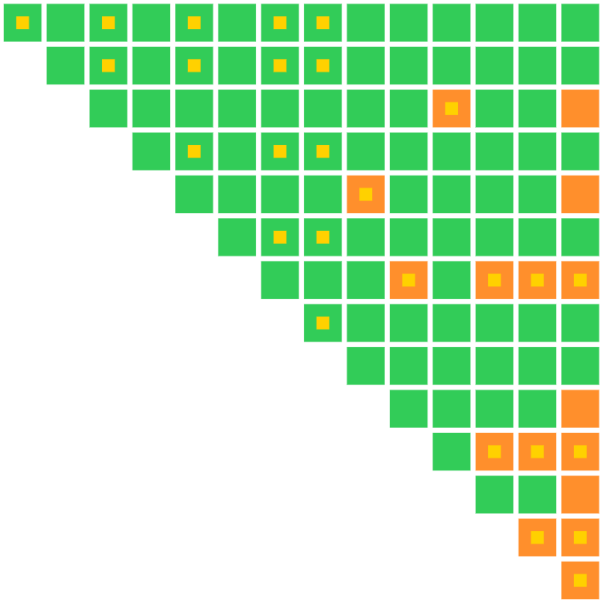


THE COMPONENT EXPONENT OF A TWO-LETTER
KRONECKER WORD

CARLO MITCHENER

MrlyProd, Inc.
First published 2026-09-02, revised 2026-09-08



ABSTRACT. Nest one small black-and-white pattern inside another, then inside a third, and keep going, choosing at each step one of two patterns according to an infinite word. Count the connected pieces after L steps. The count depends on the order of the word, so the question is whether its growth rate does. We give an exact closed formula for the piece count on every one of the 105 two-letter alphabets over the fifteen nonempty two-by-two designs, in five regimes, proved from six geometric lemmas. It follows that when both letters occur with positive frequency the growth rate exists, depends on nothing but the two frequencies, and is therefore blind to the order; on 89 of the 105 it is just the growth rate of the cell count. The one linear rule exact on constant words is refuted at every interior frequency on 78 of the 105. Along the Thue-Morse word over a gasket-and-domino alphabet, one three-cell design against one two-cell bar, the rate is exactly $\frac{1}{2} \log 6$, with a two-sided certificate and no fit. The positive-frequency hypothesis is load-bearing: at a boundary frequency, three words with the same letter frequencies have rates 0, $\log 2$, and no rate at all.

1. INTRODUCTION

Take a two-by-two square and colour some of its four cells black. That is a *design*, and there are fifteen nonempty ones. Substitute a shrunk copy of a second design into every black cell of the first, then a third into every black cell of that, and so on. After L steps you have a black-and-white picture on a $2^L \times 2^L$ grid: the Kronecker product $A_{c_1} \otimes \cdots \otimes A_{c_L}$ of the designs you chose, outermost factor first. The list $w = (c_1, \dots, c_L)$ is a *word*, and the picture is its *value*.



FIGURE 1. The smallest order-sensitive pair: four black cells either way, four specks one way and two dominoes the other. Every cell is computed by `scripts/figure.py`, and the counts are read off the drawing.

The number of black cells does not care about the order of the word. The number of connected black pieces does. Figure 1 is the smallest witness: the same two designs, nested both ways, give four separate specks one way and two dominoes the other. That phenomenon, its exact scope at lengths two and three, and a rank-4 linear representation of the piece count are the subject of a companion paper [1]; this one starts where that one stops.

Since the order matters at every fixed length, ask about the rate. Fix a two-letter alphabet and an infinite word w over it, write $\text{comp}(A_{w_1 \dots w_L})$ for the number of 4-connected black pieces of the length- L prefix, and ask whether

$$\chi(w) = \lim_{L \rightarrow \infty} \frac{1}{L} \log \text{comp}(A_{w_1 \dots w_L})$$

exists, and if it does, whether it can see anything the multiset of letters cannot. The hope, stated plainly so that its refutation is visible: a word like Thue-Morse, aperiodic but with perfectly balanced letter frequencies, ought to return a different number from a periodic word with the same frequencies, and that difference would be a genuinely non-stationary effect in an object whose dimension theory is already known to be a frequency average.

The answer has three parts, and the third one kills the hope.

Every alphabet has a closed form. There is an exact formula for $\text{comp}(A_w)$ on each of the $\binom{15}{2} = 105$ two-letter alphabets, and five regimes, proved from six geometric lemmas, cover them all (Theorem 3.1). Two-thirds of the alphabets fall to a single lemma: if either letter has no contact with a copy of itself in either direction, the piece count is the cell count of a prefix, and nothing else needs to be said. The other four regimes are two letters that never break apart, two dominoes at right angles, a run-counting law, and a row-block law.

At an interior frequency the rate exists and is order-blind. If both letters occur with strictly positive frequency, then $\chi(w)$ exists, is a function of the frequency vector alone, and is read off the closed form (Theorem 3.3). Existence is earned from the formula: no ergodic theorem is used, and Proposition 3.14 says why the matrix-norm route could not supply one. Consequently Thue-Morse returns exactly what a periodic word of the same frequencies returns, and the hoped-for non-stationary effect is not there.

And most of the rate is not news. On 89 of the 105 alphabets the rate equals the growth rate of the cell count, which was order-blind to begin with (Theorem 3.4). The one class of alphabets whose rate sits strictly between the constant-word values and that ceiling is a domino against the full tile, where at equal frequencies the rate is $\frac{1}{2} \log 2$ against per-letter values 0, 0 and a ceiling of $\frac{3}{2} \log 2$.

What does fail, cleanly and everywhere, is the obvious guess. A constant word c^L has $\text{comp}(A_{c^L}) = \text{comp}(A_c)^L$, and $\text{comp}(A_c) = 2$ for exactly two of the fifteen designs, so the unique linear functional on the frequency simplex that is exact on constant words is $\Phi(f) = (f_6 + f_9) \log 2$. It is refuted at every interior frequency on 78 of the 105 alphabets (Theorem 3.5), and changing its coefficients does not repair it: the interior rate is linear, so some linear functional does match it there, but on those alphabets that functional continues to a positive value at a vertex where the constant word has rate 0, so no continuous frequency functional whatever gives the rate of every word whose letter frequencies and rate both exist. Figure 2 is the whole verdict in one square.

The hypothesis inside every statement above is that both letters occur with positive frequency, and it is not decoration. Over the alphabet $\{3, 6\}$ at the frequency vector $(1, 0)$ the constant word has rate 0, the word carrying the second letter at the square places has rate $\log 2$, and the word carrying

it at the powers of 2 has no rate at all: its prefix rates accumulate on a whole interval (Theorem 3.11). The same three-word pathology occurs on an alphabet carrying no diagonal letter, so it is not a property of one design class.

Where this sits. Level-varying constructions are not new, and the honest position is that what is owned here is the alphabet and the experiment, not the construction. A word is a non-autonomous iterated function system in the sense of [7], restricted to a parity-rule alphabet; at similarity maps it is a Moran construction [5], and the dimension formulas for varying-ratio constructions [6], specialised to a common base, give the scale dimension of a word as a frequency average. Those are cited, not claimed, and the Feng, Wen and Wu article was unreachable to the author, so its statements are known here only through verbatim restatements in citing papers. In the plane the pictures are generalised Sierpinski carpets in the sense of [3], and mixed labyrinth fractals [4] already own the word *mixed* for level-varying patterns. A randomised schedule is a V -variable fractal at $V = 1$ [8], whose dimension theory runs through products of random matrices in the companion [9]. Several distinct ratios inside one level, which this construction cannot express, is the multiscale substitution programme [10]. And a word is an S-adic directive sequence in symbolic dynamics [11], where the question of which aperiodic words behave well is a mature subject.

What is specific here is narrower. The alphabet is finite, enumerated and classified by symmetry, where the metric-geometric literature takes each level's rule as an arbitrary given subset; the composition law is the Kronecker product, which makes contact multiplicativity one line rather than a graph-directed computation; and order is run as a controlled experiment, holding side, fill, density and dimension fixed while only the letter order moves, which is an experimental design the varying-ratio literature has no reason to run. The scale dimension of a word is a frequency average by the cited formulas, and the point of the component count is that it is the first observable on this construction that is not.

