag}(7/80, 29/80), \tag{3.4}$$

$$U_\varepsilon = R_\varepsilon X^{-1}, \quad V_\varepsilon = Y^{-1} R_\varepsilon, \quad J_\varepsilon = Y^{-1} R_\varepsilon X^{-1}. \tag{3.5}$$

Use U, V for entrance and exit flows, $-(J_\varepsilon)_{12}, -(J_\varepsilon)_{21}$ for hidden internal flows, and hidden masses from D_h. Keep the visible masses and direct flows. Rows of X and columns of Y sum to one, giving $J_\varepsilon \mathbf{1} = V_\varepsilon \mathbf{1}$ and $\mathbf{1}^T J_\varepsilon = \mathbf{1}^T U_\varepsilon$. These establish hidden balance and identify the killed flow matrix. Visible balance follows from $(R_\varepsilon)_{12} - (R_\varepsilon)_{21} = \varepsilon/2$.

All these flows are affine in ε. At zero,

$$U_0 = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}, \quad V_0 = \frac{1}{76}\begin{pmatrix} 217 & 14 \\ 87 & 290 \end{pmatrix}, \quad J_0 = \frac{1}{76}\begin{pmatrix} 420 & -189 \\ -116 & 493 \end{pmatrix}.$$

At $\varepsilon = 1/1000$ the complete four-state flow table is

$$\frac{1}{304000}\begin{pmatrix} 0 & 304000 & 911943 & 304285 \\ 304152 & 0 & 303981 & 912095 \\ 868105 & 56315 & 0 & 755475 \\ 347971 & 1159913 & 463971 & 0 \end{pmatrix}. \tag{3.6}$$

Positivity at both endpoints proves reciprocal complete support throughout. Finally

$$U_\varepsilon J_\varepsilon^{-1} V_\varepsilon = R_\varepsilon, \qquad U_\varepsilon J_\varepsilon^{-1} D_h J_\varepsilon^{-1} V_\varepsilon = Y D_h X = L.$$

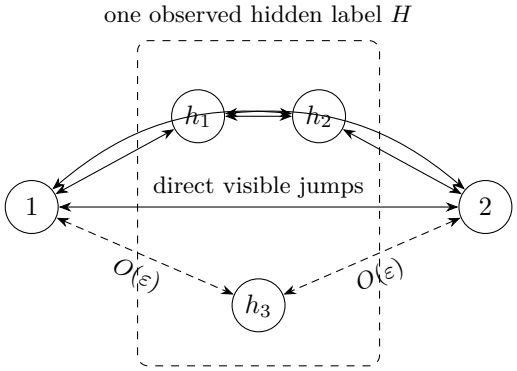


FIGURE 1. True hidden subnetwork in (3.1). The fast pair h_1, h_2 remains strongly connected to both visible states, while the rates through h_3 are proportional to ε. Its stationary mass stays $1/20$, so its long residence times contribute a finite amount to the first moments. All depicted links are reciprocal. The direct visible link is also present.

Thus these are exact feasible fits, not approximate numerical reconstructions.

3.2. **A weighted excursion bound.** For $s > 0$ let $W_{ij}(s) = \int_0^\infty (1 - e^{-st}) f_{ij}(t)\, dt$. Pairing each microscopic path with its reverse gives a nonnegative integrand $(w - w^{\mathrm{rev}}) \log(w/w^{\mathrm{rev}})$. Multiplication by the reversal-invariant weight $1 - e^{-st} \in [0, 1]$ decreases it. Discard returns and apply log-sum on the cross class. This proves

$$\sigma(Q) \geq (W_{12} - W_{21}) \log(W_{12}/W_{21}) \geq \frac{2(W_{12} - W_{21})^2}{W_{12} + W_{21}}. \tag{3.7}$$

Stationary normalization is kept in this argument: W consists of rates, not separately normalized conditional probabilities.

For any two-hidden-state exact fit, (7.2) gives $W = R - R(R+L)^{-1}R$ at $s = 1$ in the fixed inverse-time unit. At $\varepsilon = 1/1000$,

$$W = \frac{1}{996369600} \begin{pmatrix} 82443151 & 109353773 \\ 86314757 & 143879311 \end{pmatrix},$$

and hence every fit satisfies

$$\sigma(Q') \geq \frac{2764563845043}{507703580647625} > \frac{1}{200} > \sigma(Q_{1/1000}) + \frac{1}{400}. \tag{3.8}$$

As $\varepsilon \downarrow 0$, the forced matrix converges to

$$W_0 = \frac{1}{62277} \begin{pmatrix} 5153 & 6835 \\ 5395 & 8993 \end{pmatrix},$$

whose bound in (3.7) is $138240/25388257 > 1/200$. Continuity therefore keeps the fitted infimum bounded below while (3.2) tends to zero. The relative overestimate diverges. No true endpoint generator at $\varepsilon = 0$ is used as an admissible model. Theorem 3.1 below replaces this endpoint-and-limit argument by a bound that holds uniformly on the whole parameter interval and is larger by an order of magnitude.

3.3. **Every reduced fit overestimates the true entropy production.** Write d_ε for the observed data of Q_ε. The true chain has three hidden states, so $Q_\varepsilon \in \mathcal{C}_3(d_\varepsilon)$ and $I_3(d_\varepsilon) \leq \sigma(Q_\varepsilon)$.

Theorem 3.1 (Two-hidden-state fits overestimate the true value)**.** *For every $0 < \varepsilon \leq 10^{-3}$ the class $\mathcal{C}_2(d_\varepsilon)$ is nonempty, its infimum is attained, and*

$$I_\infty(d_\varepsilon) = I_4(d_\varepsilon) \leq I_3(d_\varepsilon) \leq \sigma(Q_\varepsilon) < \frac{11}{10^4}, \qquad I_2(d_\varepsilon) \geq I_\infty(d_\varepsilon) + \frac{7}{100}. \tag{3.9}$$

Consequently every exact fit Q' with at most two hidden states satisfies

$$\sigma(Q') > \sigma(Q_\varepsilon) + \frac{6}{100}, \qquad \sigma(Q') > 60\, \sigma(Q_\varepsilon), \qquad \frac{I_2(d_\varepsilon)}{\sigma(Q_\varepsilon)} \geq \frac{7}{100\, \varepsilon(\log 3 + \varepsilon/4)} \xrightarrow[\varepsilon \downarrow 0]{} \infty.$$

Proof. The fits (3.5) show $\mathcal{C}_2(d_\varepsilon) \neq \varnothing$. Since $R_\varepsilon = R_0 + \varepsilon S$,

$$\det R_\varepsilon = \frac{16 - \varepsilon}{6} > 0, \qquad (R_\varepsilon)_{12} + (R_\varepsilon)_{21} = \frac{10}{3} + \frac{5\varepsilon}{8},$$

so Theorems 4.1, 4.2 and 4.3 apply: $I_\infty = I_4$, the reduced infimum is attained, and (4.3) holds. The inequalities $I_4 \le I_3 \le \sigma(Q_\varepsilon)$ hold because $\mathcal{C}_3 \subset \mathcal{C}_4$ and $Q_\varepsilon \in \mathcal{C}_3(d_\varepsilon)$. Direct substitution of R_ε and the fixed L into the quantities of Theorem 4.3 gives

$$\Delta m = \frac{32(15-16\varepsilon)}{5(80+3\varepsilon)(80+27\varepsilon)}, \qquad a = \frac{1}{60}\Big(5 - \frac{3}{16-\varepsilon}\Big), \qquad b = \frac{19}{1200(16-\varepsilon)}.$$

On $[0, 10^{-3}]$, Δm and a decrease and b increases. Hence $\Delta m \ge \Delta m(10^{-3}) = 95897600/6402400081 > 149/10^4$, $a \le a(0) = 77/960$ and $b \le b(10^{-3}) = 95/95994 < 10^{-3}$. Since $77/960 + 2\cdot 10^{-3/2} < 1435/10^4$, (4.3) gives

$$I_2(d_\varepsilon) - I_\infty(d_\varepsilon) \ge \frac{2\cdot\frac{10}{3}\,(149/10^4)^2}{(1435/10^4)^2} = \frac{88804}{1235535} > \frac{7}{100}.$$

By (3.2) and $\log(1+x) \le x$, $\sigma(Q_\varepsilon) \le \varepsilon \log 3 + \varepsilon^2/4$, which is below $11/10^4$ because $\log 3 < 1.0987$. As $I_\infty \ge 0$, $I_2(d_\varepsilon) \ge 7/100$, which exceeds $\sigma(Q_\varepsilon) + 6/100$ and $60\,\sigma(Q_\varepsilon)$, and the displayed ratio bound follows. Every $Q' \in \mathcal{C}_2(d_\varepsilon)$ has $\sigma(Q') \ge I_2(d_\varepsilon)$. □

Thus restricting the hidden dimension from three to two does more than lose the lower-bound property: at the observed data of this chain, the minimal reduced fit reports more than sixty times the entropy production of the chain that generated the data. Theorem 3.1 concerns this explicit family; it is not a claim that $I_2(d(Q)) > \sigma(Q)$ for every true chain Q with unequal cross means, since $\sigma(Q)$ may exceed $I_\infty(d(Q))$ by more than the reduction bias. The constant $7/100$ is a lower bound, not the value of $I_2(d_\varepsilon)$; the explicit fits (3.5) have entropy production approximately 2.06 at $\varepsilon = 10^{-3}$, which bounds $I_2(d_\varepsilon)$ from above. Theorem 3.2 below raises the lower bound to $7/5$, using a computer-assisted step in exact arithmetic.

3.4. **A sharper bound by direct parametrization.** Theorem 4.3 uses only one Laplace weight of the cross-excursion durations. A direct description of all two-hidden-state fits gives a much larger bound.

