

An exactly solvable Markov chain with a fully characterized lumping lattice: equilateral geometry, three-eigenvalue spectrum, and exact coarse-grained relaxation

Tairan Miranda Krügel
Independent researcher, Giruá, RS, Brazil

August 2026

Abstract

Coarse-graining a Markov chain by merging microstates (lumping) almost always destroys Markovianity: the reduced dynamics acquires memory, and exact reduced descriptions require semi-Markov constructions. We introduce and solve a class of reversible finite Markov chains for which the opposite extreme holds. The class $\text{LD}(u, a, \Delta, s)$ is a Metropolis walk, at inverse temperature x , on a “double-pencil” family of subsets: the u one-point deletions and a one-point additions of a reference configuration U , with two energy levels separated by Δ , and pair proposals drawn from an ambient alphabet of s sites. We prove that (i) the transition graph is the complete graph K_{u+a} with multiplicity one, a consequence of the equilaterality of the unit Hamming sphere; (ii) the spectrum consists of exactly three distinct eigenvalues, 1 , $b(x) = 1 - (u + a e^{-x\Delta})/\binom{s}{2}$ and $c = 1 - (u + a)/\binom{s}{2}$, with multiplicities 1 , u and $a - 1$, so that the relaxation gap saturates at $u/\binom{s}{2}$ instead of closing or opening exponentially (a non-Arrhenius relaxation in the presence of Arrhenius populations); and (iii) for $x > 0$ the lumping lattice is completely characterized: a partition is lumpable if and only if every class that contains excited states either is pure or contains *all* excited states (the absorbing family), which includes, as boundary cases, all refinements of the two-sector partition and the trivial partition. To our knowledge this is the first nontrivial model whose full lumping lattice is explicitly characterized, answering constructively the recently posed question of whether exact Markovian coarse-graining by lumping is possible beyond special symmetric cases. Every statement is certified by exhaustive or large-sample exact computation with integer arithmetic certificates; a worked instance $\text{LD}(28, 54, 1, 136)$, arising from a constant-energy variety of a classical $[[137, 124, 2]]$ error-correcting code ensemble, is analyzed with all quantities audited to machine precision. We also quantify exactly the share of relaxation entropy production (Kullback–Leibler decay) visible at each coarse-graining level. Limitations (finite size, equilibrium, extreme symmetry) and a driven non-equilibrium extension are discussed.

Contents

1	Introduction	2
2	The double-pencil class	3
2.1	Definition	3
2.2	Equilaterality and the complete graph	3
2.3	The spectrum	4
3	The lumping lattice, completely characterized	5

4	Relaxation thermodynamics under exact coarse-graining	6
5	An audited worked instance: LD(28, 54, 1, 136)	7
6	Discussion	8

1 Introduction

Coarse-graining is the physicist’s oldest compression algorithm: replace many microstates by one macrostate and hope the reduced dynamics remains closed. For Markov chains, the exact version of this hope is *lumpability* (also called ordinary lumpability or strong lumpability) [1, 2]: a partition of the state space is lumpable when the merged process is again Markov, for every initial condition. It is classical that the lumpable partitions of a chain form a complete lattice, and efficient algorithms compute the *coarsest* lumpable partition [3]. What is essentially never available — because generic chains admit only the trivial lattice — is a *complete, explicit characterization of all lumpable partitions* of a nontrivial model.

The question has acquired fresh urgency in stochastic thermodynamics. When experiments resolve only coarse observables, entropy production is partially hidden [5, 7, 12, 13, 14], and the structure of the coarse-grained process controls what can be inferred. In a recent systematic study, Teza, Stella and GrandPre [4] showed that lumping generically produces a non-Markovian (semi-Markov) reduced process, preserving mean entropy production but not its full statistics, and posed explicitly the question of whether a coarsened Markovian description can ever be exact beyond special symmetric cases.

