WORD-POTENTIAL MEASURES: CONJUGACY AND FREQUENCY COMPARISONS

Status. This note develops Jacob Richey's proposed programme of counting the patterns that are weighted, while varying either the allowed words or the reference measure. The conjugacy/SSE framework is classical, not a new classification theorem. The elementary implications, the unbordered-word Gibbs corollary, the obstruction, the frequency asymptotics, and the complete negative-coupling frequency order, and positive-coupling order between distinct least periods below have arguments recorded here. Uniqueness of a frequency crossing within one least-period class is proved for every $q \geq 2$. The binary argument uses period constraints to make two variance criteria exhaustive. Numerical scans are separated from exact certificates in `word-potentials-checks.json`; `word_potentials.py` reproduces them. No independent audit or novelty claim is intended.

1. The object to classify and compare

Let X be a mixing one-dimensional SFT, $a : X \to \mathbb{R}$ a locally constant background potential, and F a finite collection of words. Set

$$f_F(x) = \sum_{w \in F} \mathbf{1}_{[w]}(x), \qquad N_{F,n}(x) = \sum_{j=0}^{n-1} f_F(\sigma^j x).$$

Here $[w]$ means a copy of w starting at coordinate zero; thus n counts anchors, not necessarily the number of exposed letters. Put

$$P_{X,a,F}(\beta) = P_X(a + \beta f_F), \qquad \mu_{X,a,F,\beta} = \mu_{a+\beta f_F}, \qquad \rho_{X,a,F}(\beta) = \int f_F \, d\mu_{X,a,F,\beta} = P'_{X,a,F}(\beta).$$

Positive β rewards occurrences. For several independent couplings replace βf_F by $\sum_{i=1}^{d} \beta_i f_i$; the frequency vector is ∇P. Every such finite-range problem is a finite weighted graph after a higher-block recoding. Primitivity gives unique equilibrium measures and analytic pressure at finite parameters. Increasing a set's own common coupling increases its total frequency; comparisons between different sets or backgrounds are the nontrivial question. Individual cross-responses need not have a fixed sign.

2. Conjugacy, coboundaries, and formal weighted matrices

A conjugacy $h : X \to Y$ is an everywhere continuous invertible map commuting with the shift. A measure-preserving such map is often called a block isomorphism in the Markov-chain literature [1, 2].

Definition 1. For locally constant vector observables $f : X \to \mathbb{Z}^d$ and $g : Y \to \mathbb{Z}^d$, call (X, f) and (Y, g) equivalent if some conjugacy h satisfies

$$f - g \circ h = b - b \circ \sigma \tag{1}$$

for a locally constant $b : X \to \mathbb{Z}^d$. For general real potentials one may also allow an additive constant vector c; keeping $c = 0$ preserves the actual frequency normalization.

Date: Working note, 11 October 2026.

The reason for allowing a coboundary is exact, rather than asymptotic:

$$\sum_{j=0}^{n-1}(f - g \circ h)(\sigma^j x) = b(x) - b(\sigma^n x).$$

Counts differ by a uniformly bounded boundary term, and coincide on periodic orbits. This need not give equality of finite-window count distributions.

Proposition 2 (Gibbs rigidity for a fixed code)**.** *Let $h : X \to Y$ be a conjugacy of mixing SFTs and let f, g be locally constant real scalar potentials. The following are equivalent:*

(1) *$h_*\mu_{\beta_0 f} = \mu_{\beta_0 g}$ for one $\beta_0 \neq 0$;*
(2) *$f - g \circ h = c + b - b \circ \sigma$ for a constant c and locally constant b;*
(3) *$h_*\mu_{\beta f} = \mu_{\beta g}$ for every real β.*

Their pressures then differ by βc and frequencies by c. If their means under the respective measures of maximal entropy agree, $c = 0$. In particular $c = 0$ for equal-size sets of equal-length words on the same uniform full shift.

Proof. Pull the second potential back by h and recode so both potentials are edge functions on one finite graph. For edge $e : i \to j$, the equilibrium transition probability for a potential A is

$$Q_A(e) = \frac{e^{A(e)} r_A(j)}{\lambda_A r_A(i)},$$

where r_A is a positive Perron vector and $\lambda_A = e^{P(A)}$. Equal equilibrium measures give equal conditional probabilities on every edge. Taking logarithms gives $A - B = P(A) - P(B) + b(i) - b(j)$, with $b = \log r_A - \log r_B$. Apply this to $A = \beta_0 f$, $B = \beta_0 g \circ h$ and divide by β_0. Conversely a coboundary has zero integral under every invariant measure, so the variational principle and uniqueness give the transported equilibrium state for every β. Finally h transports the unique measure of maximal entropy. Integrating the cohomology equation against that measure determines c and gives the stated normalization. □

For affine families $a + \beta \cdot f$ and $a' + \beta \cdot g$, the same argument at $d+1$ affinely independent parameter values, using the *same* h, gives cohomology modulo constants separately for the backgrounds and every feature. It therefore transports the entire family. With zero backgrounds, d linearly independent parameter values suffice. This is a finite parameter test for a proposed code, not an algorithm to discover a code.

For a prescribed h, zero-constant cohomology is equivalent to equality of the corresponding sums on every periodic orbit. For edge functions this has an elementary proof: zero sums on directed cycles make the sum along a path from a fixed vertex to another independent of that path. Compare two paths by closing both with a return path. The resulting vertex function is b. Integer edge differences give integer b. Checking a proposed finite-radius code can therefore be reduced to a finite graph calculation.

Encode integer edge labels $f(e) \in \mathbb{Z}^d$ by

$$M_f(\mathbf{t})_{ij} = \sum_{e:i\to j} \mathbf{t}^{f(e)}, \qquad \mathbf{t}^{f(e)} = \prod_{k=1}^{d} t_k^{f_k(e)}.$$

Use the semiring $\mathbb{Z}_{\geq 0}[t_1^{\pm 1}, \ldots, t_d^{\pm 1}]$. Coefficients count distinct edges; exponents record the observable. Evaluating all variables at 1 recovers the underlying adjacency matrix. Negative exponents allow changes of cohomology representative.

Theorem 3 (Classical cocycle/SSE correspondence, specialized)**.** *Conjugacy with the zero-constant relation* (1) *is equivalent to strong shift equivalence of the formal weighted adjacency matrices over $\mathbb{Z}_{\geq 0}[t_1^{\pm 1}, \ldots, t_d^{\pm 1}]$.*

This is the $G = \mathbb{Z}^d$ case of the cocycle classification recorded in [3, Section 5.3]; the state-splitting proof is displayed for finite groups in [4, Proposition 2.7.1], whose authors explicitly note the more general statement. The argument uses finite labeled graphs, and does not require the label group itself to be finite.

Here is the elementary implication, which also explains the Gibbs meaning. If $M = RS$ and $M' = SR$, regard an old edge as a pair (r_i, s_i) and replace the path $\dots(r_i, s_i)(r_{i+1}, s_{i+1})\dots$ by $\dots(s_i, r_{i+1})(s_{i+1}, r_{i+2})\dots$. Bijections between monomial summands and labeled paths are part of this construction. If labels of the two half-edges are $a(r_i)$ and $c(s_i)$, old minus new label is $a(r_i) - a(r_{i+1})$, exactly a coboundary. Composing elementary moves proves one direction. For the converse, use the standard Williams decomposition of a conjugacy into state splittings and their inverses, carrying labels through the splittings. On a common refinement the remaining coboundary is a diagonal monomial similarity $M' = DMD^{-1}$. This is itself an elementary SSE, using $R = MD^{-1}$ and $S = D$. The decomposition theorem is the classical input; we do not claim a new proof of that theorem.

Remark 4 (Keep weights formal). Numerical SSE over $\mathbb{R}_{\geq 0}$ is much too weak. With $0 < p < 1$,

$$\begin{pmatrix} p & 1-p \\ p & 1-p \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} p & 1-p \end{pmatrix}, \qquad \begin{pmatrix} p & 1-p \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = (1).$$

This would identify a Bernoulli full shift with a one-point system. Formal sums of edge weights must not be collapsed by ordinary addition of their numerical values. General backgrounds can instead be encoded in the positive group semiring of the additive group generated by their edge potentials, together with the integer feature labels; see [3, Section 5.4].