Theorem 3.2 (Sharper bound, computer-assisted)**.** *For every $0 < \varepsilon \le 10^{-3}$, every $Q' \in \mathcal{C}_2(d_\varepsilon)$ satisfies $\sigma(Q') \ge 7/5$. Hence*

$$I_2(d_\varepsilon) \ge \frac{7}{5} > I_\infty(d_\varepsilon) + \frac{139}{100}, \qquad \sigma(Q') > 1200\,\sigma(Q_\varepsilon).$$

Proof. Parametrization. Let $Q' \in \mathcal{C}_2(d_\varepsilon)$. A single hidden state gives a rank-one R by (7.1), and $\det R_\varepsilon \ne 0$, so Q' has exactly two hidden states. Take A, C, K as in (7.1), let $D_h = \operatorname{diag}(d_1, d_2)$ hold the hidden masses, and put $X = K^{-1}C$. The entrance, exit and killed flow matrices of Q' are $U = A$, $V = D_h C$ and $J = D_h K$. Zero row sums give $K\mathbf{1} = C\mathbf{1}$, so $X\mathbf{1} = \mathbf{1}$. The entry X_{gj} is the probability that an excursion started in g exits to j, so $X \ge 0$, and $R_\varepsilon = AX$ makes X invertible. Relabel the hidden states so that X has rows $(x, 1-x)$ and $(y, 1-y)$ with $x > y$, and put $\delta = x - y$. By (7.1), $U = R_\varepsilon X^{-1}$ and $AK^{-1} = LX^{-1}$. Stationarity of the hidden masses is $(d_1, d_2)K = \mathbf{1}^T A$, hence $(d_1, d_2) = \mathbf{1}^T L X^{-1}$. Also $K^{-1} = A^{-1}LX^{-1} = XR_\varepsilon^{-1}LX^{-1}$, so

$$V = D_h X L^{-1} R_\varepsilon, \qquad J = D_h X L^{-1} R_\varepsilon X^{-1}.$$

With $Y = LX^{-1}D_h^{-1}$ these are the formulas (3.5); the fits displayed there have $x = 2/3$, $y = 1/3$. Substituting R_ε and L gives $d_1 = (43 - 108y)/(240\delta)$, $d_2 = (108x - 43)/(240\delta)$,

$$U = \frac{1}{48\delta}\begin{pmatrix} 112 - 192y - \varepsilon(36y - 9) & 192x - 112 + \varepsilon(36x - 9) \\ 80 - 192y - \varepsilon(12y - 3) & 192x - 80 + \varepsilon(12x - 3) \end{pmatrix},$$

$V_{1j} = (43 - 108y)\varphi_j(x)/(912\delta)$ and $V_{2j} = (108x - 43)\varphi_j(y)/(912\delta)$, where

$$\varphi_1(t) = 336t - 100 + \varepsilon(48t - 17), \qquad \varphi_2(t) = 3\big(24 - 32t + \varepsilon(48t - 17)\big).$$

The hidden internal flows are $-J_{12} = (108y - 43)\,\psi(x)/(912\delta^2)$ and $-J_{21} = (108x - 43)\,\psi(y)/(912\delta^2)$, where $\psi(t) = 240t^2 - 364t + 100 + \varepsilon(192t^2 - 116t + 17)$.

Location. Positive hidden masses give $y < 43/108 < x$. The condition $U_{12} \ge 0$ gives $x \ge (112 + 9\varepsilon)/(192 + 36\varepsilon)$. This bound decreases in ε, so $x \ge 583/1000$. Since $43 - 108y > 0$, $V_{12} \ge 0$ gives $x \le (24 - 17\varepsilon)/(32 - 48\varepsilon)$, which increases in ε, so $x \le 752/1000$. Since $108x - 43 > 0$, $V_{21} \ge 0$ gives $y \ge (100 + 17\varepsilon)/(336 + 48\varepsilon) \ge 25/84 > 2976/10^4$. Finally $-J_{21} \ge 0$ gives $\psi(y) \ge 0$. Now ψ is convex in t and affine in ε, and $\psi(361/1000)$ and $\psi(43/108)$ are negative for $\varepsilon = 0$ and for $\varepsilon = 10^{-3}$. Hence $\psi < 0$ on $[361/1000, 43/108]$, and $y < 361/1000$. Every fit therefore has (x, y) in the rectangle $\mathfrak{B} = [583/1000, 752/1000] \times [2976/10^4, 361/1000]$, on which $\delta > 1/5$.

Entropy bound. Drop the nonnegative direct term. For a reciprocal pair with flows $a, b > 0$, $\log r \ge 2(r-1)/(r+1)$ for $r \ge 1$ gives $(a-b)\log(a/b) \ge g(a,b) := 2(a-b)^2/(a+b)$. An absent pair has $a = b = 0$

by reciprocal support and contributes zero; set $g(0,0)=0$. Hence $\sigma(Q')\geq G$, where G is the sum of g over the five hidden-incident pairs (U_{11},V_{11}), (U_{12},V_{21}), (U_{21},V_{12}), (U_{22},V_{22}) and $(-J_{12},-J_{21})$.

Finite check. On $\{a\geq b\geq 0\}$ the function g increases in a and decreases in b, and symmetrically on $\{b\geq a\geq 0\}$. Hence, if $a\in[a_-,a_+]$ and $b\in[b_-,b_+]$ with $a_->b_+\geq 0$, then $g(a,b)\geq g(a_-,b_+)$, and similarly with the roles exchanged. The rectangle $\mathfrak{B}$ was bisected adaptively into 2824 subrectangles. On each subrectangle, times $\varepsilon\in[0,10^{-3}]$, every closed-form flow above was enclosed by interval arithmetic with exact rational endpoints. Negative lower endpoints were replaced by 0, because all flows of a fit are nonnegative. This gave a lower bound at least $7/5$ for G on every subrectangle. This step is computer-assisted. It uses only exact rational arithmetic and is performed by the script `anc/certify_I2.py` supplied with the source files. The script also re-derives the closed forms symbolically. Hence $\sigma(Q')\geq 7/5$.

The remaining claims follow from $I_\infty(d_\varepsilon)\leq\sigma(Q_\varepsilon)<11/10^4$ (Theorem 3.1): $7/5-11/10^4>139/100$ and $1200\cdot 11/10^4<7/5$. □

Numerical minimization over $\mathfrak{B}$ suggests $I_2(d_\varepsilon)\approx 1.52$, attained on the boundary $U_{12}=V_{21}=0$. This value is not proved and is not used.

4. Main theorems for positive determinant

The next three statements isolate the positive-determinant case. Theorem 9.1 later removes that determinant restriction, and Theorem 10.1 and Corollary 10.3 give the complete positive-data attainment and feasible two-hidden-state safety classification. The simpler case is treated first because it supplies the quantitative bias estimate.

Theorem 4.1 (Unrestricted first-moment infimum)**.** *Under these assumptions,*

$$I_\infty(d)=I_4(d)=B(d). \tag{4.1}$$

In particular, the conditions themselves guarantee a nonempty four-hidden-state class. One can approach the infimum while matching the same fixed P,F,R,L *exactly, using four hidden states, no hidden-to-hidden edges, and a uniform upper bound on all transition rates. The construction permits some rates to tend to zero. Four is a sufficient state budget, not a claimed minimum. If the two cross-excursion means differ, no finite compatible chain attains this unrestricted infimum.*

Theorem 4.2 (Complete two-hidden-state safety dichotomy)**.** *Assume additionally $\mathcal{C}_2(d)\neq\varnothing$. The reduced infimum is attained, and*

$$I_2(d)=I_\infty(d)\quad\Longleftrightarrow\quad\frac{L_{12}}{R_{12}}=\frac{L_{21}}{R_{21}}. \tag{4.2}$$

Thus equality of the two conditional cross-excursion means is necessary and sufficient for safety of the infimum *within this regime. It is not a test that every individual fit has the same dissipation, nor a test identifying the true hidden network. Feasibility of the two-hidden-state class is an explicit hypothesis, not inferred from the single equality in* (4.2)*.*

Theorem 4.3 (Quantitative reduction bias)**.** *Define*

$$\Delta m=\frac{L_{12}}{R_{12}}-\frac{L_{21}}{R_{21}},\qquad a=\operatorname{tr}(R^{-1}L),\qquad b=\frac{\det L}{\det R}.$$

Under the hypotheses of Theorem 4.2, $a,b>0$ and

$$I_2(d)-I_\infty(d)\geq\frac{2(R_{12}+R_{21})(\Delta m)^2}{(a+2\sqrt{b})^2}. \tag{4.3}$$

For $\Delta m\neq 0$ the gap is strictly positive and applies to every exact two-hidden-state fit. This bound is not asserted to be the optimal gap constant.

5. A shared entropy lower bound

For a specified microscopic hidden excursion, its stationary occurrence-rate density divided by the density of its reversed excursion is the product of stationary flow ratios along its jumps. Holding-time factors cancel on reversal and stationary masses telescope. Summing the logarithmic ratio over all excursion paths recovers entropy production on edges incident with the hidden set. Direct visible jumps supply the first term of (2.1). For each fixed finite irreducible chain these sums are integrable: hidden exit times have finite moments and there are finitely many nonzero rate logarithms.

Log-sum after forgetting hidden paths and durations yields $\sigma(Q) \geq B(d)$. This is the standard path-coarse-graining bound; see [15, 17, 20].

The same $B(d)$ is the informed partial entropy-production rate of the marginal-observer framework [3, 16]. Deleting only the direct visible pair gives a stalled stationary ratio $p_2^{\mathrm{st}}/p_1^{\mathrm{st}} = R_{12}P_2/(R_{21}P_1)$ by hidden-block elimination. With $J = F_{12} - F_{21} = R_{21} - R_{12}$, its rate is $J\log(F_{12}R_{21}/(F_{21}R_{12})) = B(d)$. The excursion data determine this ratio without requiring an experimental stalling intervention.