What this paper does not claim. It does not claim a non-stationary result, and it retires the conjecture that a gap against Φ would be one. It does not identify χ for a word whose letter frequencies fail to exist, and Conjecture 3.16 records the evidence that positive lower density is not the right weaker hypothesis. It says nothing about alphabets of three or more letters, about bases above 2, about dimensions above 2, or about the other order-sensitive observables of [1]. It does not decide connectedness: whether a level-varying plane carpet is connected at all, which is the question of whether the count equals 1, has published necessary and sufficient conditions [3], and the formulas here are counts and rates rather than a criterion. The publisher's text of that paper was not reachable to the author either, so its definition, statement and method are known here only through the authors' own companion paper on totally disconnected generalised carpets. No prior

art for these counts and rates was found in the literature searched, which is a report on a search and not a claim of novelty.

2. DEFINITIONS

Definition 2.1 (Designs and codes). A *design* is a subset of the four cells of a 2×2 grid, encoded by a number $c \in \{1, \dots, 15\}$ whose bit i is the cell at row $\lfloor i/2 \rfloor$ and column $i \bmod 2$. So code 3 is the top row, code 5 the left column, code 9 the main diagonal, code 6 the antidiagonal, code 7 the L-shape missing the bottom-right cell, and code 15 the full tile. Write $A_c \subseteq \{0, 1\}^2$ for the set of filled cells and $k_c = |A_c|$ for its *fill*.

Definition 2.2 (Words and their pictures). For a word $w = (c_1, \dots, c_L)$ the picture $A_w \subseteq \{0, \dots, 2^L - 1\}^2$ is the Kronecker product $A_{c_1} \otimes \dots \otimes A_{c_L}$, first letter outermost: a cell with row digits $(r_1, \dots, r_L)$ and column digits $(s_1, \dots, s_L)$ in base 2 is filled exactly when $(r_i, s_i) \in A_{c_i}$ for every i . Its fill is $\text{fill}(A_w) = \prod_i k_{c_i}$, which does not depend on the order. We write $\text{comp}(A_w)$ for the number of 4-connected components of A_w and $w_{1..i}$ for the prefix of length i .

Definition 2.3 (Contacts). For a picture X of side n , $h(X)$ is the number of rows r with both $(r, 0)$ and $(r, n-1)$ filled, and $v(X)$ the number of columns s with both $(0, s)$ and $(n-1, s)$ filled. These count the places at which two side-by-side, respectively two stacked, copies of X touch. A design is *contact-free* if $h = v = 0$, which happens exactly for the four one-cell designs 1, 2, 4, 8 and the two diagonal designs 6, 9. It is *heavy* if it is connected and $h, v \geq 1$, which happens exactly for the four three-cell designs 7, 11, 13, 14 and the full tile 15. The remaining four designs 3, 12, 5, 10 are the *dominoes*, each with exactly one contact direction; 3 and 12 are the row dominoes, 5 and 10 the column dominoes; two dominoes are *like* when they share an orientation and *unlike* when they do not.

The six classes named above, the units $\{1, 2, 4, 8\}$, the row dominoes $\{3, 12\}$, the column dominoes $\{5, 10\}$, the diagonals $\{6, 9\}$, the gaskets $\{7, 11, 13, 14\}$ and the full tile $\{15\}$, are the equivalence classes of every statement in this paper. There are $\binom{15}{2} = 105$ two-letter alphabets and $\binom{6}{2} + 5 = 20$ pairs of classes, of which 15 pair distinct classes.

Definition 2.4 (Frequency and exponent). An infinite word w over $\{a, b\}$ has *letter frequencies* $f = (f_a, f_b)$ if the number of occurrences of each letter in $w_{1..L}$, divided by L , converges. The frequency is *interior* if $f_a > 0$ and $f_b > 0$. The *component exponent* is $\chi(w) = \lim_L \frac{1}{L} \log \text{comp}(A_{w_{1..L}})$ when the limit exists, and the *fill exponent* is $f_a \log k_a + f_b \log k_b$, which always exists when f does and never depends on the order.

Definition 2.5 (The constant-word functional). Write $\Phi(f) = (f_6 + f_9) \log 2$. By Lemma 4.6 below it is the unique linear functional on the frequency simplex that agrees with the rate of every constant word.

3. RESULTS

3.1. Five regimes close all 105 alphabets.

Theorem 3.1 (The closed forms). *Let $\{a, b\}$ be one of the 105 two-letter alphabets and w a word over it of length L . Then $\text{comp}(A_w)$ is given exactly by the table below. The first regime is every alphabet with a contact-free letter; the other four assume neither letter is contact-free, so that both are dominoes, gaskets or the full tile, and the five are mutually exclusive and exhaustive.*

<i>regime</i>	$\text{comp}(A_w)$	<i>alphabets</i>
one letter contact-free	$\text{fill}(A_{w_{1..p}})$, with p the last contact-free place, and 1 if there is none	69
two heavy letters, or two like dominoes	1	12
two unlike dominoes	2^{L-r} , with r the terminal run	4
a domino and the full tile	2^{n-j} , with n the number of full letters and j their terminal run	4
a domino and a gasket	$1 + \sum \text{fill}(A_{w_{1..i-1}})$ over the gasket places $i \leq m$, with m the last domino place	16

In the last row the count is 1 if w carries no domino. The five counts sum to 105.

The first regime is one lemma, the zero-contact cut, and it settles two-thirds of the table. The proof of the whole theorem is in Section 4.

Fact 3.2 (The forms against the drawing). `scripts/verify.py` renders A_w cell by cell from Definition 2.2 by an independent route and counts its 4-connected components by union-find over row runs. On all 26,670 words of length at most 7 over all 105 alphabets, and on all 4,080 words of length at most 8 over the eight named alphabets (3, 7), (5, 7), (3, 15), (3, 6), (3, 5), (7, 15), (1, 3), (6, 9), the drawn count equals Theorem 3.1 with zero mismatches.

3.2. The exponent at an interior frequency.

Theorem 3.3 (Existence, order-blindness, frequency-only). *Let w be an infinite word over one of the 105 alphabets whose letter frequencies exist and are interior. Then $\chi(w)$ exists, depends only on f , and equals $f_a \epsilon_a + f_b \epsilon_b$ with the per-letter weights ϵ given by the regime:*

- one letter contact-free: $\epsilon_c = \log k_c$ for both letters;
- two heavy letters, or two like dominoes: $\epsilon_a = \epsilon_b = 0$;

- two unlike dominoes: $\epsilon_a = \epsilon_b = \log 2$;
- a domino and the full tile: $\epsilon_{\text{domino}} = 0$ and $\epsilon_{\text{full}} = \log 2$;
- a domino and a gasket: $\epsilon_{\text{domino}} = \log 2$ and $\epsilon_{\text{gasket}} = \log 3$.

In particular χ is the same along Thue-Morse, along every periodic word with those frequencies, along every strictly ergodic word in which both letters occur, and at almost every Bernoulli word with $0 < p < 1$.