3. A positive family theorem and an obstruction

Corollary 5 (Unbordered words)**.** *For equal-length unbordered words on a q-symbol full shift, the Gibbs families are conjugate by a single full-shift automorphism in the following cases:*

- *$q \geq 3$, with no further exclusions;*
- *$q = 2$, with both words outside $\{10^{m-1}, 01^{m-1}, 1^{m-1}0, 0^{m-1}1\}$.*

The code can preserve the occurrence indicators exactly, not merely modulo a coboundary.

Proof. Two unbordered words with no cross-overlaps can be simultaneously interchanged wherever they occur. The occurrences are disjoint; the overlap assumptions prevent new occurrences across a replacement boundary. This gives a bounded radius, shift-commuting involution h with $\mathbf{1}_{[u]} = \mathbf{1}_{[v]} \circ h$. Alphabet permutations have the same property. The chains of precisely these moves in [5, Theorems 4.3 and 4.9] connect the words in the stated classes. Compose the codes, then use Proposition 2, or directly use the variational principle. This is a Gibbs corollary of those existing conjugacies. □

Proposition 6 (Equal pressure and periodic weights do not classify)**.** *The binary word families for $u = 0001$ and $v = 0011$ have equal pressure and equal periodic count polynomials at every parameter, but no topological conjugacy transports their Gibbs measures at any common nonzero coupling.*

Proof. Both words are unbordered of length four. Their weighted de Bruijn matrices satisfy, with z the orbit-length variable and t the occurrence weight,

$$\det(I - zM_u(t)) = \det(I - zM_v(t)) = 1 - 2z + (1-t)z^4.$$

Thus all $\operatorname{tr} M_w(t)^n$ agree, as do their Perron roots and pressures. If a conjugacy transported the Gibbs measures, Proposition 2 would give zero-constant cohomology and therefore preserve zero-count periodic points. For u, the closure of these points is

$$\{0^{\mathbb{Z}}\} \ \cup \ X_{\{000\}}.$$

Indeed a periodic word containing 000 and some 1 necessarily contains 0001. In this closure $0^{\mathbb{Z}}$ is isolated. For v, the avoidance graph is strongly connected and has self-loops, so its periodic points are dense in the mixing SFT $X_{\{0011\}}$, which has no isolated point. These closures cannot be homeomorphic. The determinant and graph assertions are also checked exactly by the accompanying script. □

Classification target. Determine the weighted SSE classes within a fixed nontrivial border class, beginning with the components generated by the existing swap moves. Seek either additional moves or an invariant beyond the weighted zeta function. For example, the source paper singles out 110110 and 011011 as an unresolved avoidance-conjugacy pair; their full word-potential families provide a stronger lifting question. No classification of that pair is asserted here. A coboundary preserves periodic zero counts and minimizing measures; without extra hypotheses it need not send the entire hard-avoidance space pointwise onto the other one, because transient points can have bounded nonzero counts.

Proposition 7 (Explicit bordered marker families)**.** *Fix a symbol c, integers $r, k \geq 1$, and let D consist of the k-letter words which start and end with a symbol other than c and contain no c^r. For $v \in D$ put $w_v = c^r v c^r$. Every permutation π of D is realized by a full-shift automorphism h_π such that*

$$\mathbf{1}_{[w_v]} = \mathbf{1}_{[w_{\pi(v)}]} \circ h_\pi \qquad (v \in D).$$

Thus all these word-potential families are conjugate, and the corresponding formal weighted matrices are strong shift equivalent. More generally h_π conjugates the whole vector-potential family after permuting its couplings. Every w_v has proper border set $\{1, \ldots, r\}$.

Proof. Keep every occurrence of the marker c^r fixed. Between consecutive marker occurrences with a gap of length k and payload $v \in D$, replace the payload by $\pi(v)$. Equivalently, make this replacement at each occurrence of $c^r v c^r$. Distinct selected payloads are disjoint, though their endpoint markers may overlap. No marker lies inside a selected payload. The first and last payload letters are not c, so a replacement creates no marker across its endpoints; the new payload also contains no marker. Consequently the entire marker-position field is unchanged, as are the selected gaps. The inverse rule uses π^{-1}. The rule has finite radius and commutes with the shift, so it is an automorphism and has the claimed pointwise indicator identity. The Gibbs and weighted-SSE conclusions follow from Proposition 2 and Theorem 3. For the border assertion, c^r occurs in w_v only at its two endpoints. A proper border longer than r would force an additional marker occurrence strictly between them, which is impossible. □

In the binary case $c = 1, r = 2, k = 3$, swapping 000 and 010 gives a radius-three involution between 1100011 and 1101011, both with borders $\{1, 2, 7\}$. Explicitly, flip x_i exactly when

$$(x_{i-3}, x_{i-2}, x_{i-1}) = 110, \qquad (x_{i+1}, x_{i+2}, x_{i+3}) = 011.$$

The checker exhausts all 2^{13} input windows, verifying both involutivity and the pointwise occurrence-indicator identities. These words have different letter multiplicities, so this is not a letter-permutation conjugacy. The construction is a concrete marker automorphism, not a classification of all words with these border sets. Any locally constant background depending only on the marker-position field is fixed by the same code, so the result also holds with such a background added to the vector potential.

4. An abstract description of frequency ordering

Fix a finite directed graph, a scalar edge feature $f(e)$, and a background edge potential $a(e)$. A stationary edge flow η is a nonnegative vector with $\sum_e \eta_e = 1$ and equal incoming and outgoing

mass at every vertex. Let η_i be this common mass at vertex i. Define

$$H_{a,f}(r) = \max_{\substack{\eta \text{ stationary flow} \\ \sum_e \eta_e f(e)=r}} \left[-\sum_{e:i\to j} \eta_e \log \frac{\eta_e}{\eta_i} + \sum_e \eta_e a(e) \right].$$

Use $0 \log 0 = 0$. Markovization maximizes entropy at a prescribed edge flow, so the finite-graph variational principle gives

$$P(\beta) = \max_r \{H_{a,f}(r) + \beta r\}. \tag{2}$$

The same construction gives a vector profile for several marked patterns. This is standard thermodynamic/exponential-family structure, not a new classification result; see also [6].

Where the scalar profiles are differentiable and strictly concave, the inverse frequency curve is $\beta(r) = -H'(r)$. For two models with the same attainable frequency interval, frequency order for every common coupling is equivalent to derivative order of these profiles, or to monotonicity of their difference. Specifically $\rho_B(\beta) \geq \rho_A(\beta)$ for all β is equivalent to $H'_B(r) \geq H'_A(r)$ throughout the interval. One-sided versions require the corresponding common frequency ranges. Pointwise profile or pressure order alone is weaker. Conjugacy with zero-constant cohomology preserves the profile.

There is also a useful finite classification fact. For fixed integer features and fixed rational positive background edge weights, the transfer matrix has Laurent-polynomial entries in $t = e^\beta$. Its Perron root and frequency are algebraic functions of t. Two frequency curves therefore either coincide identically or meet at finitely many positive values of t. One proof is that their graphs and equality set are semialgebraic, while analyticity excludes a nontrivial interval of equality unless they agree everywhere. With several couplings the equality loci are semialgebraic sets in the positive fugacity orthant. This gives finite exact comparison problems, not an efficient general algorithm for conjugacy.

Theorem 8 (No universal crossing bound for finite-range backgrounds)**.** *For every $N \geq 1$ there are two locally constant backgrounds a_A, a_B on the binary full shift, with positive rational local weights, such that for the single fixed observable $f = \mathbf{1}_{[1]}$,*

$$\rho_{a_A,f}(0) = \rho_{a_B,f}(0) = \tfrac{1}{2},$$

but $\rho_{a_A,f} - \rho_{a_B,f}$ has at least N sign reversals on $(0,\infty)$. The backgrounds can both be invariant under global bit complement. Thus matching reference frequencies does not give a universal crossing bound, even with one fixed alphabet, shift, and counted word.

Proof. Put $L = 2(N+2)$, $k_i = L/2 + i$ for $0 \leq i \leq N+1$, and $v_i = 1^{k_i}0^{L-k_i}$. In the binary de Bruijn graph on L-letter states, use the periodic cycles of each v_i and its complement. Each cycle has length L: its cyclic word has exactly two changes of symbol, precluding a nontrivial repetition. Distinct weights k_i give disjoint cycles. The central cycle $i = 0$ equals its complementary cycle; all other pairs are disjoint. The full de Bruijn graph has $D = 2^L$ vertices.