Keeping the duration gives a useful stronger version. Let $p = f_{12}/R_{12}$ and $q = f_{21}/R_{21}$ be the conditional duration densities. Then

$$\sigma(Q) - B(d) \geq R_{12}D(p\|q) + R_{21}D(q\|p). \tag{5.1}$$

For this observation rule, the added duration term is the cross-excursion waiting-time divergence of [14, Eq. (11)]; optimization over compatible state budgets is a separate question. Indeed the cross-excursion term decomposes exactly into the count term and the two weighted relative entropies. Contributions from return-path irreversibility, which were discarded, are nonnegative. This does not require that fitted and true models have the same full path law.

6. Exact first moments with arbitrarily little extra dissipation

We first record an elementary channel construction. Suppose $M = xy^T$ is a nonnegative nonzero rank-one 2×2 matrix, $y^T\mathbf{1} = 1$, and $x_i > 0$ iff $y_i > 0$. Write $s = \mathbf{1}^T x$. Introduce one hidden state with mean residence time $t > 0$, stationary mass st, entrance flows $n_{ig} = x_i$, and exit flows $n_{gi} = sy_i$. Its excursion matrices are exactly M, tM. Its hidden-incident entropy contribution is

$$(M_{12} - M_{21})\log(M_{12}/M_{21}), \tag{6.1}$$

when the cross entries are positive, and zero for a diagonal-only channel. This follows since $x_1 - sy_1 = M_{12} - M_{21}$ and the current at visible state 2 is its negative. A collection of these channels fits the data whenever its occurrence and moment matrices sum to R, L: hidden balance holds separately, the consistency identities supply visible balance, and the total hidden mass is $\mathbf{1}^T L\mathbf{1}$.

Choose once and for all

$$0 < \delta < \min_{i,j} L_{ij}/R_{ij}.$$

For a small positive rate ε with $\delta\varepsilon < 1$, define

$$R_f(\varepsilon) = \frac{R - \varepsilon L}{1 - \delta\varepsilon}, \qquad S(\varepsilon) = \frac{L - \delta R}{1 - \delta\varepsilon}. \tag{6.2}$$

For all sufficiently small ε, R_f is entrywise positive with positive determinant and S is entrywise positive. The exact identities are

$$R_f + \varepsilon S = R, \qquad \delta R_f + S = L. \tag{6.3}$$

Write $R_f = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$. Its following two-channel factorisation is rational whenever its entries are rational:

$$M_{f,1} = \begin{pmatrix} A & B \\ C & BC/A \end{pmatrix}, \qquad M_{f,2} = \begin{pmatrix} 0 & 0 \\ 0 & D - BC/A \end{pmatrix}. \tag{6.4}$$

The first matrix has reciprocal full support and the second only a visible-2 return. Give both channels residence time δ. Their entropy is exactly $(B - C)\log(B/C)$.

The positive matrix S has a two-channel factorisation with full reciprocal support. To see this, select a fixed $\eta \in (0, 1/2)$ less than every entry of each row of $L - \delta R$ divided by that row sum. Put

$$Y = \begin{pmatrix} 1-\eta & \eta \\ \eta & 1-\eta \end{pmatrix}, \qquad X = SY^{-1}.$$

Every entry of X is positive, since its row entries are $(S_{ij} - \eta\sum_j S_{ij})/(1 - 2\eta)$. If $S_g = X_{\cdot g}Y_{g\cdot}$, then $S = S_1 + S_2$ and both S_g are positive rank one. Give occurrence matrices $\varepsilon S_1, \varepsilon S_2$ residence time $1/\varepsilon$.

The four channels fit (6.3) exactly. Each connects to at least one visible state; the direct visible edge and the positive first channel ensure irreducibility. Every nonzero transition has its reverse. All stationary masses are positive and sum to one. The hidden-incident entropy is

$$(R_{f,12} - R_{f,21}) \log(R_{f,12}/R_{f,21}) + \varepsilon \sum_{g=1}^{2} (S_{g,12} - S_{g,21}) \log(S_{g,12}/S_{g,21}). \tag{6.5}$$

The first term tends to the cross-count term of (2.1); the second tends to zero. The channel matrices are continuous and have the required positive limits, so the excess is $O(\varepsilon)$. The direct term is fixed. The lower bound in Section 5 proves (4.1).

The fast departure rates are $1/\delta$, the slow departure rates are ε, and visible departure rates are fixed by F, P. Hence all rates are bounded above uniformly. The slow stationary masses remain positive while their entrances and exits vanish. The endpoint is reducible and its excursion first moments are not obtained by naively substituting $\varepsilon = 0$. No invalid interchange of an unbounded duration moment and a parameter limit is made.

7. Nonsingular two-hidden-state fits cannot escape to infinity

For any nonempty hidden set, irreducibility implies eventual exit: a supported path to the visible set gives a common positive exit probability in a bounded time from every hidden state. Iteration makes the killed survival tail exponential. Thus its negative killed generator is invertible, and integrating the killed semigroup gives the excursion identities below. In particular, $f(t) = Ae^{-Kt}C$, $\int_0^\infty e^{-Kt}dt = K^{-1}$ and $\int_0^\infty te^{-Kt}dt = K^{-2}$. These are the standard phase-type resolvent and moment identities [22, Section 2], here with stationary entrance flows and endpoint-specific exits. The square-factorization consequence is a continuous-time relative of the reconstruction principle in [6].

For a two-hidden-state model, let

$$A = \operatorname{diag}(P_1, P_2)Q_{VH}, \quad C = Q_{HV}, \quad K = -Q_{HH}.$$

The hidden block is transient, K is invertible, and

$$R = AK^{-1}C, \qquad L = AK^{-2}C. \tag{7.1}$$

These are the marked phase-type resolvent and moment identities, with stationary entrance rates in A rather than a normalized initial law [13, Section 3.1, Eqs. (17), (19)]. The nonsingular square-matrix consequences needed here follow directly below.

Nonsingularity of R forces A, C invertible and consequently L invertible. Straight multiplication gives

$$R(R + sL)^{-1}R = A(K + sI)^{-1}C, \tag{7.2}$$

and

$$\operatorname{tr} K = \operatorname{tr}(RL^{-1}) =: \tau, \qquad \det K = \det R / \det L =: \Delta. \tag{7.3}$$

Both τ, Δ are positive. The two hidden eigenvalues are real and positive: the discriminant is $(K_{11} - K_{22})^2 + 4q_{g_1g_2}q_{g_2g_1} \geq 0$, and trace and determinant are positive.

Proposition 7.1 (Attainment at fixed nonsingular data)**.** *If R is nonsingular and a reciprocal irreducible two-hidden-state fit exists with positive visible probabilities and positive direct visible flows, its entropy-production infimum is attained. This proposition does not require* $\det R > 0$.

Proofs of Proposition 7.1 and the sufficient direction of Theorem 4.2 are given in Appendix B.

8. A computable strict gap from unequal cross means

Apply (5.1) and the randomized map that, conditional on a duration t, outputs 1 with probability e^{-st}. Log-sum gives the Bernoulli relative-entropy lower bounds. For fixed $q \in (0, 1)$, the second derivative of binary relative entropy in its first argument is $1/[p(1-p)] \geq 4$, and its value and first derivative vanish at $p = q$. Hence each relative entropy is at least twice the squared difference of the corresponding Laplace expectations. Thus

$$\sigma(Q) - B(d) \geq 2(R_{12} + R_{21}) \left(\frac{\mathcal{K}_{12}(s)}{R_{12}} - \frac{\mathcal{K}_{21}(s)}{R_{21}} \right)^2. \tag{8.1}$$

For a two-hidden-state fit, (7.2) fixes $\mathcal{K}(s)$. The 2×2 adjugate is linear in the entries; therefore

$$\mathcal{K}(s) = R - s \, \frac{(\det R)L + s(\det L)R}{\det(R + sL)}.$$

Dividing its cross entries by their occurrence rates gives exactly

$$\frac{\mathcal{K}_{12}(s)}{R_{12}} - \frac{\mathcal{K}_{21}(s)}{R_{21}} = -\frac{s\Delta m}{1 + as + bs^2}. \tag{8.2}$$

Here $a = \operatorname{tr} K^{-1} > 0$ and $b = 1/\det K > 0$, by (7.1). The derivative of $s/(1 + as + bs^2)$ has sign $1 - bs^2$; its maximum over $s > 0$ is $(a + 2\sqrt{b})^{-1}$ at $s = b^{-1/2}$. Equations (8.1), (8.2) and Theorem 4.1 prove Theorem 4.3 and the necessary direction of Theorem 4.2.

9. The exact count optimum and all determinant regimes

We now remove the positive-determinant restriction from the unrestricted infimum. Throughout this section all entries of R, L, the three observed masses, and the direct visible flows are positive and satisfy the consistency identities of Section 2. Put

$$R = \begin{pmatrix} a & x \\ y & b \end{pmatrix}, \quad s = x + y, \quad j = x - y, \quad h = \min\{\sqrt{ab}, \sqrt{xy}\}, \quad \vartheta = \sqrt{1 - 4h^2/s^2}. \tag{9.1}$$

All square roots are real, $0 \le \vartheta < 1$, and $\vartheta \ge |j|/s$. Define, with its continuous value at $\vartheta = 0$,

$$\Psi(R) = 2s\vartheta \operatorname{artanh} \vartheta, \qquad \mathcal{B}(d) = (F_{12} - F_{21}) \log(F_{12}/F_{21}) + \Psi(R). \tag{9.2}$$

If $\det R \ge 0$, then $\vartheta = |j|/s$ and $\mathcal{B} = B$. If $\det R < 0$, then $\mathcal{B} > B$. These expressions are continuous, including at $\det R = 0$ and $x = y$.