This paper answers that question constructively and, in a precise sense, maximally. We define a two-parameter-energy class of reversible Metropolis chains — the *double-pencil class* $\text{LD}(u, a, \Delta, s)$ — whose state space is a family of subsets at Hamming distance one from a reference configuration (Section 2), and we prove:

- (i) **Complete-graph geometry** (Theorem 3): the transition graph is K_{u+a} with multiplicity one, because the unit Hamming sphere is an equilateral set — every pair of distinct states differs by exactly one site-pair swap.
- (ii) **Three-eigenvalue spectrum** (Theorem 5): the chain has exactly three distinct eigenvalues, in closed form, with an explicit eigenbasis. The relaxation gap $\gamma(x) = (u + a e^{-x\Delta}) / \binom{s}{2}$ interpolates between $(u+a) / \binom{s}{2}$ at $x = 0$ and $u / \binom{s}{2}$ as $x \rightarrow \infty$: it *saturates* rather than vanishing or diverging, although equilibrium populations obey an Arrhenius law. Relaxation is entropic, not activated — the suppressed direction is uphill, and downhill escape is free.
- (iii) **The full lumping lattice** (Theorem 7): for $x > 0$, a partition is lumpable if and only if every class that contains excited states either is pure (contained in the excited sector) or contains *all* excited states, possibly absorbing arbitrary ground states — the *absorbing family*, whose existence relies on the asymmetry of the Metropolis rule. All refinements of the two-sector partition are included as the boundary case of zero absorbed ground states.

The chain is reversible, so its steady-state entropy production vanishes; its nonequilibrium content lives in relaxation. In Section 4 we quantify exactly how much of the relaxation entropy production (Kullback–Leibler decay relative to equilibrium [6, 12]) remains visible after each admissible coarse-graining, using the chain-rule decomposition of the KL divergence.

Every theorem is accompanied by an exact computational certificate (Section 5): the spectra of 800 independently constructed instances match the closed forms to better than 5×10^{-15} ; the lumping dichotomy was verified against measurement on 160,000 random partitions

with zero counterexamples; and a fully audited worked instance, $\text{LD}(28, 54, 1, 136)$ — which arises as the constant-energy variety of a classical $[[137, 124, 2]]$ error-correcting code ensemble — anchors all constructions at machine precision. Section 6 discusses scope, limitations, and a driven extension aimed at hidden entropy production in genuine nonequilibrium steady states.

2 The double-pencil class

2.1 Definition

Let $\mathcal{S} = U \sqcup A \sqcup R$ be a finite set of *sites*, with $|U| = u \geq 1$, $|A| = a \geq 1$, and $|R| = r \geq 0$ inert ambient sites, $s := u + a + r$. Configurations are subsets of $\mathcal{S}$. The *double pencil* around the reference configuration U is the family

$$\mathcal{V} = \{G_e := U \setminus \{e\} : e \in U\} \cup \{X_f := U \cup \{f\} : f \in A\}, \quad (1)$$

of $n := u + a$ states: the u one-point deletions (the *ground sector* $\mathbb{G}$, energy 0) and the a one-point additions (the *excited sector* $\mathbb{X}$, energy $\Delta > 0$).

Definition 1 (The class $\text{LD}(u, a, \Delta, s)$). Fix $x \geq 0$ (inverse temperature) and write $N := \binom{s}{2}$, $m := e^{-x\Delta}$. One step of the chain: draw a pair $\{p, q\} \subset \mathcal{S}$ uniformly among the N pairs, and propose $S \mapsto S \triangle \{p, q\}$ (symmetric difference). If the result is not in $\mathcal{V}$, the move is rejected (self-loop). If it is in $\mathcal{V}$, accept with probability 1 if it does not increase the energy, and with probability m if it does ($\mathbb{G} \rightarrow \mathbb{X}$).