On each cycle of pair $i > 0$, put base background weight 2^{α_i} on the edge starting at v_i or its complement, and weight one on its other edges. Give the central cycle weight one on every edge. Other de Bruijn edges initially have weight zero. Tilting the edge out of state u by $e^{\beta u_0}$ counts the symbol 1. The two cycle pressures are

$$\frac{\alpha_i \log 2 + k_i\beta}{L}, \qquad \frac{\alpha_i \log 2 + (L-k_i)\beta}{L}.$$

Choose the integers α_i separately for the two backgrounds, as follows. Let $B_0 = 4(N+2)^2$. In background A, select indices 0 and the odd positive indices; in background B, select 0 and the even positive indices. Set $\alpha_0 = 0$. For successive selected indices $i' < i$, set

$$\alpha_i = \alpha_{i'} - (k_i - k_{i'})(2i-1).$$

Give unselected indices value $-B_0$. The selected exponents have absolute value at most $(N+1)(2N+1) < B_0$. Consecutive selected positive-slope lines intersect at $(2i-1)\log 2$, in increasing order.

Complementary-slope lines cannot exceed the central line $\beta/2$ for $\beta > 0$. Unselected lines stay below it throughout the interval used below, since $B_0 > (N+1)(2N+5/2)$.

At $\beta_j = 2j \log 2$, $1 \le j \le N+1$, the maximum of the cycle pressures therefore has slope k_j/L in one background and k_{j-1}/L in the other. Background A has the larger slope exactly when j is odd. Each maximum remains affine on $[\beta_j - h, \beta_j + h]$, where $h = (\log 2)/2$.

Now add the same positive rational ε to the background weight of every de Bruijn edge, including those already in cycles. These weights define finite-range backgrounds on the binary full shift: a local weight is attached to each $(L+1)$-letter word. Their transfer matrices are

$$M_\varepsilon(\beta) = M_0(\beta) + \varepsilon \operatorname{diag}(e^{\beta u_0}) A_L,$$

where A_L is the unweighted de Bruijn adjacency matrix. They are primitive. Complementing every state conjugates $M_\varepsilon(\beta)$ to $e^\beta M_\varepsilon(-\beta)$. Thus each pressure satisfies $P(\beta) = \beta + P(-\beta)$ and each reference frequency is exactly $1/2$.

For an explicit finite choice, set

$$T = 2N + 3, \quad E = 2L(B_0 + T) + T + 1 + \lceil \log_2(64LD) \rceil, \quad \varepsilon = 2^{-E}.$$

For $0 \le \beta \le (2N + 5/2) \log 2$, each nonzero base edge weight lies in $[2^{-B_0}, 2^T]$. Normalize a positive Perron vector on each cycle by setting one coordinate to one. Successive coordinates have ratio equal to the cycle Perron root divided by an edge weight, bounded between 2^{-B_0-T} and 2^{B_0+T}. Set coordinates at unused states to one. The resulting vector $z > 0$ has every coordinate between R^{-1} and R, where $R = 2^{L(B_0+T)}$. If λ_0 is the largest cycle Perron root, then $M_0 z \le \lambda_0 z$, $\lambda_0 \ge 1$, and the Perron comparison bound gives

$$0 \le \lambda_\varepsilon - \lambda_0 \le \varepsilon 2^{T+1} D R^2 \le \frac{1}{64L}.$$

Consequently the perturbed pressure lies between the cycle-pressure maximum and that maximum plus $1/(64L)$. Convex secant bounds at $\beta_j \pm h$ show that its derivative at β_j differs from the corresponding cycle slope by at most $1/(64Lh) < 1/(16L)$, since $\log 2 > 1/2$. The limiting slopes differ by $1/L$, so the perturbed differences still alternate sign at the $N+1$ sample couplings. Continuity gives at least N sign reversals. All local background weights are positive and rational. □

For $N = 3$, take $L = 10$, $D = 1024$, and $\varepsilon = 2^{-2210}$. The backgrounds depend on eleven consecutive bits. At $\beta_j = 2j \log 2$, $1 \le j \le 4$, the frequency pairs are within $1/160$ of

$$(3/5, 1/2), \quad (3/5, 7/10), \quad (4/5, 7/10), \quad (4/5, 9/10).$$

These are bounds for the displayed finite rational model. The checker records its background weights and exact rational frequency intervals. It also checks the cycle disjointness, parameter formula, and affine envelopes for $1 \le N \le 12$; the proof holds for arbitrary N.

5. A set-of-words extension of the convex-order mechanism

Let $F \subset \mathcal{A}^m$ have k words, under uniform iid reference letters. For $1 \le d < m$ suppose

$$c_F(d) := \#\{v \in F : u_{d+1} \cdots u_m = v_1 \cdots v_{m-d}\}$$

is independent of the choice of $u \in F$. This is a concrete condition on the row sums of the cross-overlap matrices. Transitivity of an alphabet-permutation group on F is one sufficient symmetry condition.

Proposition 9 (Common return law for a marked set)**.** *Suppose $F, G \subset \mathcal{A}^m$ have the same size and both satisfy this row-sum condition. If $c_F(d) \le c_G(d)$ for every $1 \le d < m$, then their stationary reference occurrence counts in every finite anchor window satisfy $N_F \le_{\mathrm{cx}} N_G$. Their reference return gaps are also in convex order. Consequently their finite-volume partition functions are ordered for both signs of coupling. If the profiles differ, the limiting pressures are strictly ordered at every finite nonzero coupling. No frequency order under Gibbs tilting is asserted by this proposition.*

Proof. Conditional on any marked word just ending, the all-return probability to the set after d steps is $c_F(d)q^{-d}$ for $d < m$, and kq^{-m} for $d \geq m$. It is independent of the current marked word. The first-return decomposition, inductively in time, then shows that the first return has the same law from every marked word. The strong Markov property makes successive gaps iid.

Let $U_F(z)$ be the all-return generating function (constant term one), $g_F(z) = \mathbb{E}z^{R_F}$ the first-return generating function, and $S_F(z) = (1 - g_F(z))/(1 - z)$ its survival series. Then $U_F = (1 - g_F)^{-1}$, and

$$\frac{g_G - g_F}{(1-z)^2} = (U_G - U_F)S_F S_G$$

has nonnegative coefficients. The means are both q^m/k by stationary renewal frequency. The integrated-tail characterization gives $R_F \leq_{\mathrm{cx}} R_G$. For a window of L anchors, counting an occurrence and its jth successor gives

$$\mathbb{E}(N_F - j)_+ = kq^{-m}\mathbb{E}(L - R_{F,1} - \cdots - R_{F,j})_+.$$

Independent sums preserve convex order, and the counts have equal means. This proves the count assertion. Apply the convex exponential function for finite volume and the return normalization $\mathbb{E}e^{\beta - \psi(\beta)R} = 1$ for the strict pressure assertion, as in the existing single-word renewal proof. □

Proposition 10 (Translation orbits reduce to one word)**.** *Identify the alphabet with $\mathbb{Z}/q\mathbb{Z}$. For w of length $m \geq 2$, let $F_w = \{w + c : c \in \mathbb{Z}/q\mathbb{Z}\}$ and write*

$$(Dx)_i = x_{i+1} - x_i, \qquad \partial w = (w_1 - w_0, \ldots, w_{m-1} - w_{m-2}).$$

Then $f_{F_w} = \mathbf{1}_{[\partial w]} \circ D$. For finite strings of length $n \geq m$, the occurrence-count partition functions satisfy

$$Z_{n,F_w}(t) = qZ_{n-1,\partial w}(t).$$

Consequently $P_{F_w}(\beta) = P_{\partial w}(\beta)$ and $\rho_{F_w}(\beta) = \rho_{\partial w}(\beta)$ at every coupling. All the single-word frequency-ordering theorems below therefore apply to these q-word sets, using the border data of their increment words. The Gibbs measure for F_w is the uniform conditional lift over the q fibres of D of the Gibbs measure for ∂w.