Theorem 9.1 (Exact count optimum and fixed-first-moment infimum)**.** *Every finite compatible model satisfies $\sigma(Q) \ge \mathcal{B}(d)$. If only P, F, R are prescribed, the minimum is $\mathcal{B}(d)$ and is attained with at most two hidden states. If the full positive data $d = (P, F, R, L)$ are prescribed, then*

$$\boxed{I_\infty(d) = I_4(d) = \mathcal{B}(d).} \tag{9.3}$$

The latter infimum can be approached with the same data held exactly fixed, no hidden-to-hidden edges, and a uniform upper bound on transition rates. This statement includes positive, zero and negative determinants. Four is a sufficient budget; it is not claimed to be the smallest universal budget.

The geometric matrices J, Z, stationary path weights and Gram bound used below are defined in Appendix A, in (A.2), (A.4) and (A.5)–(A.6).

9.1. A scalar inequality and its equality condition.

For $0 < v \le 1$ let

$$f(v) = 2\sqrt{1 - v^2} \operatorname{artanh} \sqrt{1 - v^2}, \qquad f(1) = 0.$$

This function is strictly decreasing and strictly convex. In fact, writing $u = \sqrt{1 - v^2}$,

$$f'(v) = -2v \operatorname{artanh}(u)/u - 2/v < 0, \tag{9.4}$$

$$f''(v) = \frac{2}{u^3}\left(\frac{u}{1 - u^2} - \operatorname{artanh} u\right) > 0 \quad (0 < v < 1). \tag{9.5}$$

The last bracket has derivative $2u^2/(1 - u^2)^2 > 0$ and vanishes at zero. The limiting derivatives at $v = 1$ preserve strict convexity.

For each microscopic excursion sequence $\gamma : 1 \to H \to 2$, let $A_\gamma = w(\gamma) > 0$ and $B_\gamma = w(\bar{\gamma}) > 0$, with w as in (A.4). Their total rates are x, y. Set

$$p_\gamma = \frac{A_\gamma + B_\gamma}{s}, \quad u_\gamma = \frac{A_\gamma - B_\gamma}{A_\gamma + B_\gamma}, \quad v_\gamma = \sqrt{1 - u_\gamma^2}.$$

Then $\sum p_\gamma = 1$, $\sum p_\gamma u_\gamma = j/s$ and

$$\sum_\gamma (A_\gamma - B_\gamma) \log \frac{A_\gamma}{B_\gamma} = s \sum_\gamma p_\gamma f(v_\gamma) \ge s f\left(\frac{2G_{12}}{s}\right). \tag{9.6}$$

Jensen holds on this countable probability space because the left side is integrable for every fixed finite chain. The Gram bound (A.6) and Cauchy–Schwarz on cross paths imply $G_{12} \le h$. Since f decreases, (9.6) is at least $sf(2h/s) = \Psi(R)$. Return-path entropy is nonnegative and direct jumps give the remaining term. This proves the universal lower bound.

For $\det R < 0$, equality forces, separately, zero return-path entropy, $G_{12} = \sqrt{ab}$, and $|u_\gamma| = \vartheta$ for every supported cross path. This follows from strict decrease and strict Jensen equality. Both signs occur, because $\vartheta > |j|/s$. If $\det R \ge 0$, equality with B instead forces the familiar log-sum equality: every cross-path ratio equals x/y, and every return path has the same rate as its reverse.

9.2. **Constructing the two optimal channels.** For $\det R < 0$ define positive weights

$$w_\pm = \frac{s \pm j/\vartheta}{2} \tag{9.7}$$

and matrices

$$M_\pm = w_\pm \begin{pmatrix} a/s & (1 \pm \vartheta)/2 \\ (1 \mp \vartheta)/2 & b/s \end{pmatrix}. \tag{9.8}$$

They are strictly positive, sum to R, and have determinant zero since $(1 - \vartheta^2)/4 = ab/s^2$. Their cross affinities are respectively $\pm 2 \operatorname{artanh} \vartheta$. Their combined entropy contribution is exactly $\Psi(R)$ by (6.1). The opposite affinities are required for this optimum; they are not inferred from a nonzero net current, which can vanish.

For $\det R > 0$, use the two matrices (6.4) with $R_f = R$; the second is a diagonal-only channel. For $\det R = 0$, R itself is one positive rank-one channel. Their cost is the cross term in B. To fit arbitrary positive P, F consistent with R, give every channel the common residence time $P_H/(\mathbf{1}^T R \mathbf{1})$ and use the channel construction of Section 6. This proves exact attainment for the count-only problem, with at most two hidden states.

The symmetric specialization is established prior work: for $R_{11} = R_{22} = 1-p$ and $R_{12} = R_{21} = p > 1/2$, the result reduces to the two-channel formula in Skinner–Dunkel [20, Author manuscript, Appendix Eqs. (A28)–(A29)]. Indeed, if $u = \sqrt{2p-1}$, then $\vartheta = 2u/(1+u^2)$ and $\Psi(R) = 4u \log((1+u)/(1-u))$. We do not claim that special formula or its counterflow mechanism as new. The proof above treats arbitrary positive asymmetric R without assuming a symmetric optimizer, and the following first-moment and attainment results concern the specified fixed-data inverse problem.

9.3. **Restoring arbitrary first moments exactly.** Choose $0 < \delta < \min_{ij} L_{ij}/R_{ij}$ and define R_f, S by (6.2). Both are entrywise positive for small positive ε, whether or not either determinant is positive. Realize R_f by the at-most-two optimal channels just constructed, with residence time δ. Realize the positive S by the two positive rank-one factors S_1, S_2 constructed below (6.4), with occurrence matrices εS_g and residence time $1/\varepsilon$. The identities (6.3) prove exact agreement with R, L, and consistency recovers P, F. Zero channels at determinant zero are omitted.

The entropy equals the fixed direct term plus

$$\Psi(R_f(\varepsilon)) + \varepsilon \sum_{g=1}^{2} (S_{g,12} - S_{g,21}) \log(S_{g,12}/S_{g,21}).$$

Continuity of Ψ and the positive limiting slow factors give convergence to $\mathcal{B}(d)$. Every fast total exit rate is $1/\delta$, every slow total exit rate is ε, and visible departure rates are fixed by the observations. All masses and nonzero rates are finite and positive at every positive parameter. This proves (9.3). It does not replace the endpoint by a reducible chain or exchange a singular limit with a duration moment.

10. When is the unrestricted infimum attained?

The preceding theorem gives an infimum for every strictly positive consistent first-moment data set. Finite attainment has a more restrictive, exactly describable boundary.

Theorem 10.1 (Finite-attainment classification)**.** *Under the positivity and consistency assumptions of Theorem 9.1, some finite compatible model attains $\mathcal{B}(d)$ if and only if the condition in its determinant regime holds:*

$$\det R > 0: \quad L_{12}/R_{12} = L_{21}/R_{21} \text{ and } \det L > 0; \tag{10.1}$$

$$\det R = 0: \quad L = tR \quad \text{for some } t > 0; \tag{10.2}$$

$$\det R < 0: \quad L = t_+ M_+ + t_- M_- \quad \text{for some } t_+, t_- > 0. \tag{10.3}$$

The respective sufficient hidden-state budgets are four, one and two. For negative determinant, the last condition is equivalently

$$\frac{L_{11}}{a} = \frac{L_{22}}{b} = \frac{L_{12} + L_{21}}{s}, \qquad |L_{12} - L_{21}| < \vartheta(L_{12} + L_{21}). \tag{10.4}$$

Failure means nonattainment in every finite hidden dimension, not failure of the unrestricted infimum formula.

Proof. We use a Laplace-weighted version of the geometric matrix. With $\Pi = \operatorname{diag}(\pi_h)$, set

$$G(z) = Z(J + z\Pi)^{-1}Z^T, \qquad z \geq 0. \tag{10.5}$$

The path expansion multiplies each visit to h in (A.4) by $d_h/(d_h + z\pi_h)$. This is the conditional Laplace transform of its exponential holding time, and its product is unchanged by path reversal. Consequently $G(z)$ is precisely the sum of geometric forward/reverse path rates with that common duration weight. In particular, when every return sequence has equal forward/reverse rate, $\mathcal{K}_{ii}(z) = G_{ii}(z)$.

Suppose first $\det R \geq 0$ and equality is attained. Every cross-path rate ratio is $c = x/y$. Thus for all $z \geq 0$,

$$D_c^{-1/2}\mathcal{K}(z)D_c^{1/2} = G(z), \qquad D_c = \operatorname{diag}(c, 1). \tag{10.6}$$

If $\det R > 0$, then $G(0)$ is positive definite, so Z has row rank two. Therefore

$$-G'(0) = ZJ^{-1}\Pi J^{-1}Z^T$$

is positive definite as well. Equation (10.6) makes $D_c^{-1/2}LD_c^{1/2}$ symmetric positive definite. This is exactly the equal-cross-mean condition and $\det L > 0$, since the diagonal entries of L are positive.

Conversely assume (10.1). Both symmetrized matrices R, L are entrywise positive and positive definite. Choose $\delta > 0$ so small that $L - \delta R$ has both properties, and then choose $\varepsilon > 0$ so small that $R - \varepsilon L$ does too. The matrices R_f, S in (6.2) consequently have positive determinant and the same cross ratio c. Factor each using (6.4); its non-diagonal factor has ratio c and its other factor is diagonal-only. Give the fast factors time δ and the slow factors time $1/\varepsilon$, with slow occurrences multiplied by ε. Every cross channel now has ratio c, not merely in a limit. The four-hidden-state realization matches d and attains $B(d)$ exactly.