The chain is finite, irreducible on $\mathcal{V}$ (Theorem 3), and reversible with respect to

$$\pi_i = \frac{1}{Z} \begin{cases} 1, & i \in \mathbb{G}, \\ m, & i \in \mathbb{X}, \end{cases} \quad Z = u + am, \quad (2)$$

since each accepted proposal is a single pair-swap and the acceptance satisfies detailed balance.

Remark 2. Only site-pairs contained in $U \cup A$ ever produce moves; pairs touching R are always rejected. The ambient size s therefore enters only through the normalization $N = \binom{s}{2}$. Allowing $s > u + a$ is not cosmetic: the audited physical instance of Section 5 has $u + a = 82$ states embedded in $s = 136$ proposal sites.

2.2 Equilaterality and the complete graph

Recall that the Hamming distance between configurations is $d_H(S, S') = |S \triangle S'|$. The unit sphere $\mathcal{S}_1(U) = \{U \triangle \{z\} : z \in \mathcal{S}\}$ is *equilateral*: for distinct z, z' ,

$$(U \triangle \{z\}) \triangle (U \triangle \{z'\}) = \{z, z'\}, \quad d_H = 2, \quad (3)$$

whether z, z' are both in U , both outside, or mixed. Equilaterality is hereditary to subsets, and $\mathcal{V} \subset \mathcal{S}_1(U)$. This elementary observation — two-distance sets in Hamming space are classical objects [10] — has an exact dynamical consequence:

Theorem 3 (Complete graph, multiplicity one). *For every $u, a \geq 1$, $\Delta > 0$, $s \geq u + a$ and every $x \geq 0$, the transition graph of $\text{LD}(u, a, \Delta, s)$ on $\mathcal{V}$ is the complete graph K_n , $n = u + a$: each state has exactly $n - 1$ outgoing moves, one to each other state, each realized by exactly one site pair.*

Proof. By (3), distinct states $S, S' \in \mathcal{V}$ satisfy $S \triangle S' = \{z, z'\}$ for a unique unordered pair, so the proposal $\{z, z'\}$ connects them and no other pair does. It remains to check that this pair always exists within the alphabet, i.e. that the symmetric difference of any two members of $\mathcal{V}$ is a pair of sites: deletion–deletion gives $\{e, e'\} \subset U$; addition–addition gives $\{f, f'\} \subset A$; deletion–addition gives $\{e, f\} \in U \times A$. Uniqueness is the uniqueness of the symmetric difference. $\square$

Remark 4. The off-diagonal rates take exactly four values. With $N = \binom{s}{2}$:

$$P_{ij} = \begin{cases} 1/N, & i \in \mathbb{G}, j \in \mathbb{G}, j \neq i, \text{ or } i \in \mathbb{X} \rightarrow j \in \mathbb{G}, \text{ or } i \in \mathbb{X}, j \in \mathbb{X}, j \neq i, \\ m/N, & i \in \mathbb{G} \rightarrow j \in \mathbb{X} \quad (\text{uphill}). \end{cases} \quad (4)$$

and the diagonal completes the row sums: $P_{ii} = 1 - (u - 1 + am)/N$ for $i \in \mathbb{G}$ and $P_{ii} = 1 - (u + a - 1)/N$ for $i \in \mathbb{X}$. Note the asymmetry that will drive Theorem 7: the suppressed rate m/N appears *only* in the uphill direction.

2.3 The spectrum

Theorem 5 (Three-eigenvalue spectrum). *For $x > 0$ (so $m < 1$), the transition matrix of $\text{LD}(u, a, \Delta, s)$ has exactly three distinct eigenvalues:*

$$\lambda_0 = 1 \quad (\times 1), \quad b(x) = 1 - \frac{u + am}{N} \quad (\times u), \quad c = 1 - \frac{u + a}{N} \quad (\times a - 1), \quad (5)$$

with $1 > b(x) > c$. An explicit eigenbasis is: (i) the stationary vector; (ii) the $u - 1$ vectors supported on $\mathbb{G}$ with zero sum, plus one two-sector mode obtained by lumping $\{\mathbb{G}, \mathbb{X}\}$; (iii) the $a - 1$ vectors supported on $\mathbb{X}$ with zero sum.