Proof. Two words have the same increments exactly when they differ by a constant translation. Every increment string has q preimages, proving the finite identity and both limiting identities. In the coordinates $(y, s) = (Dx, x_0)$, the shift is $(y, s) \mapsto (\sigma y, s + y_0)$. Uniform conditional measure on s gives an invariant lift of every invariant base measure. The entropy rates agree, since specifying s contributes at most $\log q$ to the entropy of any block. The variational principle and uniqueness of the finite-range equilibrium measure prove the last assertion. □

Proposition 11 (Lifting an increment conjugacy)**.** *Let h be an automorphism of the increment full shift. There exists an automorphism H with $DH = hD$ commuting with global symbol translations if and only if*

$$(hy)_0 - y_0 = b(\sigma y) - b(y) \quad \text{in } \mathbb{Z}/q\mathbb{Z} \tag{3}$$

for a locally constant b. Equivalently, h preserves the sum of the symbols modulo q on every periodic orbit. If h transports a word-potential Gibbs family, such a lift transports the associated translation-orbit family. For $q = 2$, every automorphism lifting h automatically commutes with the global complement, so the criterion characterizes all lifts over h.

Proof. In the preceding coordinates an equivariant lift necessarily has the form $(y, s) \mapsto (hy, s + b(y))$. Commuting with the skew shift is exactly (3); this form also proves invertibility. A continuous finite-valued function on the full shift is locally constant. Telescoping gives the periodic condition. Conversely recode the difference in (3) as edge labels in $\mathbb{Z}/q\mathbb{Z}$. Zero sums on all closed walks make sums along directed paths independent of the path, after comparison with a common return path. These sums define the required vertex function b. The measure assertion follows from the uniform

conditional lifts in Proposition 10. On a two-element fibre every permutation commutes with its transposition, proving the final assertion. □

For instance $\{00, 11\}$ and $\{01, 10\}$ have the identical pressure $\log(1+e^\beta)$ and frequency $e^\beta/(1+e^\beta)$, but their Gibbs families are not topologically conjugate at any common $\beta \neq 0$. Their equal uniform means force the constant in Proposition 2 to vanish, whereas their observables equal respectively one and zero on every fixed point. The base letter complement exchanges their increment potentials but violates (3) on fixed points. The formal transfer matrices exhibit the lost information directly:

$$\begin{pmatrix} t & 1 \\ 1 & t \end{pmatrix}, \qquad \begin{pmatrix} 1 & t \\ t & 1 \end{pmatrix}.$$

Their Perron roots agree, but their traces $2t$ and 2 preclude formal SSE.

Marker permutations in Proposition 7 do lift whenever they preserve each payload's symbol sum modulo q. Integrate the changes in the increments across each selected gap, starting with zero before it. The total change over the gap is zero, so the resulting b vanishes outside the gap and is a finite-radius function satisfying (3). A binary example swaps increment payloads 00000 and 01010 between markers 11. Its radius-five lift conjugates the complement-pair potentials

$$\{0100000010, 1011111101\}, \qquad \{0100110010, 1011001101\}.$$

At an occurrence of any of these four ten-letter words, toggle positions four and five, indexed from zero. The pairs of positions selected at distinct occurrences are disjoint. The checker verifies the involution and both set-indicator identities on all 2^{21} radius-ten input windows, and $DH = hD$ on all 2^{12} relevant input windows, using integer arithmetic.

In matrix terms one retains both the integer occurrence label and the increment label modulo q. Their formal semiring is $\mathbb{Z}_{\geq 0}[\mathbb{Z} \times (\mathbb{Z}/q\mathbb{Z})]$, equivalently monomials $t^n z^c$ with $z^q = 1$. The classical cocycle/SSE theorem over an abelian group [3, Section 5.3] classifies base conjugacies preserving both labels up to coboundaries. This is the matrix version of retaining the Gibbs observable together with the lifting condition.

6. Single-word frequency ordering

Return to a single word of length m on the uniform q-shift. Let $\mathcal{B}(w)$ contain all nonempty borders including m, and put $C_w(q) = \sum_{b \in \mathcal{B}(w)} q^{b-1}$. Define its least period $p_w = m - b_w$, where b_w is its longest proper border, or zero if unbordered. Write $\rho_w(\beta) = \mu_{\beta \mathbf{1}_{[w]}}([w])$.

The existing pressure theorem gives $P_v > P_u$ away from zero if $C_u(q) < C_v(q)$, and its quadratic expansion gives $\rho_v - \rho_u > 0$ for small positive coupling and the reverse sign for small negative coupling. A global positive-coupling density order is false.

Proposition 12 (Large positive coupling)**.** *For every word, $P_w(\beta) = \beta/p_w + o(1)$ and $\rho_w(\beta) \to 1/p_w$ as $\beta \to +\infty$. If $C_u(q) < C_v(q)$ and $p_u = p_v = p$, let $b = m - p$ be the common longest proper border and $c < b$ the largest differing border. Put $d = b - c$. Then*

$$P_v(\beta) - P_u(\beta) = \frac{1}{p} e^{-d\beta/p} + O(e^{-(d+1)\beta/p}),$$

$$\rho_v(\beta) - \rho_u(\beta) = -\frac{d}{p^2} e^{-d\beta/p} + O(e^{-(d+1)\beta/p}).$$

In particular every such pair has at least one positive-coupling frequency crossing, despite its fixed pressure order.

Proof. Let $D_w(x) = \sum_{b \in \mathcal{B}(w),\, b<m} x^{b-1}$ and $t = e^\beta$. The characteristic equation from the existing pressure proof is

$$(x - q)x^{m-1} = (t - 1)[1 + (x - q)D_w(x)].$$

For the Perron root at large t, this gives $t = x^{p_w}(1 + O(x^{-1}))$; the same statement holds with $p_w = m$ if $D_w = 0$. After setting $x = t^{1/p_w} y$, the equation has a simple limiting root $y = 1$. The implicit-function theorem gives an analytic expansion in t^{-1/p_w}, which may be differentiated.

For the pair with common b, $D_v(x) - D_u(x) = x^{c-1} + O(x^{c-2})$, with the remainder zero when $c = 1$. The logarithms of the two expressions for t therefore differ, at a common large x, by $\log t_u(x) - \log t_v(x) = x^{-d} + O(x^{-d-1})$. Their derivatives with respect to $\log x$ are $p + O(x^{-1})$. Inverting at the same t yields the first displayed difference. Differentiate the convergent local expansions to obtain the second. The known sign near zero and the eventual negative sign force a crossing. □

For $u = 01001$, $v = 00100$, the border sets are $\{2, 5\}$ and $\{1, 2, 5\}$. Both least periods are three. The exact characteristic factors are

$$F_u(x, t) = (x - 2)x^4 - (t - 1)(x - 1)^2, \qquad F_v(x, t) = (x - 2)x^4 - (t - 1)(x^2 - x - 1).$$

At the largest positive root λ, the frequency is $-tF_t(\lambda, t)/(\lambda F_x(\lambda, t))$. Exact rational root isolation and interval evaluation certify

t	$\rho_u(\log t)$	$\rho_v(\log t)$
2	$(0.060284386, 0.060284387)$	$(0.066836786, 0.066836787)$
64	$(0.307729696, 0.307729697)$	$(0.300579885, 0.300579886)$.

Thus even proper border inclusion does not rescue global positive density ordering. This example is separate from the older, unreproduced 000/010 fixed-sign claim in the historical derivative note.

Theorem 13 (Penalized occurrences)**.** *For equal-length words on the uniform full shift, $C_u(q) < C_v(q)$ implies $\rho_u(\beta) > \rho_v(\beta)$ for every $\beta < 0$. Equal border polynomials give equal frequencies at every coupling.*

Proof. The strict case requires $m \geq 2$. Put $L = m - 1$, $t = e^{\beta} \in (0, 1)$, $s = 1 - t$, and $\Delta = D_v - D_u$. Its leading coefficient is one and its other coefficients belong to $\{-1, 0, 1\}$. Interpolate $D_\theta = D_u + \theta\Delta$, $0 \leq \theta \leq 1$. All coefficients of D_θ lie in $[0, 1]$. Let r be the largest solution of $(q - r)r^L = 1$. There is a unique $x = x_\theta(t) \in (r, q)$ satisfying

$$\frac{1}{s} = x^{-L}\left(\frac{1}{q - x} - D_\theta(x)\right). \tag{4}$$

Indeed, at r the right side is at most one, it tends to infinity at q, and it is strictly increasing on (r, q): the derivative of $x^{-L}/(q - x)$ is positive there, whereas each term of $x^{-L}D_\theta$ is decreasing. At $\theta = 0, 1$ these roots are the Perron roots, by the characteristic equation and the largest-positive-root characterization. Write $y = q - x$. Throughout this interval,

$$yx^L < 1, \qquad x > Ly. \tag{5}$$

For $q = 2$ and $L \geq 2$ we also have $x > r \geq \varphi = (1 + \sqrt{5})/2$; for $q \geq 3$ we have $x > q - 1 \geq 2$. These follow directly from the equation defining its largest root; for $q = 2$, it is the multinacci root already used in the pressure comparison.