If $\det R = 0$, (10.6) implies that $G(0)$ has rank one. Since J is positive definite, Z has rank one. Formula (10.5) then has the form $G(z) = g(z)uu^T$, so (10.6) implies $\mathcal{K}(z) = a(z)R$ with $a(0) = 1$. Differentiation yields $L = tR$, with $t > 0$ because all hidden durations are positive. Conversely the single channel R with time t gives a saturating realization.

Finally suppose $\det R < 0$. Equality in (9.6) forces $u_\gamma = \pm\vartheta$, return-path reversibility, and $G_{12}(0) = \sqrt{ab}$. Since $G_{ii}(0) = R_{ii}$, $G(0)$ has rank one and hence Z has rank one. Therefore

$$\frac{\mathcal{K}_{11}(z)}{a} = \frac{\mathcal{K}_{22}(z)}{b} = \frac{G_{12}(z)}{\sqrt{ab}}.$$

On each cross path, $\sqrt{A_\gamma B_\gamma} = (\sqrt{ab}/s)(A_\gamma + B_\gamma)$. The common duration weight gives

$$G_{12}(z) = \frac{\sqrt{ab}}{s}\big(\mathcal{K}_{12}(z) + \mathcal{K}_{21}(z)\big).$$

All conditional first duration moments are finite for a fixed finite chain, so the nonnegative path sums may be differentiated at zero by monotone convergence. Differentiating proves the two equalities in (10.4). Let $\ell_\pm$ be the sums of $(A_\gamma + B_\gamma)\mathbb{E}[T \mid \gamma]$ over paths with sign $\pm\vartheta$. Both are strictly positive: both sign classes have positive occurrence rate and every supported duration has positive mean. Then

$$\ell_+ + \ell_- = L_{12} + L_{21}, \qquad \vartheta(\ell_+ - \ell_-) = L_{12} - L_{21}.$$

This proves the strict inequality. Conversely set

$$\ell_\pm = \frac{L_{12} + L_{21} \pm (L_{12} - L_{21})/\vartheta}{2}, \qquad t_\pm = \ell_\pm/w_\pm > 0. \tag{10.7}$$

The two equalities in (10.4) and the sum/difference formulas give $L = t_+M_+ + t_-M_-$ entry by entry. The two channels of (9.8) with these times provide a finite attaining model. This also proves equivalence of the two formulations. □

10.1. **The singular two-hidden-state boundary.** The nonsingular compactness theorem cannot be extended to rank-one data. The exact alternative is as follows.

Theorem 10.2 (Rank-one feasibility and safe nonattainment)**.** *Let R be strictly positive with $\det R = 0$, and keep the other positive consistency hypotheses. A model with at most two hidden states exists if and only if*

$$\operatorname{rank}[R\ L] = 1 \quad \textit{or} \quad \operatorname{rank}\begin{pmatrix} R \\ L \end{pmatrix} = 1. \tag{10.8}$$

Whenever this condition holds, $I_2(d) = I_\infty(d) = B(d)$. The infimum is attained if and only if $L = tR$ for some $t > 0$, in which case one hidden state suffices. Thus two-hidden-state reduction is safe throughout its feasible positive rank-one slice, even when the cross means differ; it can be safe without possessing a minimizer.

Proof. For an actual two-hidden-state fit, $R = AK^{-1}C$ with K invertible. Since $\det R = 0$, A or C has rank at most one, and neither has rank zero. In the first case the columns of both R, L lie in the one-dimensional image of A; in the second their rows lie in the row space of C. A one-hidden-state fit has both properties. This proves necessity of (10.8).

Conversely choose δ, ε as in (6.2). The positive R_f, S inherit the common rank-one column or row factor. They are therefore themselves positive rank-one channels. Give them occurrence matrices $R_f, \varepsilon S$ and times $\delta, 1/\varepsilon$. This is an exact two-hidden-state fit with entropy tending to $B(d)$. Theorem 10.1 supplies the precise attainment criterion, in every finite dimension. □

Corollary 10.3 (Safety in all feasible two-hidden-state regimes)**.** *Suppose a model with at most two hidden states fits the positive data. Safety, meaning $I_2(d) = I_\infty(d)$, holds precisely as follows:*

Occurrence matrix	*Necessary and sufficient safety condition*
$\det R > 0$	*Equal conditional cross-excursion means.*
$\det R = 0$	*Always safe; attainment requires $L = tR$.*
$\det R < 0$	*The two equalities and strict inequality in* (10.4)*.*

Proof. The positive case is Theorem 4.2 and the zero case is Theorem 10.2. For negative determinant, Proposition 7.1 makes the two-hidden-state infimum attained. Equality of infima is therefore equivalent to finite attainment of $\mathcal{B}(d)$; Theorem 10.1 both characterizes it and gives a two-hidden-state witness. □

10.2. **Examples separating safety, attainment and feasibility.** For the negative-determinant example of Appendix A, $s = 17/2$, $\vartheta = 15/17$, and $L = R/50$. Theorem 9.1 gives its exact optimum:

$$I_\infty(d) = I_2(d) = \frac{15}{2}\log 16. \tag{10.9}$$

This symmetric value reproduces the established two-channel count mechanism discussed above. It serves as an exact check of the general formula and of the first-moment attainment criterion.

A feasible but unsafe negative-determinant example uses

$$N_1 = \begin{pmatrix} 2 & 8 \\ 1/4 & 1 \end{pmatrix}, \quad N_2 = \begin{pmatrix} 1 & 1/4 \\ 4 & 1 \end{pmatrix}, \qquad t_1 = 1/100, \quad t_2 = 1/50.$$

Then

$$R = \begin{pmatrix} 3 & 33/4 \\ 17/4 & 2 \end{pmatrix}, \quad L = \begin{pmatrix} 1/25 & 17/200 \\ 33/400 & 3/100 \end{pmatrix},$$

with $P_H = 19/80$, $P_1 = P_2 = 61/160$, and $F_{12} = 1, F_{21} = 5$. The channels prove feasibility. Here $L_{11}/R_{11} = 1/75 \neq 3/200 = L_{22}/R_{22}$, so $I_2 > I_\infty = \mathcal{B}$. Here $\vartheta = 23/25$ and $I_\infty = 4\log 5 + (23/2)\log 24$. The positive reduction gap follows from attained reduced minimization; no numerical value for that gap is claimed.

For a singular safe but unattained example take

$$R = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad L = \frac{1}{100}\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \quad P = (47/100, 47/100, 3/50), \quad F_{12} = F_{21} = 1.$$

The common column space verifies (10.8); the cross means are $1/50$ and $1/100$. Consequently $I_2 = I_\infty = 0$, but every finite compatible model has strictly positive entropy production. This demonstrates why the nonsingular assumption in the earlier equal-mean safety theorem is essential.

11. A fixed reduction bias can be hidden at finite observation times

The singular channels approaching the first-moment infimum also obstruct uniform finite-time detection. Rare mean-changing contamination is classical [2, Theorem 1 and its proof]; here it is realized by stationary reciprocal chains at fixed prescribed data. A positive reduction bias and determinant, even with a fixed hidden-state budget, do not ensure uniform detection power from continuously observed stationary labels in a fixed time unit. Write

$$\mathfrak{b}(d) = I_2(d) - I_\infty(d)$$

only when the two-hidden-state fitting class is nonempty.

Proposition 11.1 (Fixed-bias finite-window non-testability)**.** *There are two fixed strictly positive consistent data $d^{\mathrm{S}}, d^{\mathrm{U}}$, both two-hidden-state feasible and sharing the same P, F, R, such that*

$$\det R = \frac{21}{4}, \qquad \mathfrak{b}(d^{\mathrm{S}}) = 0, \qquad \mathfrak{b}(d^{\mathrm{U}}) \geq g_* := \frac{5(17 - 12\sqrt{2})}{18} > 0.$$

For each $0 < \varepsilon \le 1$ there are reciprocal irreducible chains $Q^{\mathrm{S}}_\varepsilon, Q^{\mathrm{U}}_\varepsilon$, each with four hidden states, realizing their respective data exactly and having a common finite upper bound on transition rates. If $\mathsf{P}^{\ell}_{\varepsilon,T}$ is the law of their stationary observed label trajectory on $[0,T]$, then for finite constants c_0, c_1 independent of ε, T,

$$\left\|\mathsf{P}^{\mathrm{S}}_{\varepsilon,T} - \mathsf{P}^{\mathrm{U}}_{\varepsilon,T}\right\|_{\mathrm{TV}} \le c_0\varepsilon + c_1\varepsilon T. \tag{11.1}$$

In particular, for every fixed $T < \infty$ this distance tends to zero.

Fix $0 < \alpha < 1$ and a possibly randomized test $0 \le \phi_T \le 1$ declaring an unsafe reduction. If $\mathbb{E}_Q\phi_T \le \alpha$ for every chain with safe two-hidden-state-feasible data, then

$$\inf_{Q:\, d(Q)=d^{\mathrm{U}}} \mathbb{E}_Q\phi_T \le \alpha. \tag{11.2}$$

The assertion remains valid when the common exact P, F, R are supplied to the test.

Proof. Use the unsafe datum of Appendix B.4.1, and denote its moment matrix by L^{U}. Thus

$$R = \begin{pmatrix} 9/2 & 3 \\ 2 & 5/2 \end{pmatrix}, \qquad L^{\mathrm{U}} = \begin{pmatrix} 1/20 & 1/25 \\ 3/100 & 9/200 \end{pmatrix},$$
$$P = (2/5, 87/200, 33/200), \qquad F_{12} = 1, \quad F_{21} = 2.$$

Its two-channel witness and Theorem 4.3 give the stated lower bound g_*. Since $\mathbf{1}^T R\mathbf{1} = 12$, set

$$t_0 = \frac{P_H}{\mathbf{1}^T R\mathbf{1}} = \frac{11}{800}, \qquad L^{\mathrm{S}} = t_0 R.$$

The safe datum has a two-channel witness

$$N_1 = \begin{pmatrix} 9/2 & 3 \\ 2 & 4/3 \end{pmatrix}, \qquad N_2 = \begin{pmatrix} 0 & 0 \\ 0 & 7/6 \end{pmatrix},$$

both with residence time t_0. Their hidden entropy is $\log(3/2)$; the direct edge contributes $\log 2$. Hence the witness attains $B(d^{\mathrm{S}}) = \log 3$, and Theorem 4.1 proves safety.