Proof. Family (iii): let v vanish on $\mathbb{G}$ with $\sum_{j \in \mathbb{X}} v_j = 0$. For $i \in \mathbb{X}$, $(Pv)_i = \frac{1}{N} \sum_{j \in \mathbb{X}, j \neq i} v_j + (1 - \frac{u+a-1}{N})v_i = -\frac{v_i}{N} + (1 - \frac{u+a-1}{N})v_i = cv_i$, and $(Pv)_i$ for $i \in \mathbb{G}$ receives only $\frac{m}{N} \sum_{j \in \mathbb{X}} v_j = 0$. Hence v is an eigenvector with eigenvalue c : an $(a - 1)$ -dimensional space.

Family (ii): let w vanish on $\mathbb{X}$ with $\sum_{j \in \mathbb{G}} w_j = 0$. For $i \in \mathbb{G}$, $(Pw)_i = -\frac{w_i}{N} + (1 - \frac{u-1+am}{N})w_i = b(x)w_i$, and for $i \in \mathbb{X}$, $(Pw)_i = \frac{1}{N} \sum_{j \in \mathbb{G}} w_j = 0$: a $(u - 1)$ -dimensional eigenspace with eigenvalue $b(x)$.

The subspace $\text{span}\{\mathcal{K}_{\mathbb{G}}, \mathcal{K}_{\mathbb{X}}\}$ is invariant (lumpability of the two-sector partition, a special case of Theorem 7); the restricted 2×2 chain has row sums am/N ($\mathbb{G} \rightarrow \mathbb{X}$) and u/N ($\mathbb{X} \rightarrow \mathbb{G}$), hence eigenvalues 1 and $1 - (u + am)/N = b(x)$, supplying the stationary vector and the last copy of b . The dimensions add up: $1 + (u - 1) + 1 + (a - 1) = n$, and $b(x) - c = a(1 - m)/N > 0$ for $m < 1$ ensures three *distinct* values. $\square$

Corollary 6 (Non-Arrhenius relaxation). *The spectral gap is*

$$\gamma(x) = 1 - b(x) = \frac{u + ae^{-x\Delta}}{\binom{s}{2}}, \quad \gamma(0) = \frac{u + a}{N} \searrow \gamma(\infty) = \frac{u}{N}. \quad (6)$$

Although the equilibrium population of the excited sector obeys an Arrhenius law ($\propto ae^{-x\Delta}$), the relaxation rate never closes and never diverges: it saturates at u/N , set by the free downhill escape $\mathbb{X} \rightarrow \mathbb{G}$ and the intra-sector mixing. The barrier Δ controls populations, not timescales.

3 The lumping lattice, completely characterized

Recall the Kemeny–Snell criterion [1]: a partition $\Pi = \{C\}$ of $\mathcal{V}$ is *lumpable* for the chain P if for every pair of classes M, C the sum $\sum_{j \in C} P_{ij}$ is the same for all $i \in M$. The lumped matrix is then $\hat{P}_{MC} = \sum_{j \in C} P_{ij}$ (any $i \in M$), and the lumped process is exactly Markov for every initial condition.

Theorem 7 (The full lumping lattice). *Let $x > 0$ and $\Delta > 0$ ($m < 1$). A partition Π of $\mathcal{V}$ is lumpable for $\text{LD}(u, a, \Delta, s)$ if and only if every class $M \in \Pi$ with $M \cap \mathbb{X} \neq \emptyset$ satisfies one of*

$$M \subseteq \mathbb{X} \quad \text{or} \quad \mathbb{X} \subseteq M. \quad (7)$$

In words: classes containing excited states are either pure excited classes, or contain all excited states — in which case they may additionally absorb an arbitrary subset $S \subseteq \mathbb{G}$ of ground states. All remaining classes are subsets of $\mathbb{G}$, arbitrary.