Implicit differentiation of (4) gives

$$\partial_\theta \log x = \frac{\Delta(x)}{T_\theta(x)}, \qquad T_\theta(x) = \frac{x}{y^2} - \frac{L}{y} + H_\theta(x), \qquad H_\theta = LD_\theta - xD'_\theta. \tag{6}$$

Both Δ and T_θ are positive on this interval. The positivity of Δ is the first-differing-coefficient bound from the pressure proof, and $T_\theta > 0$ also follows from $x > Ly$ and $H_\theta \geq 0$. We show the two logarithmic derivative bounds

$$\frac{\Delta'}{\Delta} < \frac{1}{y} < \frac{T'_\theta}{T_\theta}. \tag{7}$$

For the first, put $k = \deg \Delta \le L-1$ and $A_k(x) = x^k - \sum_{j=0}^{k-1} x^j$ (so $A_0 = 1$). Since $x - jy > 0$ for $0 \le j \le k$, minimizing the coefficients of $\Delta - y\Delta'$ gives

$$\Delta - y\Delta' \ge A_k - yA_k'.$$

For $q = 2$, direct differentiation, using $y = 2 - x$, gives for $k \ge 1$

$$A_k - yA_k' = \frac{1 - yx^{k+1} + ky^2(x-1)x^{k-1}}{(x-1)^2} > 0,$$

by $yx^{k+1} \le yx^L < 1$; $k = 0$ is immediate. For $q \ge 3$, $x > 2$ and $A_k = [1 + (x-2)x^k]/(x-1) > 0$. If $k \ge 1$, then

$$\frac{A_k'}{A_k} = \frac{x^k + k(x-2)x^{k-1}}{1 + (x-2)x^k} - \frac{1}{x-1} < x^k + \frac{k}{x} \le x^{k+1} \le x^L < \frac{1}{y}.$$

Here $k \le (x-1)x^{k+1}$ for $x \ge 2$, since $k \le 2^{k+1}$. This proves the first bound, again with $k = 0$ immediate.

For the second, write $D_\theta = \sum_{j=0}^{L-1} d_j x^j$ with $0 \le d_j \le 1$. Then $H_\theta = \sum_j (L-j) d_j x^j$ and

$$T_\theta' - \frac{T_\theta}{y} = \frac{q}{y^3} + H_\theta' - \frac{H_\theta}{y} \ge \frac{q - y^2 H_\theta}{y^3}.$$

It remains to bound $y^2 H_\theta < q$. For $x > 1$,

$$H_\theta \le \sum_{j=0}^{L-1} (L-j)x^j < \frac{x^{L+1}}{(x-1)^2}, \qquad y^2 H_\theta < \frac{yx}{(x-1)^2}.$$

For $q \ge 3$, $x > 2$ and $y < x^{-L} \le x^{-1}$ make the last bound at most one. For $q = 2$, $L \ge 2$, it is $x(2-x)/(x-1)^2 \le \varphi < 2$, since $x \ge \varphi$. Finally, if $q = 2$, $L = 1$, use $H_\theta \le 1$ and $y < 1$ directly. This proves (7) in every case.

Consequently Δ / T_θ is strictly decreasing in x. Equation (4) shows that $x_\theta(e^\beta)$ is strictly increasing in β. Differentiating (6) therefore gives

$$\partial_\theta \partial_\beta \log x_\theta(e^\beta) < 0.$$

Integrating over $\theta \in [0,1]$ proves $\rho_v < \rho_u$. All differentiations are justified by the strictly positive derivative in (4). Equality for identical border polynomials follows from their identical characteristic equations. □

The proof uses only the binary coefficients of the border polynomials, not their realizability as word borders. In particular it supplies the same negative-coupling derivative order for the interpolated algebraic root curves.

Theorem 14 (Different least periods)**.** *If $p_v < p_u$, then $\rho_v(\beta) > \rho_u(\beta)$ for every $\beta > 0$.*

Proof. Put $L = m - 1$, $t = e^\beta > 1$, and, for $x > q$, set

$$K_w(x) = D_w(x) + \frac{1}{x - q}, \qquad E_w(x) = \frac{xK_w'(x)}{K_w(x)}.$$

The characteristic equation reads $t - 1 = x^L / K_w(x)$. At its Perron root $x = \lambda_w(t)$, logarithmic differentiation therefore gives

$$\rho_w(\log t) = \frac{t}{t-1} \frac{1}{L - E_w(\lambda_w(t))}. \tag{8}$$

The denominator is positive. Moreover E_w is strictly increasing on (q,∞): as a function of $r=\log x$, $K_w(e^r)$ is a positive sum of exponentials, since

$$\frac{1}{x-q}=\sum_{n\ge 1} q^{n-1}x^{-n}.$$

The second derivative of its logarithm is the strictly positive variance of the exponents under the normalized summand weights.

We next compare E_v and E_u at the same $x>q\ge 2$. If a nonzero binary-coefficient polynomial D has degree k, its logarithmic derivative satisfies

$$k-1<\frac{xD'(x)}{D(x)}\le k \qquad (x\ge 2). \tag{9}$$

For the lower bound, subtract $(k-1)D$ from xD' and use

$$xD'-(k-1)D\ge x^k-\sum_{j=0}^{k-2}(k-1-j)x^j>0.$$

Indeed, dividing the sum by x^k and extending it to an infinite series gives $\sum_{r\ge 2}(r-1)x^{-r}=1/(x-1)^2\le 1$; the finite sum is strictly smaller. The cases $k=0,1$ follow from the same empty-sum convention.

The hypothesis means $\deg D_v>\deg D_u$, with $\deg 0=-1$. If $D_u\neq 0$, (9) shows that the logarithmic derivative of D_v strictly exceeds that of D_u, while the coefficient comparison gives $D_v(x)>D_u(x)$. Adding the common positive term $U=1/(x-q)$ preserves this derivative order: $xU'/U=-x/(x-q)<0$ is smaller than either polynomial derivative, and D_v receives both the larger weight and the larger derivative in the weighted average for D_v+U. Thus $E_v(x)>E_u(x)$. If $D_u=0$, the same conclusion follows directly by adding D_v to U.

Finally the pressure order gives $\lambda_v(t)>\lambda_u(t)>q$. Together with the increasing functions just proved, this implies $E_v(\lambda_v(t))>E_u(\lambda_u(t))$. Equation (8) now yields $\rho_v>\rho_u$. □

The next two criteria reduce uniqueness of a frequency crossing to a variance comparison. The binary theorem below shows that, for word border profiles, these criteria together are exhaustive.

Theorem 15 (An exact criterion for a unique crossing)**.** *For a word w of length m, put $\mathcal{P}_w=\{m-b : b\in\mathcal{B}(w),\ b<m\}$. Suppose $C_u(q)<C_v(q)$ and $p_u=p_v$. Put $J=\mathcal{P}_v\setminus\mathcal{P}_u$, $a=\min J$, and let $a-\ell$, $a+j$ be the nearest exponents on either side of a in $\mathcal{P}_u\cup\{m,m+1,\dots\}$. Thus $\ell,j\ge 1$. If*

$$\sum_{n\in J-a}(n+\ell)(n-j)q^{-n}\le 0, \tag{10}$$

then there is exactly one positive frequency crossing, from $\rho_v>\rho_u$ to $\rho_v<\rho_u$. The same conclusion follows whenever the variance comparison in the proof can be established by another argument.

Proof. For $r>\log q$ define the strictly decreasing series

$$A_w(r)=\sum_{p\in\mathcal{P}_w}e^{-pr}+\sum_{n\ge m}q^{n-m}e^{-nr}.$$

The characteristic equation gives $A_w(P_w(\beta))=1/(e^\beta-1)$ for $\beta>0$. Parametrize the pair of roots by $r=P_v(\beta)$, and write $\delta(r)=P_v(\beta)-P_u(\beta)>0$. It satisfies

$$A_u(r-\delta(r))=A_v(r), \qquad r-\delta(r)>\log q.$$

By Proposition 12, δ tends to zero at both ends of $(\log q,\infty)$. Consequently it has a positive interior maximum.