Theorem 3.4 (Saturation). *At every interior frequency the component exponent equals the fill exponent on 89 of the 105 alphabets and falls strictly below it on the other 16. The alphabets on which it falls short are the 12 of the second regime and the 4 of the fourth. Among the 16, exactly the 4 of the fourth regime, that is a domino against the full tile, have an exponent strictly between $\Phi(f)$ and the fill exponent; at equal frequencies that exponent is $\frac{1}{2} \log 2$ against per-letter values 0,0 and a ceiling of $\frac{3}{2} \log 2$. Up to transposition this is the one class pair with that property.*

Theorem 3.5 (The constant-word functional is refuted). *At every interior frequency, Φ is exact on 27 of the 105 alphabets and refuted on the other 78. It is exact exactly on the 15 alphabets both of whose letters are units or diagonals, and on the 12 alphabets of the second regime. On each of the 78 the obstruction sits at a vertex rather than in the interior: changing the coefficients does repair the interior, since the interior formula is itself linear, but the repaired functional extends continuously to the closed simplex and at some vertex takes a positive value while the constant word sitting at that vertex has count 1 and rate 0. Hence on those alphabets no continuous functional of f alone gives the rate of every word whose letter frequencies exist and whose rate exists.*

Remark 3.6 (Two different eighty-nines). The 89 of Theorem 3.4 is the set of alphabets whose exponent saturates. There is a second set of size 89 below, in Corollary 4.10: the alphabets outside the fifth regime, on which every count is 3-smooth, that is of the form $2^\alpha 3^\beta$. The two are different sets, and since each holds 89 of the same 105 they must differ by exactly sixteen on each side. The sixteen of the fifth regime saturate and are not smooth. The sixteen of the second and fourth regimes are smooth and do not saturate: the second because its count is always 1, and the fourth because its count 2^{n-j} is a power of two, while Theorem 3.3 gives it the exponent $f_{\text{full}} \log 2$ against a fill exponent of $(1 + f_{\text{full}}) \log 2$, so it falls short at every interior frequency. The coincidence of size is a coincidence.

3.3. Thue-Morse, exactly. Let t be the Thue-Morse sequence [2], t_n the parity of the binary digit sum of n , and read an infinite word over a gasket-and-domino alphabet off it by sending $t_n = 0$ to one letter and $t_n = 1$ to the other, starting at $n = 0$, so that w_i is the letter of t_{i-1} . There are two readings, and they give different words.

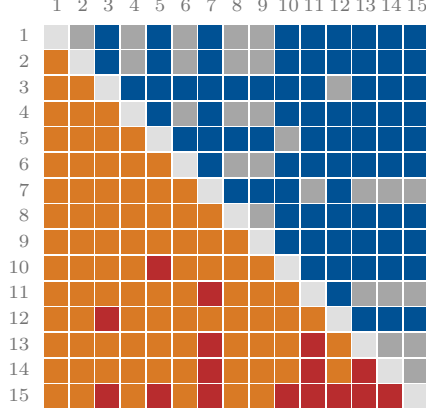


FIGURE 2. All 105 two-letter alphabets, indexed by design code along both axes. Lower left triangle: orange where the component exponent meets the fill ceiling (89 alphabets), red where it falls short (16). Upper right triangle: grey where the constant-word rule Φ is exact (27), blue where it is refuted (78). The diagonal is blank, since an alphabet has two distinct letters. Every cell is computed by `scripts/figure.py` from the closed forms of Theorem 3.1, and the four counts in this caption are printed by the same script.

Theorem 3.7 (The Thue-Morse value). *Over each of the 16 gasket-and-domino alphabets, and under either reading, the Thue-Morse word has $\chi = \frac{1}{2} \log 6$, with the two-sided certificate*

$$\left| \log \text{comp}(A_{w_{1..L}}) - \frac{L}{2} \log 6 \right| \leq \log 108 + \frac{1}{2} \log \frac{3}{2} < 4.885$$

for every $L \geq 4$. In nats the value is 0.895879734614027, and in units of $\log 2$ it is 1.292481250360578; both are floats naming an exact constant.

The certificate does not come from an ergodic theorem or from a fit. It comes from two elementary properties of t , that it has no three equal consecutive letters and that its letter counts are exactly balanced, fed into a sandwich bounding comp between two multiples of one prefix fill. The two alphabets (3, 7) and (5, 7) are the two whose rate resisted identification longest, and they are covered like the other fourteen.

Fact 3.8 (The certificate, measured). `scripts/verify.py` evaluates the closed form along Thue-Morse under both readings at every length $4 \leq L \leq 2^{14}$. The largest deviation from $\frac{L}{2} \log 6$ is 4.273459 nats, against the certificate constant 4.884864. The prefix rate in units of $\log 2$ reads 1.291967463826 at $L = 4096$ and 1.292352803727 at $L = 2^{14}$ under one reading, and 1.291291597694 and 1.292183837194 under the other, against the limit 1.292481250360578; all six are floats.

Fact 3.9 (The saturation ratio over one finite sweep). This is the result easiest to overstate: the exponents agree with the fill exponent, and the counts themselves do not. Along Thue-Morse over $(3, 7)$ and over $1 \leq L \leq 2^{14}$, `scripts/verify.py` measures the ratio `comp / fill` as taking its smallest value 0.0113766545, above the floor $1/108$ proved in Lemma 4.8, and its largest value at $L \geq 5$ as the exact rationals $43397/186624$ under one reading and $151/648$ under the other. The cutoff at $L \geq 5$ is not cosmetic: on the short prefixes Theorem 3.1 gives the larger values $1/2$ at $L = 1$ and $1/4$ at $L = 4$ in one reading and $1/3$ at $L = 2$ in the other. At $L = 4096$ the ratio is 0.2325367033, which is neither the largest nor the smallest value of that sweep. Nothing here is a claim about the limit; Conjecture 3.17 is what the paper says about that.

3.4. Failure modes. The four statements below are what the investigation refuted, and each was at some point the expected answer. They are stated as results because each of them closes a route.

Remark 3.10 (A gap against Φ is not a non-stationary result). By Theorem 3.3, at an interior frequency χ is a function of f alone on every one of the 105 alphabets. So Thue-Morse returns exactly what a periodic word of the same frequencies returns and what almost every Bernoulli word with those frequencies returns, and every one of those stationary words misses Φ by exactly the same amount. What a gap against Φ refutes is the frequency functional, not stationarity, and no aperiodic word witnesses anything here. This retires the conjecture that the component exponent is where a level-varying construction stops being a frequency average.

Theorem 3.11 (The interior hypothesis is sharp). *Fix the alphabet $\{3, 6\}$ and the frequency vector $f = (1, 0)$. Then:*

- (1) *the constant word 3^∞ has $\chi = 0$;*
- (2) *the word carrying the letter 6 exactly at the square places has $\chi = \log 2$;*
- (3) *the word carrying the letter 6 exactly at the powers of 2 has upper rate $\log 2$, lower rate $\frac{1}{2} \log 2$, and no rate, its prefix rates accumulating on the whole interval $[\frac{1}{2} \log 2, \log 2]$ rather than on its endpoints.*

The same three statements hold over the alphabet $\{3, 7\}$, which carries no diagonal letter, with the letter 7 in place of 6; so the pathology is not a property of the diagonal class. Hence no statement of Theorem 3.3 survives dropping the interior hypothesis. The orbit closure of each of the three words is countable and uniquely ergodic, and for the two non-constant words it is not minimal, so no minimal word is implicated. The constant word is of course minimal, and it is the one of the three that does have a rate.

Fact 3.12 (The boundary rates, measured). For the powers-of-2 word over $(3, 7)$, `scripts/verify.py` reads the prefix rate in units of $\log 2$ as 0.502367981 at $L = 2048$, 1.002164269 at $L = 2049$, and 0.500219405 at

$L = 32768$, and over the single block $4097 \leq L \leq 8192$ it decreases monotonically from 1.00123 to 0.50073, which is the interval of Theorem 3.11 visible in one sweep. For the same word over $(3, 6)$ the prefix rate is exactly $2^k/(2^{k+1} - 1)$ at $L = 2^{k+1} - 1$, printing $4/7, 8/15, \dots, 8192/16383$, and exactly 1 at every $L = 2^k$.

Fact 3.13 (The matrix-norm exponent is a different number). Restricted to words over $\{3, 6\}$ the component count is a rational series of rank 2: with $\lambda = (1, 0)$, $\gamma = (1, 1)^\top$,

$$M_3 = \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix}, \quad M_6 = \begin{pmatrix} 2 & 0 \\ 4 & 0 \end{pmatrix}, \quad \text{comp}(A_w) = \lambda M_{c_1} \cdots M_{c_L} \gamma,$$

verified by `scripts/verify.py` on all 8,190 words of length at most 12, and the largest entry of M_3^L is $2^{L+1} - 1$ at every $L \leq 24$.