For both data use the exact channel construction of Section 6, with $\delta = 1/1000$:

$$R^{\ell}_f = \frac{R - \varepsilon L^{\ell}}{1 - \delta\varepsilon}, \qquad S^{\ell} = \frac{L^{\ell} - \delta R}{1 - \delta\varepsilon}, \qquad \ell \in \{\mathrm{S}, \mathrm{U}\}. \tag{11.3}$$

These matrices are positive for $0 < \varepsilon \le 1$, and $\det R^{\ell}_f > 0$. In fact,

$$\det(R - \varepsilon L^{\mathrm{U}}) = \frac{21}{4}(1 - \varepsilon/100)(1 - \varepsilon/50), \qquad \det(R - \varepsilon L^{\mathrm{S}}) = \frac{21}{4}(1 - 11\varepsilon/800)^2.$$

Split R^{ℓ}_f into the two fast matrices in (6.4), both with residence time δ. For the slow factorization take

$$Y = \begin{pmatrix} 99/100 & 1/100 \\ 1/100 & 99/100 \end{pmatrix}, \qquad X^{\ell} = S^{\ell}Y^{-1}, \qquad S^{\ell}_j = X^{\ell}_{\cdot j}Y_{j\cdot} \quad (j = 1, 2).$$

Every entry of X^{ℓ} is positive: each row's smaller entry in $L^{\ell} - \delta R$ exceeds $1/100$ of that row's sum. Give occurrence matrix εS^{ℓ}_j residence time $1/\varepsilon$. The identities

$$R^{\ell}_f + \varepsilon S^{\ell} = R, \qquad \delta R^{\ell}_f + S^{\ell} = L^{\ell}$$

prove exact agreement with R, L. The channel construction and consistency supply the prescribed masses, flows and stationarity. All supported edges are reciprocal, and the direct visible edge makes the chain irreducible. Fast exit rates are 1000, slow exit rates are ε, and visible departure rates are fixed by P, F, R. Thus all transition rates have a common finite upper bound.

It remains to compare the stationary *observed* paths. The aggregate stationary slow mass is the same in both chains:

$$m_{\mathrm{slow}}(\varepsilon) = \mathbf{1}^T S^{\ell}\mathbf{1} = \frac{P_H - \delta\mathbf{1}^T R\mathbf{1}}{1 - \delta\varepsilon} = \frac{153/1000}{1 - \varepsilon/1000}. \tag{11.4}$$

The two fast-state masses and rates are smooth in ε, with common positive limits at zero; their differences are $O(\varepsilon)$. Slow entrance rates are $O(\varepsilon)$. Couple the stationary initial categories

$$1,\ 2,\ f_1,\ f_2,\ \{\text{slow states}\}.$$

Their probabilities differ by $O(\varepsilon)$, because visible and aggregate slow masses agree exactly. If the common initial category is slow, both observed labels stay hidden until a slow exit; since each residual slow exit time is exponential of rate ε, the two exit times can be coupled to coincide, and the probability of this exit by T is at most εT. The slow microscopic states need not agree. If the initial common category is visible or fast, couple each transition using the minimum of the two rates, declaring failure at an unmatched transition or any slow entrance. Before failure the states in this common subsystem agree,

and the failure intensity is at most a constant times ε. Uniform constants follow from the positive fast limits and bounded rational coefficients on $[0,1]$. Initial mismatch and these failure probabilities prove (11.1) by the coupling inequality. Finally the difference between the expectations of a $[0,1]$-valued test is at most total variation. Apply the uniform safe level bound and let $\varepsilon \downarrow 0$ to obtain (11.2). □

The reducible endpoint $\varepsilon = 0$ serves only to bound coefficients and compare limits. Pointwise eventual detection for a fixed chain remains possible; arbitrary growing horizons are not covered. Procedures stopping by T are covered, since their output is measurable from the path on $[0,T]$. The fixed bias compares fitting infima, not a fitted value with the actual generator's entropy.

12. Discussion

The first-moment classification makes the consequence of a hidden-state budget explicit. The unrestricted fitting floor is $\mathcal{B}(d)$; Corollary 10.3 decides whether a nonempty two-hidden-state class preserves it. Feasibility, equality of infima, finite attainment and identification of the microscopic dynamics are separate conclusions. In particular, rank-one data can be safe without a finite minimizer, and matching prescribed moments need not preserve the remaining observed law.

The obstruction in Proposition 11.1 holds at fixed safe and unsafe data, fixed determinant and positive unsafe bias, with four hidden states and a common upper rate bound. Every fixed finite chain has exponentially decaying excursion tails, but their decay rates tend to zero across the slow-exit family. Rare long excursions carry the distinguishing first moments.

Open questions: the minimal hidden-state budget for the unrestricted infimum (four is sufficient); exact unsafe two-hidden-state infima; a complete reduced-feasibility test beyond the cases treated here; and data with vanishing observed entries.

AI assistance disclosure

OpenAI GPT-5.6 Sol, accessed through ChatGPT and Codex, was used to develop and audit the proofs and exact computations, and for drafting. Anthropic Claude Opus 5.5, used through Claude Code, was used to condense and edit the paper, for an independent check of the sufficiency, regime and non-testability proofs, and to derive Theorems 3.1 and 3.2. The author reviewed and takes responsibility for every statement and proof. AI systems are not authors of this work, and the work has not been peer reviewed.

References

[1] Murad Aznagulov. Thermodynamic realizability of hidden Markov processes: Attainment, observable certificates, and the price of architecture. *arXiv preprint*, page 2609.14623, 2026. arXiv:2609.14623v1, submitted 13 September 2026.

[2] R. R. Bahadur and Leonard J. Savage. The nonexistence of certain statistical procedures in nonparametric problems. *The Annals of Mathematical Statistics*, 27(4):1115–1122, 1956. doi: 10.1214/aoms/1177728077.

[3] Gili Bisker, Matteo Polettini, Todd R. Gingrich, and Jordan M. Horowitz. Hierarchical bounds on entropy production inferred from partial information. *Journal of Statistical Mechanics: Theory and Experiment*, page 093210, 2017. doi: 10.1088/1742-5468/aa8c0d. arXiv:1708.06769; Sections 3.2, 5 and 6.

[4] Peter Buchholz and Miklós Telek. On minimal representations of Rational Arrival Processes. *Annals of Operations Research*, 202(1):35–58, 2013. doi: 10.1007/s10479-011-1001-5. URL `https://link.springer.com/article/10.1007/s10479-011-1001-5`.

[5] Julius Degünther, Jann van der Meer, and Udo Seifert. Fluctuating entropy production on the coarse-grained level: Inference and localization of irreversibility. *Physical Review Research*, 6:023175, 2024. doi: 10.1103/PhysRevResearch.6.023175. arXiv:2309.07665; Sections II.C–II.E: Markovian events and trajectory snippets.

[6] Jannik Ehrich. Tightest bound on hidden entropy production from partially observed dynamics. *Journal of Statistical Mechanics: Theory and Experiment*, page 083214, 2021. doi: 10.1088/1742-5468/ac150e. arXiv:2105.08803.

[7] Massimiliano Esposito. Stochastic thermodynamics under coarse graining. *Physical Review E*, 85:041125, 2012. doi: 10.1103/PhysRevE.85.041125. URL `https://arxiv.org/abs/1112.5410`.

[8] Massimiliano Esposito and Juan M. R. Parrondo. Stochastic thermodynamics of hidden pumps. *Physical Review E*, 91:052114, 2015. doi: 10.1103/PhysRevE.91.052114. arXiv:1310.2987.

[9] Jie Gu. Excursion-resolved thermodynamic inference without observing reverse transitions. *arXiv preprint*, page 2609.39205, 2026. URL `https://arxiv.org/abs/2609.39205v1`. arXiv:2609.39205v1; unrefereed version 1, submitted 30 September 2026.

[10] Illés Horváth and Miklós Telek. A constructive proof of the phase-type characterization theorem. *Stochastic Models*, 31(2):316–350, 2015. doi: 10.1080/15326349.2015.1012912. URL https://doi.org/10.1080/15326349.2015.1012912.

[11] Oleg A. Igoshin, Anatoly B. Kolomeisky, and Dmitrii E. Makarov. Uncovering dissipation from coarse observables: A case study of a random walk with unobserved internal states. *The Journal of Chemical Physics*, 162:034111, 2025. doi: 10.1063/5.0247331.

[12] Uri Kapustin, Aishani Ghosal, and Gili Bisker. Utilizing time-series measurements for entropy-production estimation in partially observed systems. *Physical Review Research*, 6:023039, 2024. doi: 10.1103/PhysRevResearch.6.023039. arXiv:2212.13487v1 is the earlier 2022 preprint.

[13] Bo Henry Lindqvist. Phase-type models for competing risks, with emphasis on identifiability issues. *Lifetime Data Analysis*, 29(2):318–341, 2023. doi: 10.1007/s10985-022-09547-7. URL https://link.springer.com/article/10.1007/s10985-022-09547-7.

[14] Ignacio A. Martínez, Gili Bisker, Jordan M. Horowitz, and Juan M. R. Parrondo. Inferring broken detailed balance in the absence of observable currents. *Nature Communications*, 10:3542, 2019. doi: 10.1038/s41467-019-11051-w. URL https://arxiv.org/abs/1809.02084.