Proof. Write $N = \binom{s}{2}$ and use the four rates of Remark 4.

Sufficiency. Pure classes $\subseteq \mathbb{G}$ or $\subseteq \mathbb{X}$ satisfy the criterion because, within each sector, every off-diagonal rate is $1/N$ (resp. m/N for $\mathbb{G} \rightarrow \mathbb{X}$ targets, counted per class): for $i \in M \subseteq \mathbb{G}$ and any class C , $\sum_{j \in C} P_{ij} = (|C \cap \mathbb{G}| - \mathbb{1}_{C=M})/N + m|C \cap \mathbb{X}|/N + \mathbb{1}_{C \ni i} P_{ii}$, which depends on i only through the shared indicators; similarly for $M \subseteq \mathbb{X}$.

The nontrivial case is a mixed class $M = \mathbb{X} \cup S$, $S \subseteq \mathbb{G}$, with all other classes $\subseteq \mathbb{G} \setminus S$. For $i \in S \subset \mathbb{G}$:

$$\sum_{j \in M} P_{ij} = \frac{|S| - 1}{N} + \frac{am}{N} + P_{ii} = 1 - \frac{u - |S|}{N}, \quad (8)$$

since $P_{ii} = 1 - (u - 1 + am)/N$. For $i' \in \mathbb{X}$:

$$\sum_{j \in M} P_{i'j} = \frac{|S|}{N} + \frac{a - 1}{N} + P_{i'i'} = 1 - \frac{u - |S|}{N}, \quad (9)$$

since $P_{i'i'} = 1 - (u + a - 1)/N$. The two coincide: the diagonal absorbs the only suppressed rate, because the downhill escape $\mathbb{X} \rightarrow \mathbb{G}$ is free and equals the intra-ground rate. Sums to classes $C \subseteq \mathbb{G} \setminus S$ are $|C|/N$ from every row.

Necessity. Suppose a class M is mixed but misses some excited state $j^* \in \mathbb{X} \setminus M$. Let C be the class containing j^* (so $C \neq M$), and pick $i \in M \cap \mathbb{G}$, $i' \in M \cap \mathbb{X}$; note $i, i' \notin C$. Then

$$\sum_{j \in C} P_{ij} = \frac{|C \cap \mathbb{G}|}{N} + \frac{m|C \cap \mathbb{X}|}{N}, \quad \sum_{j \in C} P_{i'j} = \frac{|C \cap \mathbb{G}|}{N} + \frac{|C \cap \mathbb{X}|}{N}, \quad (10)$$

whose difference is $|C \cap \mathbb{X}|(1 - m)/N \geq (1 - m)/N > 0$, since $j^* \in C \cap \mathbb{X}$. Hence the criterion fails, and a lumpable mixed class must contain all of $\mathbb{X}$. $\square$

Corollary 8 (Boundary cases).

- (a) $S = \emptyset$: the pure case recovers all refinements of the two-sector partition $\{\mathbb{G}, \mathbb{X}\}$, down to the n singletons — the chain is lumpable with respect to the finest partition trivially, and every intermediate refinement of each sector separately.
- (b) $S = \mathbb{G}$: the trivial one-class partition.
- (c) $a = 1$: every partition is lumpable, because any class containing the unique excited state contains “all” of $\mathbb{X}$.
- (d) $u = 1$ does not trivialize: a mixed class missing some excited state fails, since uphill proposals hit excited classes at the suppressed rate.

(e) $x = 0$ ($m = 1$): every partition is lumpable (all off-diagonal rates are $1/N$); the characterization (7) is exclusive to $x > 0$.

Remark 9 (Relation to the literature). Algorithmic work computes the coarsest lumpable partition in $O(m \log n)$ [3] and studies approximate (quasi-) lumpability. Theorem 7 is of a different kind: for this class, the *entire* lattice is described in one sentence, including a non-refinement branch (the absorbing family) whose existence depends on the thermodynamic asymmetry $m < 1$. In the context of [4], the class provides exact Markovian lumping at all temperatures, with the reduced state space ranging from 2 macrostates to n .