Proposition 3.14 (The matrix-norm exponent is a different number). $\lim_L \frac{1}{L} \log \|M_3^L\| = \log 2$ in any matrix norm, while $\chi(3^\infty) = 0$. Hence no argument that returns a top Lyapunov exponent [12] or a joint spectral radius [13], and no projective-contraction argument on forward orbits, can return χ : they answer a different question. The rank-4 representation over the full alphabet, and the invariant cone it carries, belong to [1]; neither is used anywhere in this paper, and this proposition is why neither would help.

Proof. M_3 has characteristic polynomial $\lambda^2 - 3\lambda + 2$, so its eigenvalues are 1 and 2 and its spectral radius is 2; by Gelfand's formula the norm exponent is $\log 2$. Explicitly, splitting M_3 over its eigenvectors gives

$$M_3^L = \begin{pmatrix} 2 - 2^L & 2^L - 1 \\ 2 - 2^{L+1} & 2^{L+1} - 1 \end{pmatrix},$$

whose largest entry in absolute value is $2^{L+1} - 1$, which is the number Fact 3.13 measures. On the other side $\text{comp}(A_{3^L}) = 1$ at every L by Lemma 4.6, so $\chi(3^\infty) = 0$. $\square$

Remark 3.15 (Thue-Morse is not a non-existence witness). A headline that does not exist, said out loud so that it is not written later: the component exponent does not fail to exist along Thue-Morse. It converges, exactly, by Theorem 3.7. The genuine non-existence witness is the third word of Theorem 3.11, which sits at a boundary frequency and whose orbit closure is not minimal.

3.5. What stays open.

Conjecture 3.16 (Words with no letter frequencies). *Positive lower density of both letters is not enough for χ to exist. Over $(3, 7)$ the tripling word $W_{k+1} = W_k 7^{|W_k|} 3^{|W_k|}$ with $W_1 = (7, 3)$ keeps both letters at lower density at least $1/4$, and `scripts/verify.py` finds its prefix rate ranging over $[0.4792, 1.4379]$ in units of $\log 2$ on $1024 \leq L \leq 4096$ with no narrowing.*

The closed form is exact there; the non-existence of the limit is measurement, not proof. Which words without letter frequencies carry a component exponent is open.

Conjecture 3.17 (The accumulation set of the saturation). *Whether comp/fill converges at all along Thue-Morse over a gasket-and-domino alphabet is open, and the measurement of Fact 3.9 is that it does not settle over $1 \leq L \leq 2^{14}$, its values there spreading over more than a factor of twenty. Its accumulation set is unidentified, and is plausibly the attractor of the two affine maps that the two letters induce on the normalised chart. Lemma 4.8 traps every value at $L \geq 4$ in $(1/108, 5/12]$.*

Conjecture 3.18 (Where a non-stationary effect could still live). *Nothing here speaks about alphabets of three or more letters, about the Euler characteristic, boundary or hole counts, whose representations have larger rank [1], or about bases above 2. Whether any of them makes the exponent depend on more than the letter frequencies, which is what a genuinely non-stationary result would need, is open; at an interior frequency on two letters it provably does not.*

4. PROOFS

4.1. Six lemmas.

Lemma 4.1 (Contacts multiply). $h(A_w) = \prod_i h(A_{c_i})$ and $v(A_w) = \prod_i v(A_{c_i})$.

Proof. It suffices to treat $A \otimes B$ with A of side m and B of side n . A row of $A \otimes B$ is indexed by a pair (r, r') with $r < m$, $r' < n$, and its first and last cells are (r, r') -indexed cells of the leftmost and rightmost blocks, that is the cells $(r', 0)$ of B inside block column 0 of A and $(r', n - 1)$ of B inside block column $m - 1$. So row (r, r') of $A \otimes B$ has both end cells filled exactly when row r of A has both end cells filled and row r' of B has both end cells filled. Counting pairs gives $h(A \otimes B) = h(A)h(B)$, and the column statement is the transpose. Induction on the length finishes it. $\square$

Lemma 4.2 (The zero-contact cut). (a) *If $h(A_u) = v(A_u) = 0$ for a suffix $u = w_{p..L}$, then $\text{comp}(A_w) = \text{fill}(A_{w_{1..p-1}}) \cdot \text{comp}(A_u)$.* (b) *If z is contact-free then $\text{comp}(A_{zu}) = k_z \cdot \text{comp}(A_u)$ for every word u , and in particular $\text{comp}(A_z) = k_z$.*

Proof. For (a), write $A_w = A_{w_{1..p-1}} \otimes A_u$. Each filled cell of $A_{w_{1..p-1}}$ carries one translated copy of A_u inside its own block, and two cells of A_w lying in different blocks are 4-adjacent only if the two blocks sit at 4-adjacent cells of $A_{w_{1..p-1}}$, since two distinct block row-ranges overlap only when the block rows agree, and likewise for columns. Two blocks at horizontally adjacent cells contribute a 4-adjacent pair exactly when some row of A_u has both its end cells filled, that is when $h(A_u) \geq 1$; vertically adjacent blocks likewise need $v(A_u) \geq 1$. Both are zero, so distinct copies never merge, and the

component count is the number of copies times the component count of one copy.

For (b), a contact-free design is either a single cell or one of the two diagonal designs, so its filled cells are pairwise non-4-adjacent. Hence in $A_z \otimes A_u$ the k_z copies of A_u sit in pairwise non-adjacent blocks and never touch, whatever the contacts of A_u . $\square$

Lemma 4.3 (A heavy suffix letter does nothing). *If c is heavy then $\text{comp}(A_{wc}) = \text{comp}(A_w)$ for every word w .*

Proof. In $A_w \otimes A_c$ each filled cell of A_w becomes a block congruent to A_c , which is connected. Two blocks at horizontally adjacent cells meet exactly when $h(A_c) \geq 1$, two at vertically adjacent cells exactly when $v(A_c) \geq 1$, and both hold; two blocks at non-adjacent cells are disjoint and never touch. So the graph on blocks with an edge when two blocks meet is isomorphic to the graph on filled cells of A_w with an edge when two cells are 4-adjacent, and connected components correspond. The base case $\text{comp}(A_c) = 1$ is connectedness of A_c . $\square$

Lemma 4.4 (A domino suffix letter counts runs). *Let $H(X)$ be the number of maximal horizontal runs of filled cells of X , summed over rows. For a row domino p , $\text{comp}(A_{wp}) = H(A_w)$. Moreover $H(\varepsilon) = 1$, $H(A_{wp}) = H(A_w)$ at a row domino p , and $H(A_{wq}) = H(A_w) + \text{fill}(A_w)$ at a gasket q . The transposed statements hold for column dominoes with vertical runs.*

Proof. A row domino has one full row and one empty row. In $A_w \otimes A_p$ every filled cell of A_w becomes a block whose one filled row is the same row of the block for every block, and the other row of each block is empty. Rows of $A_w \otimes A_p$ therefore alternate between a copy of a row of A_w with each cell doubled and an empty row, so no two filled rows are vertically adjacent, and doubling each cell merges exactly the cells that were already horizontally adjacent. Hence components are in bijection with maximal horizontal runs of A_w , which is the first claim, and the run count is unchanged, which is the second.

For a gasket q : each of 7, 11, 13, 14 has one full row and one row with a single cell. In $A_w \otimes A_q$ the rows carrying the full row of each block reproduce the runs of A_w with each cell doubled, giving $H(A_w)$ runs; the rows carrying the one-cell row of each block hold one isolated cell per filled cell of A_w , since consecutive blocks are two columns apart and the single cells occupy the same column of each block, giving $\text{fill}(A_w)$ runs. Adding gives $H(A_{wq}) = H(A_w) + \text{fill}(A_w)$. $\square$

Lemma 4.5 (The row-block law). *Over an alphabet consisting of one row domino and the full tile, $\text{comp}(A_w) = 2^{n-j}$ with n the number of full letters in w and j the length of the terminal run of full letters. The transposed statement holds for a column domino.*

Proof. A row domino forces the row digit at its place, say to the value d , and leaves the column digit free; the full tile forces neither. So the filled

set is a product $R \times \{0, \dots, 2^L - 1\}$, where R is the set of rows whose digit equals d at every domino place, and $|R| = 2^n$. Every row of R is a full horizontal line, so two of them lie in the same component exactly when they are vertically adjacent, and the component count is the number of maximal blocks of consecutive integers in R .