[15] Eden Nitzan, Aishani Ghosal, and Gili Bisker. Universal bounds on entropy production inferred from observed statistics. *Physical Review Research*, 5:043251, 2023. doi: 10.1103/PhysRevResearch.5.043251. arXiv:2212.01783.

[16] Matteo Polettini and Massimiliano Esposito. Effective thermodynamics for a marginal observer. *Physical Review Letters*, 119:240601, 2017. doi: 10.1103/PhysRevLett.119.240601. arXiv:1703.05715.

[17] Édgar Roldán and Juan M. R. Parrondo. Estimating dissipation from single stationary trajectories. *Physical Review Letters*, 105:150607, 2010. doi: 10.1103/PhysRevLett.105.150607. arXiv:1004.2831.

[18] Édgar Roldán and Juan M. R. Parrondo. Entropy production and Kullback–Leibler divergence between stationary trajectories of discrete systems. *Physical Review E*, 85:031129, 2012. doi: 10.1103/PhysRevE.85.031129. arXiv:1201.5613.

[19] Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, 75:126001, 2012. doi: 10.1088/0034-4885/75/12/126001. arXiv:1205.4176.

[20] Dominic J. Skinner and Jörn Dunkel. Improved bounds on entropy production in living systems. *Proceedings of the National Academy of Sciences*, 118:e2024300118, 2021. doi: 10.1073/pnas.2024300118. arXiv:2011.08765.

[21] Dominic J. Skinner and Jörn Dunkel. Estimating entropy production from waiting time distributions. *Physical Review Letters*, 127:198101, 2021. doi: 10.1103/PhysRevLett.127.198101. arXiv:2105.08681.

[22] Miklós Telek and Gábor Horváth. A minimal representation of Markov arrival processes and a moments matching method. *Performance Evaluation*, 64(9–12):1153–1168, 2007. doi: 10.1016/j.peva.2007.06.001. URL https://www.sciencedirect.com/science/article/abs/pii/S0166531607000508.

[23] Gianluca Teza and Attilio L. Stella. Exact coarse graining preserves entropy production out of equilibrium. *Physical Review Letters*, 125:110601, 2020. doi: 10.1103/PhysRevLett.125.110601. arXiv:2003.08674.

[24] Jann van der Meer, Benjamin Ertel, and Udo Seifert. Thermodynamic inference in partially accessible Markov networks: A unifying perspective from transition-based waiting time distributions. *Physical Review X*, 12:031025, 2022. doi: 10.1103/PhysRevX.12.031025. arXiv:2203.12020.

Appendix A. A count-only obstruction beyond the positive-determinant regime

This section allows any finite number of hidden states and does not assume $\det R > 0$ or prescribe L. The direct and excursion rates appearing in logarithms are assumed positive. The affinity argument is an application of standard relative-entropy contraction; its matrix representation exposes the relevant determinant obstruction.

Theorem A.1 (Occurrence-determinant bound)**.** *For every finite irreducible reciprocal chain under the stated observation convention,*

$$\sigma(Q) \geq B(d) + (R_{12} + R_{21}) \left[\log \frac{R_{12}R_{21}}{R_{11}R_{22}}\right]_+, \qquad [x]_+ = \max(x, 0). \tag{A.1}$$

In particular, if $\det R < 0$*, every compatible model, irrespective of its finite hidden dimension, has dissipation strictly greater than* $B(d)$ *by the displayed amount.*

Proof. For a hidden state h, let $d_h = \sum_{y \neq h} n_{hy}$ be its total stationary departure flux. It is positive. Define the symmetric hidden matrix and visible-to-hidden matrix

$$J_{hh} = d_h, \quad J_{hk} = -\sqrt{n_{hk}n_{kh}}\ (h \neq k), \qquad Z_{ih} = \sqrt{n_{ih}n_{hi}}. \tag{A.2}$$

These matrices use flows, not transition rates. By stationarity and the arithmetic–geometric mean inequality,

$$d_h - \sum_{k \in H \setminus \{h\}} \sqrt{n_{hk}n_{kh}} \geq \frac{1}{2} \sum_{i \in V} (n_{hi} + n_{ih}) \geq 0. \tag{A.3}$$

Every hidden connected component has an edge to a visible state, because the full reciprocal graph is connected. Hence on each such component at least one row is strictly diagonally dominant. Explicitly, for a vector u,

$$u^T J u = \sum_{\{h,k\} \subset H} \sqrt{n_{hk}n_{kh}}(u_h - u_k)^2 + \sum_h \left(d_h - \sum_{k \neq h} \sqrt{n_{hk}n_{kh}} \right) u_h^2.$$

Vanishing forces u constant on each hidden component and zero at a strictly dominant row. Thus J is positive definite.

Let $\gamma = (i, h_1, \ldots, h_m, j)$ be an excursion jump sequence, allowing repeated hidden states but no consecutive repeats. After the holding times are integrated out, its stationary rate is

$$w(\gamma) = \frac{n_{ih_1}n_{h_1h_2} \cdots n_{h_{m-1}h_m}n_{h_mj}}{d_{h_1} \cdots d_{h_m}}. \tag{A.4}$$

The holding-time factors in the forward/reverse ratio cancel before integration, so keeping this jump sequence keeps its path-entropy contribution. Reversal, denoted $\bar{\gamma}$, is a bijection between opposite endpoint classes. Summing their geometric means yields

$$G_{ij} = \sum_{\gamma : i \to H \to j} \sqrt{w(\gamma)w(\bar{\gamma})} = (ZJ^{-1}Z^T)_{ij}. \tag{A.5}$$

To justify the sum, write $J = D - A$ with $D = \mathrm{diag}(d_h)$ and $A \geq 0$ symmetric. The normalized matrix $D^{-1/2}AD^{-1/2}$ has spectral radius below one. Indeed positive definiteness bounds its largest eigenvalue below one, and nonnegativity makes that eigenvalue its spectral radius. The Neumann series for J^{-1} then expands exactly into the successive hidden paths in (A.4).

The matrix G is positive semidefinite. Cauchy–Schwarz in each return-path class gives $G_{ii} \leq R_{ii}$, since reversal preserves the total return rate. Consequently

$$G_{12} \leq \sqrt{G_{11}G_{22}} \leq \sqrt{R_{11}R_{22}}. \tag{A.6}$$

On the set of $1 \to H \to 2$ sequences, normalize forward and reversed rates as $p = w/R_{12}$ and $q = w \circ \mathrm{rev}/R_{21}$. Their affinity is

$$A(p, q) = \sum_\gamma \sqrt{p(\gamma)q(\gamma)} = G_{12}/\sqrt{R_{12}R_{21}}.$$

Jensen's inequality under p gives $D(p\|q) \geq -2 \log A(p, q)$; the same inequality holds with p, q reversed. These sums are finite in entropy for every fixed admissible model, by the stationary edge-counting argument used for (5.1). Applying that argument to jump sequences gives

$$\sigma(Q) - B(d) \geq R_{12}D(p\|q) + R_{21}D(q\|p) \geq -2(R_{12} + R_{21}) \log A(p, q).$$

Use (A.6) when its ratio is below one, and the original nonnegative entropy excess otherwise. This proves (A.1). □

A.1. **An exact negative-determinant example.** Take two independent exponential channels

$$M_1 = \begin{pmatrix} 1 & 4 \\ 1/4 & 1 \end{pmatrix}, \qquad M_2 = \begin{pmatrix} 1 & 1/4 \\ 4 & 1 \end{pmatrix}, \qquad t_1 = t_2 = 1/50. \tag{A.7}$$

Both matrices are positive rank one. The channel construction gives

$$R = \begin{pmatrix} 2 & 17/4 \\ 17/4 & 2 \end{pmatrix}, \quad L = R/50, \quad P = (3/8, 3/8, 1/4), \quad F_{12} = F_{21} = 1. \tag{A.8}$$

The two hidden masses are $1/8, 1/8$, all required rates are positive, and $B(d) = 0$. Nevertheless,

$$I_\infty(d) \geq \frac{17}{2} \log \frac{289}{64} > 0. \tag{A.9}$$

The displayed model has $\sigma = (15/2) \log 16$, which is an upper witness for the present inequality. Section 9 strengthens the inequality and proves that this witness is optimal. For this model, $J = (25/4)I$ and all entries of Z equal $5/2$, so G is the matrix with all entries equal to two. Thus the Gram constraint is exact here, even though the relative-entropy inequality need not be saturated. Equal cross means by themselves do not imply equality of the unrestricted infimum with $B(d)$ outside the positive-determinant regime. The present coarse affinity inequality alone does not decide reduced safety; the equality classification in Section 10 does so.

Appendix B. First-moment realization details

B.1. **Attainment at fixed nonsingular data.**

Proof of Proposition 7.1. Every hidden rate is at most τ, since each hidden total departure rate is a positive diagonal entry of K and their sum is τ. Visible-to-hidden rates are bounded by k_i/P_i, where $k_i = \sum_j R_{ij}$ is fixed. Direct rates are fixed by F, P. Thus all rates lie in a fixed compact box.

The hidden masses have a uniform positive lower bound. Let $k_{\max} = \max(k_1, k_2)$. Since each row of C has sum at most τ, $|\det C| \leq \tau^2$. For either hidden state h, stationarity gives $\sum_i A_{ih} \leq \pi_h K_{hh} \leq \pi_h \tau$. Bounding the other column entries of A by $k_{\max}$ gives $|\det A| \leq \pi_h \tau k_{\max}$. From $\det A \det C = \det R \det K$ we obtain

$$\pi_h \geq \frac{|\det R| \Delta}{\tau^3 k_{\max}} > 0. \tag{B.1}$$

Take a minimizing sequence of finite-entropy fits and a convergent subsequence of their rates and stationary masses. Equations (7.3) keep K uniformly nonsingular, so the excursion data remain equal to R, L in the limit. The limiting stationary masses remain positive by (B.1).