4 Relaxation thermodynamics under exact coarse-graining

The chain satisfies detailed balance (2), so its steady-state entropy production rate is identically zero. The thermodynamic content is in *relaxation*: for an initial distribution p_0 , the non-adiabatic entropy production along the relaxation to equilibrium is the KL decay [6, 12]

$$\Sigma_{\text{na}} = D(p_0 \parallel \pi) - D(p_\infty \parallel \pi) = D(p_0 \parallel \pi), \quad (11)$$

since $p_t \rightarrow \pi$.

For a lumpable partition Π , let $\hat{p}_t$ be the lumped distribution on classes and $\hat{\pi}$ the lumped equilibrium. The chain rule of the KL divergence gives the *exact* decomposition, valid for any coarse-graining of any chain,

$$D(p_t \parallel \pi) = D(\hat{p}_t \parallel \hat{\pi}) + \sum_M \hat{p}_t(M) D(p_t(\cdot \mid M) \parallel \pi(\cdot \mid M)), \quad (12)$$

i.e. total relaxation EP = macroscopically visible EP + EP hidden inside the classes. What the lumpability lattice of Theorem 7 buys us is that the first term is computed by an *exactly Markovian* reduced chain — no memory kernel, no semi-Markov construction [8, 9, 4] — while the second term is a pure conditional equilibration.

Define the *captured fraction*

$$f(x; \Pi, p_0) := \frac{D(\hat{p}_0 \parallel \hat{\pi})}{D(p_0 \parallel \pi)}. \quad (13)$$

Because all states within a sector share the same equilibrium weight, f is analytically computable for simple initial conditions. Two reference choices used below: $\Pi_2 = \{\mathbb{G}, \mathbb{X}\}$ and, in the worked instance, a seven-class partition Π_7 refining $\mathbb{X}$ into its six symmetry classes (Section 5).

Remark 10 (Tripartite verdict). For the class LD: (a) steady-state EP is trivially preserved ($\sigma = 0$ at every level); (b) the macroscopic *dynamics* is preserved exactly (lumpability); (c) the relaxation EP is *not* preserved in general: $f < 1$ whenever the initial condition is not sector-uniform, and the hidden share is carried by the conditional term of (12). This is the exact, closed-form counterpart of the generic bounds of [5, 11].

For the two-sector lumping and a pure initial state $p_0 = \delta_i$, one obtains

$$f(x; \Pi_2, \delta_{i \in \mathbb{G}}) = \frac{\ln(Z/u)}{\ln Z}, \quad f(x; \Pi_2, \delta_{i \in \mathbb{X}}) = \frac{\ln(Z/(am))}{\ln(Z/m)}, \quad (14)$$

while any sector-uniform initial condition gives $f = 1$. In particular $f \rightarrow 0$ as $x \rightarrow \infty$ for ground-pure starts: at low temperature almost all of the relaxation information of a tagged ground state is the *identity* of the vacancy, invisible to the two-sector observer.

5 An audited worked instance: LD(28, 54, 1, 136)

The class was discovered by solving a concrete finite model first. The instance $u = 28$, $a = 54$, $\Delta = 1$, $s = 136$ arises as the constant-energy *variety* of a Boltzmann ensemble built on a classical $[[137, 124, 2]]$ error-correcting code (thirteen parity blocks, one frozen edge): the 82 lowest-lying site-odd configurations of the ensemble are exactly the 28 one-deletions and 54 one-additions of a reference configuration U (the fully occupied 28-edge block), the 54 additions being filtered by an energy budget among 108 candidates. Nothing in the present paper depends on that origin; we report it because the instance comes with an unusually strong audit trail, summarized in Table 1.