Index the row digits so that place i contributes 2^{L-i} . The last j places are full, hence free, and place $L - j$ is a domino, hence forced, whenever $j < L$. So R is a union of blocks of 2^j consecutive integers, one for each choice of the free digits above place $L - j$, and the integer just above such a block differs from its top element by flipping the forced digit at place $L - j$ from d , so it is not in R and the blocks are pairwise non-adjacent. There are 2^{n-j} of them. If $j = L$ then R is everything and the count is $1 = 2^{n-j}$; if $n = 0$ then $j = 0$, R is a single row and the count is 1. $\square$

Lemma 4.6 (Constant words). $\text{comp}(A_{cL}) = \text{comp}(A_c)^L$ for every code c , and $\text{comp}(A_c) = 2$ exactly for $c \in \{6, 9\}$ and 1 for the other thirteen.

Proof. A unit design has fill 1, so A_{cL} is one cell. A domino word of one orientation is a full line, hence connected, by Lemma 4.4 applied repeatedly or directly: the filled set is a product of a single forced row with the whole column range. A heavy letter gives 1 at every length by Lemma 4.3. A diagonal letter is contact-free with fill 2, so Lemma 4.2 applies at every suffix and $\text{comp}(A_{cL}) = 2^L$. Finally $\text{comp}(A_c)$ is 1 for the nine designs of fill 1, 3, 4 and for the four dominoes, and 2 for the two diagonals, since the two cells of a diagonal design are not 4-adjacent. $\square$

4.2. Proof of Theorem 3.1.

Proof. The five regimes are mutually exclusive and exhaust the 105 alphabets. Counting: the contact-free designs are the six codes 1, 2, 4, 8, 6, 9, so the alphabets meeting the first regime number $\binom{6}{2} + 6 \cdot 9 = 15 + 54 = 69$. The remaining $\binom{9}{2} = 36$ alphabets lie among the four dominoes, the four gaskets and the full tile, and split as 10 pairs of heavy letters plus 2 pairs of like dominoes, giving 12; 4 pairs of unlike dominoes; 4 pairs of a domino with the full tile; and $4 \cdot 4 = 16$ pairs of a domino with a gasket. Then $69 + 12 + 4 + 4 + 16 = 105$.

First regime. Suppose first that both letters are contact-free. Then $p = L$, and the one-letter suffix w_L has zero contact, so Lemma 4.2(a) followed by Lemma 4.2(b) gives

$$\text{comp}(A_w) = \text{fill}(A_{w_{1..L-1}}) \text{comp}(A_{w_L}) = \text{fill}(A_{w_{1..L-1}})k_{w_L} = \text{fill}(A_{w_{1..p}}).$$

Otherwise exactly one letter z is contact-free; call the other c . If w carries no z then $w = c^L$ with c a domino, a gasket or the full tile, so $\text{comp}(A_w) = 1$ by Lemma 4.6, which is the stated value. Otherwise let p be the last place carrying z . Every letter after p equals c , so the suffix $u = w_{p..L} = z c^{L-p}$ has $h(A_u) = h(A_z)h(A_c)^{L-p} = 0$ and likewise $v(A_u) = 0$ by Lemma 4.1. Then Lemma 4.2(a) gives $\text{comp}(A_w) = \text{fill}(A_{w_{1..p-1}}) \text{comp}(A_u)$, and Lemma 4.2(b)

applied to u gives $\text{comp}(A_u) = k_z \text{comp}(A_{c^{L-p}}) = k_z$, using $\text{comp}(A_{c^{L-p}}) = 1$ from Lemma 4.6. Multiplying, $\text{comp}(A_w) = \text{fill}(A_{w_{1..p-1}})k_z = \text{fill}(A_{w_{1..p}})$.

Second regime. If both letters are heavy then $\text{comp}(A_w) = 1$ by Lemma 4.3 and induction on the length. If both are row dominoes then every letter forces the row digit at its place and leaves the column digit free, so the filled set is a single full row; if both are column dominoes the transposed argument gives a single full column. Either way $\text{comp}(A_w) = 1$.

Third regime. Let r be the length of the terminal run and suppose first $r < L$, so that place $L-r$ carries the other orientation. The suffix $u = w_{L-r..L}$ satisfies $h(A_u) = h(A_{w_{L-r}})h(A_{w_L})^r$ and $v(A_u) = v(A_{w_{L-r}})v(A_{w_L})^r$; one of the two orientations has $h = 0$ and the other has $v = 0$, so both products vanish. Lemma 4.2 gives $\text{comp}(A_w) = \text{fill}(A_{w_{1..L-r-1}}) \text{comp}(A_u) = 2^{L-r-1} \text{comp}(A_u)$. And $\text{comp}(A_u) = 2$, computed directly: up to transposition A_u is a column domino carrying in each of its two cells a copy of A_{w_L} , which is a full horizontal line at one row of its $2^r \times 2^r$ block; the two copies occupy blocks one above the other, so the two lines sit $2^r \geq 2$ rows apart and are not adjacent. Hence $\text{comp}(A_w) = 2^{L-r}$. If $r = L$ the word is constant, $\text{comp}(A_w) = 1 = 2^{L-r}$.

Fourth regime. This is Lemma 4.5.

Fifth regime. Let q be the gasket and p the domino, and suppose without loss of generality that p is a row domino, the column case being the transpose. If w carries no domino then $w = q^L$ and $\text{comp}(A_w) = 1$ by Lemma 4.3. Otherwise let m be the last domino place. Every letter after m is the gasket, which is heavy, so $\text{comp}(A_w) = \text{comp}(A_{w_{1..m}})$ by Lemma 4.3. The last letter of $w_{1..m}$ is the row domino, so $\text{comp}(A_{w_{1..m}}) = H(A_{w_{1..m-1}})$ by Lemma 4.4. Finally the run recursion of Lemma 4.4 telescopes: starting from $H(\varepsilon) = 1$ and reading $w_{1..m-1}$ letter by letter, each domino leaves H unchanged and each gasket at place i adds $\text{fill}(A_{w_{1..i-1}})$. Hence $H(A_{w_{1..m-1}}) = 1 + \sum \text{fill}(A_{w_{1..i-1}})$ over gasket places $i \leq m-1$, and since m itself is a domino place this is the sum over gasket places $i \leq m$. $\square$

4.3. Proof of Theorem 3.3.

Lemma 4.7 (Terminal runs vanish at interior frequency). *Let w be an infinite word over $\{a, b\}$ whose letter frequencies exist and are interior, and let $r(L)$ be the length of the terminal run of $w_{1..L}$. Then $r(L) = o(L)$.*

Proof. Suppose not, so there are $\varepsilon > 0$ and lengths $L_1 < L_2 < \dots$ with $r(L_i) \geq \varepsilon L_i$. Let x be a letter not equal to the repeated letter of the run at L_i , passing to a subsequence so that x is the same letter throughout, and let $N_x(L)$ count occurrences of x in $w_{1..L}$. Then $N_x(L_i) = N_x(L_i - r(L_i))$, since x does not occur in the terminal run. Set $M_i = L_i - r(L_i) \leq (1-\varepsilon)L_i$. If M_i is bounded then $N_x(L_i)$ is bounded and $f_x = 0$. Otherwise $N_x(M_i)/M_i \rightarrow f_x$, so

$$\frac{N_x(L_i)}{L_i} = \frac{N_x(M_i)}{M_i} \cdot \frac{M_i}{L_i} \leq (f_x + o(1))(1-\varepsilon),$$

while the left side tends to f_x . Hence $f_x \leq (1-\varepsilon)f_x$, so $f_x = 0$, contradicting interiority. $\square$

Proof of Theorem 3.3. Write $N_c(L)$ for the count of the letter c in $w_{1..L}$, so $N_c(L)/L \rightarrow f_c$.