Let $I = \mathcal{P}_u \cap \mathcal{P}_v$ and $K = \mathcal{P}_u \setminus \mathcal{P}_v$. For a fixed $d > 0$ set

$$Q_d(r) = \sum_{p\in I}(e^{pd}-1)e^{-pr} + \sum_{p\in K} e^{pd}e^{-pr} + \sum_{n\ge m} q^{n-m}(e^{nd}-1)e^{-nr},$$
$$B(r) = \sum_{p\in J} e^{-pr}.$$

Then $Q_{\delta(r)}(r)/B(r) = 1$, and its logarithm has strictly positive partial derivative with respect to d. All terms of Q_d are positive; its exponent support is $\mathcal{P}_u \cup \{m, m+1, \dots\}$. All series and their derivatives converge for $r > \log q + d$.

Normalize the summands of Q_d and B to define random exponents S, T. At a stationary point in r of $\log(Q_d/B)$ their means agree, say at $a + \mu$. For $x = e^r > q \ge 2$, $T - a$ has probabilities proportional to x^{-n} on the finite set $J - a$, which contains zero. Moreover $0 \le \mu < 1$:

$$\sum_{n\in J-a}(1-n)x^{-n} \ge 1 - \sum_{n\ge 2}(n-1)x^{-n} = 1 - \frac{1}{(x-1)^2} > 0.$$

The support of $S - a$ lies outside $(-\ell, j)$. Therefore

$$\mathrm{Var}(S) \ge (\mu + \ell)(j - \mu).$$

Let $V(x) = \sum_{n\in J-a}(n+\ell)(n-j)x^{-n}$. The function $x^j V(x)$ is strictly decreasing: its terms with $n < j$ are negative increasing powers of x, its terms with $n > j$ are positive decreasing powers, and its constant-index term is $-\ell j x^j$. Thus (10) implies $V(x) < 0$ for $x > q$. Dividing by the normalizing sum gives

$$\mathrm{Var}(T) < (\mu + \ell)(j - \mu) \le \mathrm{Var}(S).$$

Thus $\partial_r^2 \log(Q_d/B) > 0$ at every such stationary point.

Differentiate $\log(Q_{\delta(r)}(r)/B(r)) = 0$. At any point where $\delta'(r) = 0$, the preceding inequality and the positive d derivative give $\delta''(r) < 0$. Hence every critical point of δ is a strict local maximum; there can be only one, since two maxima would force an intervening minimum. The endpoint limits then give positive derivative before that maximum and negative derivative afterwards. Finally

$$\rho_v(\beta) - \rho_u(\beta) = \delta'(P_v(\beta))\,\rho_v(\beta),$$

and $\rho_v(\beta) > 0$, proving the claimed unique crossing and its signs. □

Corollary 16 (All larger alphabets and explicit binary cases)**.** *If $q \ge 3$, every pair with $C_u(q) < C_v(q)$ and $p_u = p_v$ has exactly one positive frequency crossing. For $q = 2$, the same conclusion holds if there are at most two new borders, or if all new overlap periods lie in one interval containing no old overlap period.*

Proof. It suffices to use the weaker lower gap $\ell = 1$ in the variance argument. For $q \ge 3$, omit the nonpositive terms with $0 < n < j$ and extend the positive tail to get

$$\sum_{n\in J-a}(n+1)(n-j)q^{-n} \le -j + q^{-j}\left[\frac{q(q+1)}{(q-1)^3} + (j+1)\frac{q}{(q-1)^2}\right]$$
$$\le -j + \frac{3(j+3)}{4\cdot 3^j} \le 0.$$

The last inequality holds for every $j \ge 1$, with equality possible in the bound only at $j = 1$. For at most two new borders, the only nontrivial case is $J - a = \{0, k\}$, $k > j \ge 1$. Then

$$-j + (k+1)(k-j)q^{-k} \le -j + j(k^2-1)2^{-k} \le 0,$$

since $k^2 - 1 \le 2^k$ for every integer $k \ge 2$. If the new periods lie in one gap of the old ones, every $n \in J - a$ satisfies $n < j$, so the certificate is strictly negative directly. □

Proposition 17 (Using the mean of the common periods)**.** *Retain the notation of Theorem 15, put* $I = \mathcal{P}_u \cap \mathcal{P}_v$, $K = \mathcal{P}_u \setminus \mathcal{P}_v$, *and define*

$$M_q = \frac{\sum_{p\in I} p^2 q^{-p}}{\sum_{p\in I} pq^{-p}}, \qquad \ell = a - M_q, \qquad j = \min(K \cup \{m\}) - a.$$

If $\ell > 0$ *and*

$$\sum_{n\in J-a} (n+\ell)(n-j)q^{-n} \le 0, \tag{11}$$

then there is exactly one positive frequency crossing. In particular this holds if $\mathcal{P}_u \subsetneq \mathcal{P}_v$ *and* $M_q < a$.

Proof. The common least period belongs to I, so M_q is defined. Lexicographic order gives $\min K > a$ when K is nonempty, hence $j \ge 1$. Split the exponent law S in Theorem 15 into its I part and its K-plus-tail part. In the first part the weights are

$$(e^{pd}-1)e^{-pr} = pe^{-p(r-d)}\frac{1-e^{-pd}}{p}.$$

The last factor decreases with p. The covariance inequality for oppositely ordered functions, followed by $r - d > \log q$, bounds the conditional mean of this part by M_q. The second part has conditional mean at least $a + j$. At a stationary ratio the overall mean is $a + \mu$, where $0 \le \mu < 1$. The variance between the two parts therefore yields

$$\operatorname{Var}(S) \ge (\mu + \ell)(j - \mu).$$

Indeed, if the two conditional means are L, H, their between-part variance is $(a+\mu-L)(H-a-\mu)$, and $L \le M_q < a \le a + \mu < a + j \le H$. The proof of Theorem 15 now applies verbatim: it used only $\ell > 0$, not that ℓ was an integer. In the inclusion case $j = m - a$, and every $n \in J - a$ satisfies $0 \le n < j$, so the certificate is strictly negative. □

Lemma 18 (Two bounds for size-biased geometric weights)**.** *For a finite set* $S \subset \mathbb{N}$ *with* $p = \min S \ge 2$,

$$\frac{\sum_{s\in S} s^2 2^{-s}}{\sum_{s\in S} s2^{-s}} < p + 2. \tag{12}$$

If z *is a nonconstant word,* R *its proper positive periods, and* $L \ge 5$, *then*

$$\sum_{n\in R} (n-1)(L+n)2^{-n} < L. \tag{13}$$

Proof. For the first assertion retain the negative term at p, omit any other negative term, and extend the positive tail:

$$\sum_{s\in S}(s-p-2)s2^{p-s} \le -2p + \sum_{k\ge 3}(k-2)(p+k)2^{-k} = \frac{5-3p}{2} < 0.$$

For the second put $h = |z|$ and $f_L(n) = (n-1)(L+n)2^{-n}$. If R is empty there is nothing to prove. Otherwise let $r = \min R \ge 2$. We use Fine–Wilf: periods r, s with $h \ge r + s - \gcd(r, s)$ force $\gcd(r, s)$ to be a period; see [7, Theorem 2.5]. If $r = 2$, nonconstancy forces alternating letters, and all periods are even. If $r = 3$ and $h \ge 7$, periods $4, 5$ are impossible; for $h \le 6$ the largest possible upper sum is $f_L(3) + f_L(4)$ (at $h = 6$, period 4 is impossible and $f_L(5) < f_L(4)$). If $r = 4$ and $h \ge 8$, period 5 is impossible; for $h \le 7$ only $4, 5, 6$ can contribute. For $r \ge 5$, bound by the full tail starting at five.