For this instance $N = \binom{136}{2} = 9180$ and the spectrum is

$$1 (\times 1), \quad b(x) = 1 - \frac{28 + 54 e^{-x}}{9180} (\times 28), \quad c = 1 - \frac{82}{9180} (\times 53). \quad (15)$$

At $x = 12$: $b(12) = 0.9969498549251\dots$ and $c = 0.9910675381\dots$, matching direct diagonalization of the code-derived chain to 2.0×10^{-15} . The lumping lattice contains, besides the refinements of Π_2 , the absorbing family $M = \mathbb{X} \cup S$; a refined seven-class partition Π_7 (ground sector plus six symmetry classes of the excited sector) is lumpable as well. Audited captured fractions (13) for this instance are collected in Table 2.

Table 1: Certification summary. “Spectrum, class” : 200 random instances (u, a, Δ) , $u \in \{2, \dots, 30\}$, $a \in \{2, \dots, 60\}$, $\Delta \in \{0.25, 0.5, 1, 2\}$, four values of x each, chains built independently of the code machinery; “lumping dichotomy”: the prediction of Theorem 7 compared against measured row-sum deviations on random partitions (all 200 instances $\times$ 4 values of $x \times$ 200 partitions).

Statement	Test	Result	Max. deviation
Complete graph + spectrum (Thm. 3, 5)	800/800 instances	pass	4.6×10^{-15}
Lumping dichotomy (Thm. 7)	160,000/160,000 partitions	pass	prediction = measurement 3.2×10^{-15}
Absorbing family lumpable ($X \cup S$)	exhaustive within sample	pass	
Anchoring: instance = code-derived chain	spectra at $x = 12$	pass	2.0×10^{-15}
Reversibility $\sigma(x)$	code-derived chain, 4 values of x	$\sigma = 0$	$< 10^{-19}$

Table 2: Audited captured fractions f (Eq. (13)) for LD(28, 54, 1, 136). Pure starts: g_0 (a specific ground state), e_0 (a specific excited state); uniform: sector-uniform. Π_7 : seven-class lumpable partition of the excited sector.

x	$f(\Pi_2, g_0)$	$f(\Pi_2, e_0)$	$f(\Pi_7, e_0)$	$f(\Pi_2, \text{unif.})$
0.5	0.189	0.134	—	1.000
12	4×10^{-6}	0.740	0.857	1.000

Two features deserve emphasis. First, at $x = 12$ a tagged ground state leaves only $f \approx 4 \times 10^{-6}$ of its relaxation EP visible to the two-sector observer — the information is the vacancy identity — whereas a tagged excited state leaves 74% (86% at the Π_7 resolution). Second, the closed forms (14) reproduce the audited values exactly (e.g. $\ln(Z/(am))/\ln(Z/m)$ with $Z = 28 + 54e^{-12}$ gives 0.7398), as they must: for this class the KL decomposition involves no approximation.

6 Discussion

A constructive answer to an open question. Teza, Stella and GrandPre [4] asked whether a coarse-grained description obtained by lumping can ever remain exactly Markovian beyond special symmetric cases, having shown that generically it cannot (memory kernels appear; the mean entropy production survives, the fluctuations do not). The double-pencil class answers in the affirmative with an extreme example: exact Markovian lumping at *every* temperature and for an *explicitly characterized lattice* of partitions, including a non-obvious absorbing branch. The mechanism is transparent: equilateral geometry forces a complete transition graph with only four off-diagonal rates, and the Kemeny–Snell row-sum criterion then reduces to the dichotomy of Theorem 7.

Physical interpretation. The chain is a two-medium system with a single-step barrier: a ground state is the reference configuration with one vacancy; an excited state is the reference configuration with one adatom. Uphill moves are Arrhenius-suppressed, downhill moves are free, and each medium is perfectly mixed (complete graph). The consequences are cleanly separated: populations are activated, relaxation is entropic and saturates (Corollary 6); and the coarse-graining lattice has an absorbing branch precisely because downhill escape is free (Eqs. (8)–(9)).