First regime. By Theorem 3.1, $\log \text{comp}(A_{w_{1..L}}) = \log \text{fill}(A_{w_{1..p(L)}})$ with $p(L)$ the last contact-free place. If both letters are contact-free then $p(L) = L$ and the right side is $\sum_{i \leq L} \log k_{w_i}$, whose average tends to $f_a \log k_a + f_b \log k_b$. Otherwise $L - p(L)$ is the terminal run of the non-contact-free letter, hence $o(L)$ by Lemma 4.7, and

$$\log \text{fill}(A_{w_{1..p(L)}}) = \log \text{fill}(A_{w_{1..L}}) - (L - p(L)) \log k_c = \sum_{i \leq L} \log k_{w_i} + o(L),$$

giving the same limit. Both cases are the fill exponent, which is the stated weight table since $\epsilon_c = \log k_c$.

Second regime. $\text{comp} \equiv 1$, so $\chi = 0$.

Third regime. $\log \text{comp} = (L - r(L)) \log 2 = L \log 2 + o(L)$ by Lemma 4.7.

Fourth regime. $\log \text{comp} = (n(L) - j(L)) \log 2$ with $n(L)$ the number of full letters, so $n(L)/L \rightarrow f_{\text{full}}$, and $j(L)$ a terminal run, hence $o(L)$ by Lemma 4.7. The limit is $f_{\text{full}} \log 2$.

Fifth regime. Let $m(L)$ be the last domino place and $i^*(L)$ the last gasket place at or before $m(L)$; both exist for large L because both frequencies are positive. Then $L - m(L)$ is a terminal run of gaskets, hence $o(L)$, and $m(L) - i^*(L)$ is the terminal run of $w_{1..m(L)}$, hence $o(m(L)) = o(L)$, both by Lemma 4.7. Write $T(L) = \text{fill}(A_{w_{1..i^*(L)-1}})$. By Lemma 4.8 below, $T(L) < \text{comp}(A_{w_{1..L}}) \leq 1 + \frac{3}{2}T(L)$, so $\log \text{comp} = \log T(L) + O(1)$. And

$$\begin{aligned} \log T(L) &= \sum_{i < i^*(L)} \log k_{w_i} = \sum_{i \leq L} \log k_{w_i} - O(L - i^*(L)) \\ &= L(f_p \log 2 + f_q \log 3) + o(L), \end{aligned}$$

since $L - i^*(L) = o(L)$ and each $\log k_{w_i} \leq \log 3$. Dividing by L gives the claim.

In all five cases the limit is a function of f alone. The final sentence follows: Thue-Morse, every periodic word, and every strictly ergodic word carrying both letters have letter frequencies, interior in the last two cases by minimality and in the first because both letters have frequency $\frac{1}{2}$; and a Bernoulli word has frequencies $(p, 1-p)$ almost surely by the strong law. $\square$

Lemma 4.8 (The sandwich). *Over a gasket-and-domino alphabet, let m be the last domino place of w , let i^* be the last gasket place at or before m , and put $T = \text{fill}(A_{w_{1..i^*-1}})$. Then $T < \text{comp}(A_w) \leq 1 + \frac{3}{2}T$. Moreover $\text{fill}(A_w)/T \in [6, 108]$ whenever no three consecutive letters of w are equal, and then $\text{comp}(A_w)/\text{fill}(A_w) \in (1/108, 5/12]$, with no side condition, since $T \geq 1$ always.*

Proof. By Theorem 3.1, $\text{comp}(A_w) = 1 + \sum_i \text{fill}(A_{w_{1..i-1}})$ over gasket places $i \leq m$. The largest term is the one at $i = i^*$, namely T , so $\text{comp}(A_w) > T$. If $i < i'$ are two consecutive gasket places at or before m then $\text{fill}(A_{w_{1..i'-1}})/\text{fill}(A_{w_{1..i-1}}) = \text{fill}(A_{w_{i..i'-1}}) \geq k_q = 3$, since w_i is a gasket

and the other letters of the block have fill at least 2. Reading the terms backwards from T they therefore decrease by factors of at least 3, so their sum is at most $T(1 + \frac{1}{3} + \frac{1}{9} + \dots) = \frac{3}{2}T$, giving $\text{comp}(A_w) \leq 1 + \frac{3}{2}T$.

For the second claim, the suffix $w_{i^*..L}$ has the shape $qp^\alpha q^\beta$ with $\alpha = m - i^* \geq 1$ and $\beta = L - m \geq 0$, because i^* is the last gasket place at or before m and m is the last domino place. So $\text{fill}(A_w)/T = 3 \cdot 2^\alpha 3^\beta$. If no three consecutive letters agree then $\alpha \leq 2$ and $\beta \leq 2$, so this lies in $[3 \cdot 2, 3 \cdot 4 \cdot 9] = [6, 108]$. Finally $1 + \frac{3}{2}T \leq \frac{5}{2}T$ as soon as $T \geq 1$, which holds for every word since a fill is at least 1, so $\text{comp} \leq \frac{5}{2}T$, so $\text{comp} / \text{fill} \leq (\frac{5}{2}T)/(6T) = \frac{5}{12}$, and $\text{comp} / \text{fill} > T/(108T) = 1/108$. $\square$

4.4. Proof of Theorems 3.4 and 3.5.

Proof. Both are a walk through the five regimes with the weights of Theorem 3.3 in hand. Write $\phi_c = \log 2$ if c is a diagonal and 0 otherwise, so that $\Phi(f) = f_a \phi_a + f_b \phi_b$, and write ϵ_c for the exponent weight and $\log k_c$ for the fill weight. Since one frequency is free on the open simplex, χ equals the fill exponent at every interior f if and only if $\epsilon_c = \log k_c$ for both letters, and Φ is exact at every interior f if and only if $\epsilon_c = \phi_c$ for both letters.

First regime, 69 alphabets. $\epsilon_c = \log k_c$ for both letters, so all 69 saturate. And $\epsilon_c = \phi_c$ requires $\log k_c = \phi_c$ for each letter, which holds for a unit ($k = 1, \phi = 0$) and a diagonal ($k = 2, \phi = \log 2$) and fails for a domino ($k = 2, \phi = 0$), a gasket ($k = 3$) and the full tile ($k = 4$). So Φ is exact on the $\binom{6}{2} = 15$ alphabets both of whose letters are units or diagonals and refuted on the other 54.

Second regime, 12 alphabets. $\epsilon_a = \epsilon_b = 0$; the letters are dominoes, gaskets or the full tile, all of fill at least 2, so saturation fails on all 12. No letter is a diagonal, so $\phi_a = \phi_b = 0$ and Φ is exact on all 12.

Third regime, 4 alphabets. $\epsilon = \log 2 = \log k$ for both letters, so all 4 saturate; $\phi = 0 \neq \log 2$, so Φ is refuted on all 4.

Fourth regime, 4 alphabets. $\epsilon_{\text{domino}} = 0 \neq \log 2 = \log k_{\text{domino}}$, so none saturates; $\epsilon_{\text{full}} = \log 2 \neq 0 = \phi_{\text{full}}$, so Φ is refuted on all 4.

Fifth regime, 16 alphabets. $\epsilon = \log k$ for both letters, so all 16 saturate; $\phi = 0$ for both while $\epsilon \neq 0$, so Φ is refuted on all 16.

Adding, saturation holds on $69 + 4 + 16 = 89$ and fails on $12 + 4 = 16$; Φ is exact on $15 + 12 = 27$ and refuted on $54 + 4 + 4 + 16 = 78$.

For the strictly-between claim, an alphabet has $\Phi(f) < \chi(f) < \text{fill}$ exponent at interior f exactly when saturation fails and Φ is refuted, since otherwise one of the two inequalities is an equality. Saturation fails only in the second and fourth regimes, and Φ is exact throughout the second, so only the fourth regime survives: the 4 alphabets of a domino against the full tile, one class pair up to transposition. At $f = (\frac{1}{2}, \frac{1}{2})$ its exponent is $\frac{1}{2} \log 2$, against $\Phi = 0$ and a fill exponent of $\frac{1}{2} \log 2 + \frac{1}{2} \log 4 = \frac{3}{2} \log 2$.