Extend $(x - y) \log(x/y)$ to the nonnegative quadrant by assigning zero at $(0, 0)$ and $+\infty$ when exactly one argument is positive. This function is lower semicontinuous. A finite-entropy sublevel limit therefore has reciprocal support. It is irreducible: the direct visible edge connects the visible states, and a hidden component disconnected from them would make the killed block K singular. Entropy lower semicontinuity now gives a minimizing admissible model. This proves the proposition. Notice that a general nonclosed reciprocal-support class is *not* being declared compact; compactness is used together with finite entropy and the nonsingular-data constraints. □

B.2. **Equal means admit a saturating two-state realization.** Assume now $\mathcal{C}_2(d) \neq \varnothing$, $\det R > 0$ and $\Delta m = 0$. Set

$$c = (R_{12}/R_{21}, 1)^T, \qquad D_c = \operatorname{diag}(c), \qquad G = D_c^{-1} R, \quad H = D_c^{-1} L.$$

The cross-mean condition makes both G, H symmetric. Their diagonal entries are positive; $\det G > 0$, and $\det H > 0$ because $\det L = \det R / \det K > 0$. Consequently both are positive definite.

Their unique symmetric positive square roots define

$$S = H^{-1/2} G H^{-1/2}, \qquad Z = H^{-1/2} G.$$

Both S and H are positive definite and Z is invertible. The forced symmetrised Laplace kernel is

$$D_c^{-1} \mathcal{K}(s) = G(G + sH)^{-1} G = Z^T (S + sI)^{-1} Z. \tag{B.2}$$

Its time-density is $\mathcal{F}(t) = Z^T e^{-St} Z$. Because (B.2) agrees with the kernel of an actual feasible model, uniqueness of the Laplace transform of finite exponential matrix functions makes $\mathcal{F}(t)$ entrywise nonnegative. In particular $\mathcal{F}_{12}(0) \geq 0$.

We need the following elementary two-dimensional realization lemma.

Lemma B.1 (Positive two-mode rotation)**.** *If S is a 2×2 positive-definite symmetric matrix, Z is invertible, and $[Z^T e^{-St} Z]_{12} \geq 0$ for every $t \geq 0$, there is an orthogonal change of hidden coordinates giving matrices M, B with*

$$M > 0, \quad M_{12} \leq 0, \quad B \geq 0 \text{ entrywise}, \qquad B^T (M + sI)^{-1} B = Z^T (S + sI)^{-1} Z. \tag{B.3}$$

Proof. Diagonalise S with eigenvalues $0 < \lambda_1 \leq \lambda_2$. For distinct eigenvalues write the transformed rows of Z as u, v. Positivity at large t implies $u_1 u_2 \geq 0$, and at zero implies $u_1 u_2 + v_1 v_2 \geq 0$. Choose the sign of u so $u \geq 0$.

If $u_1 u_2 = 0$, the zero-time inequality gives $v_1 v_2 \geq 0$, so choose its sign to have $v \geq 0$; the diagonal coordinates already work. The same applies when $u > 0$ and $v_1 v_2 \geq 0$.

In the remaining case choose the sign of v so $v_1 < 0 < v_2$. The interval

$$-v_1/u_1 \leq t_0 \leq u_2/v_2$$

is nonempty precisely because of the zero-time inequality. Let $\theta \in (0, \pi/2)$ satisfy $\tan\theta = t_0$, and rotate the hidden rows by

$$O = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Both columns of $B = O\binom{u}{v}$ are nonnegative. Also

$$[O \operatorname{diag}(\lambda_1, \lambda_2) O^T]_{12} = (\lambda_1 - \lambda_2) \sin\theta \cos\theta \leq 0.$$

This proves the assertion in the distinct case, including the case of a degenerate allowable interval. If the eigenvalues coincide, $S = \lambda I$. The positive-definite Gram matrix $Z^T Z$ has nonnegative off-diagonal entry and therefore has a nonnegative triangular Gram factor B (for a Gram matrix $\left(\begin{smallmatrix} a & c \\ c & d \end{smallmatrix}\right)$, use $B = \left(\begin{smallmatrix} \sqrt{a} & c/\sqrt{a} \\ 0 & \sqrt{d - c^2/a} \end{smallmatrix}\right)$). Two invertible matrices with the same Gram matrix differ by an orthogonal factor. Take $M = \lambda I$. □

Apply the lemma to (B.2). Define

$$W = M^{-1} B, \quad t = W\mathbf{1}, \quad \pi_H = \operatorname{diag}(t) W c,$$

and

$$K_* = \operatorname{diag}(t)^{-1} M \operatorname{diag}(t), \quad C_* = \operatorname{diag}(t)^{-1} B, \quad A_* = D_c B^T \operatorname{diag}(t). \tag{B.4}$$

The inverse of a symmetric positive-definite 2×2 matrix with nonpositive off-diagonal entries is entrywise nonnegative. Since B is invertible and nonnegative, every row is nonzero, so t and Wc are strictly positive. Thus both hidden masses are positive. Moreover $K_* \mathbf{1} = C_* \mathbf{1}$, and off-diagonal entries of K_* are nonpositive, so $-K_*$ is a valid killed generator with exit matrix C_*.

Use the prescribed visible masses, entrance rates $Q_{VH} = \operatorname{diag}(P_1, P_2)^{-1} A_*$ and direct rates F_{ij}/P_i. The hidden stationarity identity follows from $K_*^T \pi_H = A_*^T \mathbf{1}$. Also

$$\mathbf{1}^T \pi_H = c^T W^T W \mathbf{1} = c^T H \mathbf{1} = P_H.$$

The realized matrices are

$$A_* K_*^{-1} C_* = D_c B^T M^{-1} B = R, \qquad A_* K_*^{-2} C_* = D_c B^T M^{-2} B = L.$$

Consequently the coarse flows and visible stationarity also match. Reciprocal support holds at every boundary edge since both directions vanish exactly when the corresponding entry of B vanishes, and at the hidden pair because M is symmetric. Each hidden state connects to a visible state because its row of B is nonzero. Thus the full chain is irreducible.

Finally introduce positive reference weights $w_i = P_i/c_i$ at visible states and $w_h = t_h^2$ at hidden states. Every edge with a hidden endpoint obeys

$$w_x q_{xy} = w_y q_{yx}.$$

This is reversibility on the hidden-containing subnetwork with respect to reference weights, not a claim that the stationary whole model is reversible. Along any excursion $i \to H \to j$, its stationary path-rate ratio to its reverse is c_i/c_j. Return excursions have ratio one. The cross-excursion entropy is therefore exactly $(R_{12} - R_{21}) \log(c_1/c_2)$, and direct jumps provide the other term in (2.1). The constructed model attains $B(d)$. This proves the sufficient direction of Theorem 4.2.

B.3. **Second moments restore a dimension-independent positive gap.** Let $M_p = \mathbb{E}_p T^2$ and $M_q = \mathbb{E}_q T^2$. Every finite model has finite values, but a compatible-model class need not have a uniform bound. The elementary inequality

$$(x - y) \log(x/y) \geq 2(x - y)^2/(x + y)$$

and Cauchy–Schwarz give

$$D(p\|q) + D(q\|p) \geq 2 \int \frac{(p - q)^2}{p + q} \geq \frac{2(\Delta m)^2}{M_p + M_q}.$$

Combining with (5.1) proves, in *any* finite hidden dimension,

$$\sigma(Q) - B(d) \geq \frac{2 \min(R_{12}, R_{21})(\Delta m)^2}{M_p + M_q}. \tag{B.5}$$

Since every finite admissible chain has finite second moments, (B.5) also proves the nonattainment assertion in Theorem 4.1 whenever $\Delta m \neq 0$. A known finite upper bound on the sum of conditional second moments restores a strictly positive unrestricted lower bound whenever $\Delta m \neq 0$.

When $\Delta m \neq 0$, any sequence at these fixed data with entropy excess $e_n = \sigma(Q_n) - B(d) \to 0$ must have

$$M_{p,n} + M_{q,n} \geq \frac{2 \min(R_{12}, R_{21})(\Delta m)^2}{e_n} \longrightarrow \infty. \tag{B.6}$$

B.4. **An explicit example.**

B.4.1. *Feasible but unsafe reduced data.* Take

$$R = \begin{pmatrix} 9/2 & 3 \\ 2 & 5/2 \end{pmatrix}, \quad L = \begin{pmatrix} 1/20 & 1/25 \\ 3/100 & 9/200 \end{pmatrix},$$

$$P = (2/5, 87/200, 33/200), \qquad F_{12} = 1, \quad F_{21} = 2.$$

All other coarse flows are determined by consistency. A two-hidden-state fit consists of channels

$$M_1 = \begin{pmatrix} 4 & 2 \\ 1 & 1/2 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1/2 & 1 \\ 1 & 2 \end{pmatrix}, \quad t_1 = 1/100, \quad t_2 = 1/50.$$

Here $B(d) = \log 3$, this displayed witness has entropy $\log 4$, $\Delta m = -1/600$, $a = 3/100$ and $b = 1/5000$. Hence

$$I_\infty = I_4 = \log 3, \qquad I_2 \geq \log 3 + \frac{5(17 - 12\sqrt{2})}{18}. \tag{B.7}$$

The last additive term is approximately 0.0081770143. The construction in Section 6 supplies exact six-state fits with these same data and any sufficiently small positive ε, for example with fast time $\delta = 1/1000$.

Department of Physics, École normale supérieure, Université PSL, Paris, France
Email address: samuil.petkov@phys.ens.psl.eu