Limitations, stated plainly. (i) The model is finite and small; there is no thermodynamic limit here. (ii) It is an equilibrium model: $\sigma = 0$, so it does not touch the hardest problem of the area, hidden dissipation in nonequilibrium steady states [5, 13, 14]. Its thermodynamic content is relaxation (KL decay), where it provides exact numbers rather than bounds. (iii) The solvability rests on extreme symmetry; equilateral Hamming families are classical [10], and the contribution is the dynamical consequence, not the combinatorial object. (iv) The value of the model is therefore that of a *benchmark*: a case where every coarse-graining question — Markovianity, spectrum, lattice, KL shares — has a closed-form answer with integer certificates, against which approximate and inference methods [11, 4] can be calibrated exactly.

Outlook: a driven variant. The natural next step is to break detailed balance while preserving the double-pencil geometry (e.g. clockwise/counter-clockwise bias on the pair alphabet), producing a nonequilibrium steady state on K_n whose lumping lattice may remain partially characterizable. If so, the hidden-EP accounting of Section 4 would extend to genuine steady-state dissipation, connecting directly with current inference bounds. Work in progress.

Reproducibility. All numerical certificates of Table 1 are produced by self-contained Python scripts (integer-exact enumeration; fixed seed) deposited together with this preprint on Zenodo. The audit protocol pre-registers target quantities before computation; lost pre-registered bets and errata are documented in the deposit.

Acknowledgments. The author thanks the AI assistant Kimi K3 (Moonshot AI), which acted as critical co-author of the research program behind this paper: independent recomputation, pre-registered falsification attempts, and the discovery of the absorbing family of Theorem 7 via a falsified conjecture.

References

- [1] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, Van Nostrand, Princeton, 1960.
- [2] P. Buchholz, Exact and ordinary lumpability in finite Markov chains, *Journal of Applied Probability* **31**, 59–75 (1994).

- [3] S. Derisavi, H. Hermanns and W. H. Sanders, Optimal state-space lumping in Markov chains, *Information Processing Letters* **87**, 309–315 (2003).
- [4] G. Teza, A. L. Stella and T. GrandPre, Coarse-graining via lumping: exact calculations and fundamental limitations, arXiv:2512.11974 [cond-mat.stat-mech] (2025).
- [5] M. Esposito, Stochastic thermodynamics under coarse graining, *Physical Review E* **85**, 041125 (2012).
- [6] T. Hatano and S. I. Sasa, Steady-state thermodynamics of Langevin systems, *Physical Review Letters* **86**, 3463–3466 (2001).
- [7] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, *Reports on Progress in Physics* **75**, 126001 (2012).
- [8] H. Mori, Transport, collective motion, and Brownian motion, *Progress of Theoretical Physics* **33**, 423–455 (1965).
- [9] R. Zwanzig, Memory effects in irreversible thermodynamics, *Physical Review* **124**, 983–992 (1961).
- [10] P. Delsarte, J. M. Goethals and J. J. Seidel, Spherical codes and designs, *Geometriae Dedicata* **6**, 363–388 (1977).
- [11] L. Cocconi, G. Salbreux and G. Pruessner, Scaling of entropy production under coarse graining in active disordered media, *Physical Review E* **105**, L042601 (2022).
- [12] C. Van den Broeck and M. Esposito, Ensemble and trajectory thermodynamics: a brief introduction, *Physica A* **418**, 6–16 (2015).
- [13] D. Hartich and A. Godec, Comment on “Inferring broken detailed balance in the absence of observable currents”, *Physical Review Letters* **127**, 089801 (2021).
- [14] J. Dieball and A. Godec, Fusing thermodynamic uncertainty relations and fluctuation theorems to estimate entropy production, *Physical Review Letters* **129**, 140601 (2022).
- [15] D. Aldous and J. A. Fill, *Reversible Markov Chains and Random Walks on Graphs*, monograph, 2002.