Finally the wrong-shape statement of Theorem 3.5. On each of the 78, pick a vertex of the simplex at which the interior formula reads a positive

value. In the first regime with a refuting letter c the interior formula is the fill exponent, which reads $\log k_c > 0$ at the vertex $f = \delta_c$ for any letter of fill at least 2; in the third it reads $\log 2$ at both vertices; in the fourth it reads $\log 2$ at the full-tile vertex; in the fifth it reads $\log 2$ or $\log 3$ at the two vertices. The constant word at that vertex has rate 0, since its letter is not a diagonal in any of the cases listed. A continuous functional agreeing with χ in the interior takes the interior formula's value at that vertex by continuity, which is positive, so it misses the constant word. $\square$

4.5. Proof of Theorem 3.7.

Proof. Two properties of the Thue-Morse sequence do all the work. First, no three consecutive terms are equal: among any three consecutive indices there is an even index $2n$ whose successor $2n + 1$ is also among them, and $t_{2n} = t_n$ while $t_{2n+1} = 1 - t_n$, so those two differ. Second, the letter counts are exactly balanced: the prefix $w_{1..L}$ reads the indices $0, \dots, L - 1$, and pairing them as $(2k, 2k + 1)$ each pair contributes one of each letter, so a prefix of even length L has exactly $L/2$ of each and a prefix of odd length has $\lfloor L/2 \rfloor$ or $\lceil L/2 \rceil$; in both cases $|N_q(L) - L/2| \leq \frac{1}{2}$.

Fix a gasket-and-domino alphabet, a reading, and $L \geq 4$, which is what it takes for a gasket place and a domino place both to occur, so that m and i^* are defined. No lower bound on T is needed, and none is available: under the reading sending 0 to the gasket the prefix of length 4 is $qppq$, whose i^* is 1 and whose T is therefore $\text{fill}(A_\varepsilon) = 1$. By Lemma 4.8 and the first property, $\text{comp}(A_{w_{1..L}}) / \text{fill}(A_{w_{1..L}}) \in (1/108, 5/12]$, hence

$$\left| \log \text{comp}(A_{w_{1..L}}) - \log \text{fill}(A_{w_{1..L}}) \right| \leq \log 108.$$

By the second property, with N_q the gasket count and $N_p = L - N_q$ the domino count,

$$\begin{aligned} \left| \log \text{fill}(A_{w_{1..L}}) - \frac{L}{2} \log 6 \right| &= \left| (N_p - \frac{L}{2}) \log 2 + (N_q - \frac{L}{2}) \log 3 \right| \\ &= \left| N_q - \frac{L}{2} \right| \log \frac{3}{2} \leq \frac{1}{2} \log \frac{3}{2}, \end{aligned}$$

using $N_p - \frac{L}{2} = -(N_q - \frac{L}{2})$. Adding the two displays gives the certificate, and $\log 108 + \frac{1}{2} \log \frac{3}{2} = 4.884864\dots < 4.885$. Dividing by L and letting $L \rightarrow \infty$ gives $\chi = \frac{1}{2} \log 6$. Nothing in the argument names the alphabet beyond the fills 3 and 2, so it covers all 16 alphabets and both readings. $\square$

4.6. Proof of Theorem 3.11.

Proof. Over $\{3, 6\}$ the letter 6 is contact-free and the letter 3 is not, so the first regime of Theorem 3.1 applies and $\text{comp}(A_{w_{1..L}}) = \text{fill}(A_{w_{1..p(L)}}) = 2^{p(L)}$ with $p(L)$ the last place of $w_{1..L}$ carrying the letter 6, both letters having fill 2. In each of the three words the letter 6 occurs on a set of density 0, so $f = (1, 0)$.

For the constant word $p(L) = 0$ and $\text{comp} \equiv 1$, so $\chi = 0$. For the squares word $p(L) = \lfloor \sqrt{L} \rfloor^2$, and $\lfloor \sqrt{L} \rfloor^2/L \rightarrow 1$, since for L in $[n^2, (n+1)^2)$ the ratio is at least $n^2/((n+1)^2 - 1) \rightarrow 1$, so $\chi = \log 2$. For the powers-of-2 word $p(L) = 2^{\lfloor \log_2 L \rfloor}$, so the prefix rate in units of $\log 2$ is $2^k/L$ for $L \in [2^k, 2^{k+1})$, which equals 1 at $L = 2^k$ and decreases to just above $\frac{1}{2}$ at $L = 2^{k+1} - 1$, in steps of size at most $2^k/L - 2^k/(L+1) \leq 1/L$. Every point of $[\frac{1}{2}, 1]$ is therefore within $1/L$ of a value taken in the k -th block, so the accumulation set of the prefix rates is exactly $[\frac{1}{2}, 1]$ in units of $\log 2$, the upper rate is $\log 2$, the lower rate is $\frac{1}{2} \log 2$, and the limit does not exist.

Over $\{3, 7\}$ the fifth regime applies. For the word carrying 7 exactly at the powers of 2 and L in $(2^k, 2^{k+1}]$, the last domino place m is L unless L is a power of 2, in which case it is $L - 1$; in either case the last gasket place at or before m is $i^* = 2^k$ for $L > 2^k$. So $T = \text{fill}(A_{w_{1..2^k-1}}) = 2^{2^k-1-k} 3^k$ and Lemma 4.8 gives $\log_2 \text{comp} = 2^k + O(k)$ throughout the block. The prefix rate is therefore $2^k/L + O(k/L)$, and the argument of the previous paragraph applies verbatim. The constant word 3^∞ has $\text{comp} \equiv 1$ by Theorem 3.1 with no domino-free reading needed, and the squares word has $\log_2 \text{comp} = \lfloor \sqrt{L} \rfloor^2 + O(\sqrt{L})$ by the same sandwich, hence rate $\log 2$. All three words have $f = (1, 0)$.

Each of the three orbit closures is the closure of the shift orbit of a word whose non-constant letters have density 0, hence countable, with the point mass at the constant word as its only shift-invariant measure, so all three are uniquely ergodic. For the two non-constant words the closure properly contains the fixed point of the shift and so is not minimal; the first word is the constant word itself, whose orbit closure is a single point and is minimal. $\square$

Fact 4.9 (A stationary control, and the one rough family). $\text{comp}(A_{(7,3)^k}) = (6^k + 4)/5$, reading 2, 8, 44, 260, 1556, 9332, checked by `scripts/verify.py` to $k = 8$ against Theorem 3.1 and to $k = 4$ against the drawing. Its per-letter rate is $\frac{1}{2} \log 6$ again, from a periodic and therefore stationary word. And the fifth regime really does leave the smooth numbers of Corollary 4.10 behind: the largest count at length 8 over the alphabet $(3, 7)$ is $1094 = 2 \cdot 547$, checked by `scripts/verify.py` over all words of that length.

Corollary 4.10 (Every count off the fifth regime is 3-smooth). *On each of the 89 alphabets outside the fifth regime, every count $\text{comp}(A_w)$ has the form $2^\alpha 3^\beta$. A gasket against a domino is therefore the only family whose counts need a prime above 3.*

Proof. Read Theorem 3.1 regime by regime. In the first the count is $\text{fill}(A_{w_{1..p}})$, a product of fills drawn from $\{1, 2, 3, 4\}$, hence of the form $2^\alpha 3^\beta$; in the second it is 1; in the third it is 2^{L-r} ; in the fourth it is 2^{n-j} . Only the fifth regime produces a sum rather than a product, and only there can another prime appear. $\square$

Remark 4.11 (Words are not recoverable from pictures). Nothing above tries to read the word off the picture, and it could not: Kronecker factorisation of a binary matrix is unique at a fixed profile of factor sizes but not once the profile may vary, and a matrix can factor at profiles of different length and at factor sizes that are not prime. That is settled in [14], and it is why the object of study here is the alphabet and the word, not the tile.

5. REPRODUCIBILITY

Two scripts, both plain `python3` with nothing beyond the standard library, both run from the paper's directory, both using relative paths only.