Summing these bounds gives

case	upper bound for $\sum_{n\in R} f_L(n)$
$r=2$	$5L/9+56/27$
$r=3,\ h\le 6$	$7L/16+3/2$
$r=3,\ h\ge 7$	$7L/16+17/8$
$r=4,\ h\le 7$	$25L/64+59/32$
$r=4,\ h\ge 8$	$3L/8+17/8$
$r\ge 5$	$5L/16+2.$

Every displayed affine function is strictly smaller than L for $L\ge 5$. □

Lemma 19 (Least periods three and four)**.** *If a word z has least period $k\in\{3,4\}$ with $k<|z|$, and proper period set R, then*

$$\sum_{s\in R} s(s-k-1)2^{k-s}<0.$$

Proof. Put $d=|z|$. Fine–Wilf shows that every nonmultiple of k in R is at least $d-k+2$. Multiples of k contribute at most

$$\sum_{j\ge 1} kj(kj-k-1)2^{k-kj}=\begin{cases}-192/343, & k=3,\\ -7168/3375, & k=4.\end{cases}$$

For $k=3$ and $d\ge 11$, the only possible additional period is $d-1\ge 10$, whose contribution is at most $15/32$. For $k=4$ and $d\ge 12$, the two possible additional periods contribute at most $83/64$. The functions $s(s-4)2^{3-s}$ and $s(s-5)2^{4-s}$ decrease respectively for $s\ge 6$ and $s\ge 7$, which justifies these uniform bounds. Both resulting totals are negative. For the remaining lengths, include all multiples of k below d and all integers from $\max(k+1,d-k+2)$ to $d-1$, counting each once. The resulting upper bounds are

k	d	successive upper bounds
3	$4,\dots,10$	$-3,-3,-7/4,-3/2,-3/16,-1/2,-51/64$
4	$5,\dots,11$	$-4,-4,-5/2,-3/4,-3/4,-11/8,-19/32.$

□

Theorem 20 (Binary single crossing)**.** *For equal-length binary words with $C_u(2)<C_v(2)$ and $p_u=p_v$, there is exactly one positive coupling β_* with $\rho_u(\beta_*)=\rho_v(\beta_*)$. The difference $\rho_v-\rho_u$ is positive before β_* and negative afterwards. Consequently this single-crossing conclusion holds for every alphabet size $q\ge 2$.*

Proof. Put $p=p_u=p_v$. Distinct border data exclude $p=1$, so $p\ge 2$. Use a,J from Theorem 15. We separate inclusion and noninclusion of the proper period sets.

Suppose first $\mathcal{P}_u\subsetneq\mathcal{P}_v$. If $a\ge p+2$, (12) gives $M_2<p+2\le a$, so Proposition 17 applies. If $a=p+1$, the prefix and suffix of v of length $b=m-p$ are constant, because their period is one. They cannot cover v: otherwise v would be constant and $p=1$. Thus $p>b$. The prefix z of u of length b is nonconstant, since $p+1\notin\mathcal{P}_u$. If R is its proper period set, the border-tail identity gives

$$\mathcal{P}_u=\{p\}\cup\{p+n : n\in R\}.$$

Here and below that identity follows by comparing borders of a word and of one of its border prefixes; equivalently see [7, Lemma 2.6]. The inequality $M_2<p+1$ is precisely

$$-p+\sum_{n\in R}(n-1)(p+n)2^{-n}<0.$$

For $p\ge 5$ this is (13). For $p<5$, we have $b<p\le 4$ and $R\subseteq\{2\}$, so the positive sum is at most $(p+2)/4<p$. Again Proposition 17 applies.

Now suppose $\mathcal{P}_u \not\subseteq \mathcal{P}_v$. Put $p_0 = a - \ell$ and $a + j$ equal to the nearest old periods on either side of a, with the usual tail starting at m. All periods below a agree. If $\ell = 1$, v has consecutive periods p_0, a, which force every period from p_0 onwards; this would imply inclusion, a contradiction. Hence $\ell \geq 2$. For $j \geq 2$, the gap certificate is bounded above by

$$-\ell j + \sum_{n>j} (n+\ell)(n-j)2^{-n} = -\ell j + \frac{6 + 2j + 2\ell}{2^j} < 0.$$

The last expression decreases with $\ell \geq 2$ and $j \geq 2$, and equals $-1/2$ at $(\ell, j) = (2, 2)$.

It remains that $j = 1$. If $a+1$ were also a period of v, its consecutive periods $a, a+1$ would again force inclusion. Therefore $a + 1 \notin \mathcal{P}_v$, and the border prefix $z = v[0, m - a)$ is nonconstant. The new offsets $J - a$ lie in $\{0\} \cup R(z)$. For $\ell \geq 5$, (13) bounds the gap certificate by

$$-\ell + \sum_{n \in R(z)} (n-1)(\ell+n)2^{-n} < 0.$$

For $\ell = 2, 3, 4$, instead take the border prefixes of u, v of common length $d = m - p_0$. Their least periods are respectively $\ell + 1$ and ℓ. Writing s for a period in these prefixes, the gap certificate becomes

$$\sum_{s \in J - p_0} s(s - \ell - 1)2^{\ell - s}.$$

For $\ell = 3, 4$, its sole negative term is the required $s = \ell$; Lemma 19 bounds it strictly below zero. For $\ell = 2$, the second prefix has exactly the even proper periods, whereas the first has every proper period divisible by three. Thus the last display is at most the infinite sum over even integers not divisible by three:

$$4\left(\sum_{k \geq 1} 2k(2k-3)2^{-2k} - \sum_{k \geq 1} 6k(6k-3)2^{-6k}\right) = -\frac{1376}{27783} < 0.$$

Only the required term $s = 2$ is negative, so extending the remaining positive terms was legitimate. This proves the gap criterion in every noninclusion case. Theorem 15 completes the binary proof, and Corollary 16 covers $q \geq 3$. □

Enumeration of border profiles. The checker enumerates profiles recursively. If b is the longest proper border of an m-letter word, all smaller borders are exactly the border set of its b-letter prefix. For each shorter profile B, test the candidate $B \cup \{m\}$ by imposing its position equalities. The equivalence classes of these equalities, assigned distinct letters, realize the candidate if and only if they force no additional border. Every realized candidate is therefore retained and every word's profile is included. This enumerates exactly the profiles over arbitrary finite alphabets, which suffices to cover all binary words. Classical alphabet independence identifies these with the binary profiles as well; see [7, Section 2.2]. Direct binary enumeration through length sixteen and ternary enumeration through length nine independently check the implementation.

Evidence and limits. Floating-point scans over one representative of every realized border class agree with Theorems 13 and 14 for binary words of lengths 2–16 (344 classes summed over lengths), ternary words through length 9 (63), or four-symbol words through length 7 (33). The negative grid was $-12, -6, -3, -1, -.2, -.02$; the positive grid was $.02, .2, 1, 3, 6, 12, 24$. For binary lengths $5, 8, 12, 16$, a 230-point positive grid from .02 to 30 detected no pair with more than one sign change. These are finite scans with explicit tolerances, and cannot certify absence of additional narrow or distant crossings. Proposition 12, not the numerical grid, establishes the eventual reversal for common-period pairs.

The recursive enumeration through length seventy yields 11,418,851 pairs of distinct border classes with a common least period, summed over the lengths. Exact rational arithmetic verifies (10) for 11,378,767 of them and (11) for the remaining 40,084. No pair is uncovered. The former exceptions

at lengths eighteen and twenty satisfy the new inclusion criterion: their common-period means are respectively $462/65 < 13$ and $52418/7321 < 15$. These finite checks corroborate Theorem 20; the proof itself does not depend on enumeration. The exact small-period sums and affine moment bounds used in that proof are also checked separately in the JSON.

The remaining programme includes auditing these working proofs and making the word-specific conjugacy classes explicit. The row-sum theorem supplies a controlled extension to sets. The following section proves a negative-coupling frequency order for biased iid words under weighted overlap domination; its positive-coupling counterparts remain conjectural. General finite-range backgrounds already admit the unbounded crossings of Theorem 8.

7. Biased iid backgrounds

Let π be a positive letter law and normalize the reference pressure to zero. For a word w of length m put

$$\eta_w = \prod_{j=0}^{m-1} \pi(w_j), \qquad \theta_w(s) = \mathbf{1}_{\{s\in\mathcal{P}_w\}} \prod_{j=m-s}^{m-1} \pi(w_j) \quad (1 \le s < m),$$
$$D_w^\pi(x) = \sum_{s=1}^{m-1} \theta_w(s) x^{m-1-s}.$$

Here $\theta_w(s)$ is the probability of another occurrence at displacement s, conditional on an occurrence now. Write P_w^π, ρ_w^π for the normalized pressure and its derivative.