`python3 scripts/verify.py` runs in about 15 seconds on a laptop and dies loudly, naming the word and both values, at the first disagreement. It checks:

- *Contacts multiply.* All 54,240 words of length 1 to 4 over the 15 codes: h and v of the drawn picture against the product of the per-letter values (Lemma 4.1).
- *The regime census.* That the five regimes partition the 105 alphabets as 69, 12, 4, 4, 16 (Theorem 3.1).
- *The closed forms against the drawing.* All 26,670 words of length at most 7 over all 105 alphabets, and all 4,080 words of length at most 8 over eight named alphabets, drawn cell by cell and counted by union-find over row runs (Fact 3.2).
- *The rate table.* For each of the 105 alphabets, the prefix rate of the closed form at length 4000, at two different letter frequencies, against $f_a\epsilon_a + f_b\epsilon_b$ from Theorem 3.3, to within 0.02.
- *The verdict counts.* The counts 89 and 16 of Theorem 3.4, 27 and 78 of Theorem 3.5, and that exactly 4 alphabets are strictly between and all four are a domino against the full tile; each re-derived alphabet by alphabet from the weights rather than copied.
- *Smoothness.* Corollary 4.10 re-checked on all words of length at most 8 over the 89 alphabets outside the fifth regime, and the largest count at length 8 over (3, 7) read as 1094 (Fact 4.9).
- *Thue-Morse.* The certificate of Theorem 3.7 at every length $4 \leq L \leq 2^{14}$ under both readings, the largest deviation 4.273459 against 4.884864, the four printed prefix rates of Fact 3.8 to twelve decimals, and the float track cross-checked against an exact big-integer evaluation of the closed form at $L = 4096$ and $L = 2^{14}$.
- *Saturation.* The minimum 0.0113766545 of comp / fill over $1 \leq L \leq 2^{14}$, its position above the proved floor 1/108, the exact rational maxima 43397/186624 and 151/648 at $L \geq 5$, and the sampled 0.2325367033 at $L = 4096$ (Fact 3.9).
- *The boundary words.* Over (3, 6), that the squares word has last-diagonal place exactly n^2 at $L = n^2$ and at $L = (n + 1)^2 - 1$ for $2 \leq n < 40$, and that the powers-of-2 word gives the exact rates

$4/7, 8/15, \dots, 8192/16383$; over $(3, 7)$, the three floats of Fact 3.12 and the monotone sweep of one block from 1.00123 to 0.50073 (Theorem 3.11).

- *The tripling word.* Its prefix-rate range $[0.4792, 1.4379]$ on $1024 \leq L \leq 4096$ (Conjecture 3.16).
- *The rank-2 cocycle.* That $\lambda M_{c_1} \cdots M_{c_L} \gamma$ reproduces the count on all 8,190 words of length at most 12 over $\{3, 6\}$, that the largest entry of M_3^L is $2^{L+1} - 1$ for $L \leq 24$, and that $\text{comp}(A_{3^L}) = 1$ throughout (Fact 3.13).
- *The stationary control.* $\text{comp}(A_{(7,3)^k}) = (6^k + 4)/5$ to $k = 8$, drawn to $k = 4$ (Fact 4.9).

`python3 scripts/figure.py` runs in well under a second and writes `figures/ledger.svg`, the two-plate picture the repository page embeds, and `figures/plates.tex`, the TikZ that Figs. 1 and 2 are drawn from. Every cell of both plates is computed: the witness pictures are rendered from Definition 2.2 and their piece counts are read off the drawing, and each of the 210 coloured cells of the square is a verdict recomputed from Theorem 3.1, as are the four counts in the caption. The script carries a verbatim copy of the drawing style shared by the papers in this collection.

Every numbered statement above is either proved, in Section 4 or in place as with Proposition 3.14, or re-checked by `scripts/verify.py`, and most are both. The exceptions are the three conjectures, which are labelled as such, and Remark 4.11, which is cited to [14] and not reproved here.

Open problem. The paper stops in four places and each is a real gap. It says nothing about a word whose letter frequencies fail to exist, and Conjecture 3.16 shows that the obvious weakening, positive lower density of both letters, is not enough: over $(3, 7)$ the tripling word's prefix rate still swings across a factor of three at length 4096 with no sign of narrowing, and whether it has a limit at all is unknown. It identifies the exponent but not the counts: the accumulation set of `comp` / `fill` along Thue-Morse is unidentified even though it is trapped in $(1/108, 5/12]$, and identifying it is a question about the attractor of two affine maps read along an aperiodic word, not about rates. It is confined to two letters, and the only route to a genuinely non-stationary result left open on this construction is an alphabet of three or more letters or a different order-sensitive observable, where nothing here applies. And it is confined to base 2 in the plane: at base q the alphabet is 2^{q^2} designs, the contact vocabulary of Definition 2.3 becomes a vocabulary of contact profiles rather than two integers, and neither the zero-contact cut nor the run recursion has been tried there.

ACKNOWLEDGMENTS

This paper was developed and verified in collaboration with Claude (Anthropic). The author takes sole responsibility for every claim.

REFERENCES

- [1] C. Mitchener, *Order Sensitivity of Kronecker Design Words: What the Perfect Shuffle Cannot See*, 2026. <https://github.com/carlomitchener/carlomitchener/tree/main/research/order-sensitivity-of-kronecker-words>
- [2] The On-Line Encyclopedia of Integer Sequences, *A010060: the Thue-Morse sequence*. <https://oeis.org/A010060>
- [3] L. L. Cristea and B. Steinsky, *Connected generalised Sierpinski carpets*, Topology and its Applications 157(7), 1157–1162, 2010. <https://doi.org/10.1016/j.topol.2010.02.005>
- [4] L. L. Cristea and B. Steinsky, *Mixed labyrinth fractals*, Topology and its Applications 229, 112–125, 2017. <https://doi.org/10.1016/j.topol.2017.06.022>
- [5] P. A. P. Moran, *Additive functions of intervals and Hausdorff measure*, Math. Proc. Cambridge Philos. Soc. 42(1), 15–23, 1946. <https://doi.org/10.1017/S0305004100022684>
- [6] D.-J. Feng, Z.-Y. Wen and J. Wu, *Some dimensional results for homogeneous Moran sets*, Science in China Ser. A 40(5), 475–482, 1997. <https://doi.org/10.1007/BF02896955>
- [7] L. Rempe-Gillen and M. Urbanski, *Non-autonomous conformal iterated function systems and Moran-set constructions*, Trans. Amer. Math. Soc. 368(3), 1979–2017, 2016. <https://doi.org/10.1090/tran/6490>
- [8] M. Barnsley, J. E. Hutchinson and O. Stenflo, *V-variable fractals: fractals with partial self similarity*, Advances in Mathematics 218(6), 2051–2088, 2008. <https://doi.org/10.1016/j.aim.2008.04.011>
- [9] M. Barnsley, J. E. Hutchinson and O. Stenflo, *V-variable fractals: dimension results*, Forum Mathematicum 24(3), 445–470, 2012. <https://doi.org/10.1515/form.2011.075>
- [10] Y. Smilansky and Y. Solomon, *Multiscale substitution tilings*, Proc. London Math. Soc. 123(6), 517–564, 2021. <https://doi.org/10.1112/plms.12404>
- [11] V. Berthe and V. Delecroix, *Beyond substitutive dynamical systems: S-adic expansions*, RIMS Kokyuroku Bessatsu B46, 81–123, 2014. <https://arxiv.org/abs/1309.3960>
- [12] H. Furstenberg and H. Kesten, *Products of random matrices*, Annals of Mathematical Statistics 31(2), 457–469, 1960. <https://doi.org/10.1214/aoms/1177705909>
- [13] R. Jungers, *The Joint Spectral Radius: Theory and Applications*, Springer LNCIS 385, 2009. <https://doi.org/10.1007/978-3-540-95980-9>
- [14] J. Voet and L. De Novellis, *Identifying Kronecker product factorizations*, arXiv:2510.25292, 2025. <https://arxiv.org/abs/2510.25292>

MRLYPROD, INC.

Email address: carlo.mitchener@gmail.com