Proposition 21 (Weighted overlap data)**.** *With $t = e^\beta$ and $x = e^{P_w^\pi(\beta)}$, the Perron branch satisfies*

$$(x-1)x^{m-1} = (t-1)\big[\eta_w + (x-1)D_w^\pi(x)\big]. \tag{14}$$

Two equal-length words have identical pressure curves, or identical frequency curves, exactly when their η and all their $\theta(s)$ agree. On positive couplings the pressure crossings are in bijection with the roots $x > 1$ of

$$H(x) = \eta_v - \eta_u + (x-1)(D_v^\pi(x) - D_u^\pi(x)),$$

unless H is identically zero. In particular there are at most $m-1$ distinct positive pressure crossings.

Proof. The all-return generating function is

$$U_w(z) = 1 + \sum_{s=1}^{m-1} \theta_w(s) z^s + \frac{\eta_w z^m}{1-z}.$$

The first-return generating function is $g_w = 1 - U_w^{-1}$, and renewal normalization gives $t g_w(x^{-1}) = 1$. Clearing denominators gives (14), initially for $\beta > 0$ and then for every finite coupling by analyticity of the Perron branch. For $x > 1$ define

$$A_w(x) = \sum_s \theta_w(s) x^{-s} + \frac{\eta_w x^{-m}}{1-x^{-1}}.$$

It decreases strictly from infinity to zero, and $t = 1 + A_w(x)^{-1}$ parametrizes all positive couplings. Its difference has the sign of $H(x)$, and equality yields a common coupling for both curves. Thus the claimed crossing correspondence is exact. Equality of the entire curves forces $H \equiv 0$, which at $x = 1$ gives equality of η and then equality of the two polynomials. Conversely identical data give the same positive pressure branch and hence the same analytic family. Frequency identity is equivalent to pressure identity because both pressures vanish at zero. □

For example, at $\pi(1) = 13/20$, the equal-composition words $u = 1110000111$ and $v = 0011111100$ have proper borders $\{1, 2, 3\}$ and $\{1, 2\}$, respectively. Their polynomial $D_u^\pi - D_v^\pi$ has the sign of

$$9800x^2 - 15600x - 3549.$$

This polynomial has exactly one positive root, lying above one. Consequently their positive pressure curves cross once: $P_v^\pi > P_u^\pi$ near zero, but $P_u^\pi > P_v^\pi$ eventually. Border order alone therefore does not extend to a biased reference law, even when reference word frequencies match.

Proposition 22 (Weighted overlap domination)**.** *Suppose $\eta_u = \eta_v = \eta$ and $\theta_u(s) \le \theta_v(s)$ for every $1 \le s < m$. The reference return gaps and finite-window occurrence counts are convex ordered in the same direction. If at least one coefficient inequality is strict, then $P_u^\pi(\beta) < P_v^\pi(\beta)$ for every finite $\beta \neq 0$.*

Proof. The difference $U_v - U_u$ has nonnegative coefficients. With S_w the return-gap survival generating function, the identity

$$\frac{g_v - g_u}{(1-z)^2} = (U_v - U_u) S_u S_v$$

proves return-gap convex order, since the means are both $1/\eta$. The stationary renewal count argument from the set-of-words proposition proves finite-window count convex order. In the strict case the return laws are distinct, so every integrable strictly convex exponential separates them. Their increasing first-return generating functions reach $e^{-\beta}$ at strictly ordered arguments for $\beta \neq 0$, proving strict pressure order. If the larger generating function has a smaller radius of convergence, its root is already below that endpoint, so the comparison still applies. □

Theorem 23 (Penalization order for biased iid words)**.** *Under the strict hypotheses of Proposition 22,*

$$\rho_u^\pi(\beta) > \rho_v^\pi(\beta) \qquad (\beta < 0).$$

Proof. Strict domination forces $m \ge 2$, and at least one word is nonconstant: equal-probability constant words have identical overlap coefficients. Every nonconstant m-letter word satisfies

$$\eta \le \frac{(m-1)^{m-1}}{m^m}.$$

Indeed, select a letter appearing $k \in \{1, \dots, m-1\}$ times. Its word probability is at most $p^k(1-p)^{m-k}$, maximized at $p = k/m$. Convexity of $k \log k + (m-k)\log(m-k)$ puts its maximum at $k = 1$ or $m - 1$.

Put $L = m - 1$, $\Delta = D_v^\pi - D_u^\pi \neq 0$, and interpolate $D_\theta = D_u^\pi + \theta\Delta$. All coefficients of D_θ and Δ are nonnegative. Let r be the largest root of $(1-r)r^L = \eta$; then $r \ge L/(L+1)$. For $t \in (0, 1)$ the interpolated root $x \in (r, 1)$ is uniquely determined by

$$\frac{1}{1-t} = x^{-L}\left(\frac{\eta}{1-x} - D_\theta(x)\right).$$

The right side is at most one at r, tends to infinity at one, and has strictly positive derivative on $(r, 1)$. Write $y = 1 - x$, so $x > Ly$. As in the uniform argument,

$$\partial_\theta \log x = \frac{\Delta(x)}{T_\theta(x)}, \quad T_\theta = \eta\left(\frac{x}{y^2} - \frac{L}{y}\right) + H_\theta, \quad H_\theta = LD_\theta - xD_\theta' \ge 0.$$

The coefficient and root bounds give

$$\frac{\Delta'}{\Delta} \le \frac{L-1}{x} < \frac{1}{y}.$$

Also $yD_\theta < \eta$, $H_\theta \le LD_\theta$, and $Ly < 1$, hence $y^2 H_\theta < \eta$. Since $H_\theta' \ge 0$,

$$T_\theta' - \frac{T_\theta}{y} = \frac{\eta}{y^3} + H_\theta' - \frac{H_\theta}{y} > 0.$$

Thus Δ/T_θ decreases strictly with x. The root increases strictly with β, so integrating the mixed derivative in θ proves $\rho_v^\pi < \rho_u^\pi$. The endpoint interpolants coincide with their Perron branches near $t = 1$ and throughout $(0, 1)$ by continuity and uniqueness of the root in $(r, 1)$. □

Conjecture 24 (Rewarding words under weighted overlap domination)**.** Under the strict hypotheses of Proposition 22, if $p_v < p_u$ then $\rho_v^\pi(\beta) > \rho_u^\pi(\beta)$ for every $\beta > 0$. If $p_v = p_u$, their frequency difference has at most one positive zero; when one exists, the sign changes from positive to negative.

Near zero the difference is positive, since its derivative there is $2\eta \sum_s (\theta_v(s) - \theta_u(s)) > 0$. When the common least-period coefficient also agrees, the pressure gap is positive and tends to zero at large coupling: (14) gives $P_w^\pi(\beta) = \beta/p_w + \log \theta_w(p_w)/p_w + o(1)$ for bordered words. Thus at least one positive frequency crossing then follows already. The uniqueness assertion remains open here. A finite scan is recorded separately; its scope is weighted overlap domination, not arbitrary biased word pairs or the finite-range backgrounds of Theorem 8.

References

[1] W. Parry and R. F. Williams, *Block coding and a zeta function for finite Markov chains*, Proc. London Math. Soc. (3) 35 (1977), 483–495. doi:10.1112/plms/s3-35.3.483.

[2] W. Parry and S. Tuncel, *On the stochastic and topological structure of Markov chains*, Bull. London Math. Soc. 14 (1982), 16–27. doi:10.1112/blms/14.1.16.

[3] M. Boyle, *Positive K-theory and symbolic dynamics*, 22 June 2001 version, Sections 5.3–5.4. Author's PDF.

[4] M. Boyle and M. C. Sullivan, *Equivariant flow equivalence for shifts of finite type, by matrix equivalence over group rings*, Proc. London Math. Soc. (3) 91 (2005), 184–214. Author's preprint, Proposition 2.7.1 and the accompanying proof.

[5] N. Chandgotia, B. Marcus, J. Richey, and C. Wu, *Shifts of finite type obtained by forbidding a single pattern.* arXiv:2409.09024, Proposition 4.1 and Theorems 4.3, 4.9.

[6] H. Nagaoka, *The exponential family of Markov chains and its information geometry.* arXiv:1701.06119.

[7] E. Rivals, M. Sweering, and P. Wang, *Convergence of the number of period sets in strings*, Algorithmica 87 (2025), 690–711. doi:10.1007/s00453-025-01295-y; author's PDF, Theorem 2.5, Lemma 2.6, and Section 2.2.