On the entropy and letter frequencies of ternary square-free words

On the Entropy and Letter Frequencies of Ternary Square-Free Words

Christoph Richard
Institut für Mathematik
Universität Greifswald
Jahnstr. 15a
17487 Greifswald, Germany
richard@uni-greifswald.de

Uwe Grimm
Applied Mathematics Department
The Open University
Walton Hall
Milton Keynes MK7 6AA, UK
u.g.grimm@open.ac.uk

19 March 2003
Keywords: Combinatorics on words, square-free words
Mathematics Subject Classifications: 68R15, 05A15

Abstract

We enumerate all ternary length-ℓ square-free words, which are words avoiding squares of words up to length ℓ, for $\ell \leq 24$. We analyse the singular behaviour of the corresponding generating functions. This leads to new upper entropy bounds for ternary square-free words. We then consider ternary square-free words with fixed letter densities, thereby proving exponential growth for certain ensembles with various letter densities. We derive consequences for the free energy and entropy of ternary square-free words.

1 Introduction

The interest in the combinatorics of pattern-avoiding [3, 2, 8], in particular of power-free words, goes back to work of Axel Thue in the early 20th century [35, 36]. The celebrated Prouhet-Thue-Morse sequence, defined by a substitution rule $a \rightarrow ab$ and $b \rightarrow ba$ on a two-letter alphabet $\{a, b\}$, proves the existence of infinite cube-free words in two letters a and b.

Here, a word of length n is a string of n letters from a certain alphabet Σ, an element of the language $\mathcal{L}(n) = \Sigma^n$ of n-letter words in Σ. The union

$$\mathcal{L} \;=\; \bigcup_{n\geq 0} \mathcal{L}(n) \;=\; \Sigma^{\mathbb{N}_0} \tag{1}$$

is the language of all words in the alphabet Σ. It is a monoid, with concatenation of words as operation, and with the empty word λ of zero length as neutral element [22].

A word w is called *square-free* if $w = xyyz$, with words x, y and z, implies that $y = \lambda$ is the empty word, and cube-free words are defined analogously. So square-free words are characterised by the property that they do not contain an adjacent repetition of any subword.

It is easy to see that there are only a few square-free words in two letters, these are the empty word λ, the two letters a and b, the two-letter words ab and ba, and, finally, the three-letter words aba and bab. Appending any letter to those two words inevitably results in a square, either of a single letter, or of one of the square-free two-letter words.

However, there do exist infinite ternary square-free words, i.e., square-free words on a three-letter alphabet. In fact, the number s_n of ternary square-free words of length n grows exponentially with n. Denoting set sets of ternary square-free words of length n by $\mathcal{A}_n$, we have

$$\begin{aligned}
\mathcal{A}_0 &= \{\lambda\}, \\
\mathcal{A}_1 &= \{a,b,c\}, \\
\mathcal{A}_2 &= \{ab,ac,ba,bc,ca,cb\}, \\
\mathcal{A}_3 &= \{aba,abc,aca,acb,bab,bac,bca,bcb,cab,cac,cba,cbc\},
\end{aligned} \tag{2}$$

and so on. So $s_0 = 1$, $s_1 = 3$, $s_2 = 6$, $s_3 = 12$, and so on, see [1] and [12] where the values of s_n for $n \leq 90$ and $91 \leq n \leq 110$ are tabulated, respectively. In [29], the sequence s_n is listed as A006156 (formerly M2550).

In this article, we consider ternary square-free words [35, 36, 38, 25, 3, 4, 5, 11, 22, 28, 21, 27, 18, 1, 10, 24, 12, 9, 32]. We are interested in the asymptotic growth of the sequence s_n. We use a series of generating functions for a truncated square-freeness condition and conjecture the presence of a natural boundary at the radius of convergence. We also consider the frequencies of letters in ternary square-free words and derive upper and lower bounds. We prove exponential growth for certain ensembles of ternary square-free words with fixed letter frequencies. We use methods of statistical mechanics [17] to prove that, subject to a plausible regularity assumption on the free energy of ternary square-free words, the maximal exponential growth occurs for words with equal mean letter frequencies, where we average over all square-free words. Some of our results are based on extensive exact enumerations of square-free ternary words of length $n \leq 110$ [12] and on constructions of generalised Brinkhuis triples [11, 12].

2 Ternary square-free words

Denote the number of ternary square-free words by s_n and the corresponding generating function by $S(x)$,

$$S(x) \;=\; \sum_{n=0}^{\infty} s_n\, x^n\,. \tag{3}$$

Since the language of ternary square-free words is subword-closed, we conclude that the sequence s_n is submultiplicative,

$$s_{n+m} \;\leq\; s_n\, s_m\,. \tag{4}$$

A standard argument, compare [1, Lemma 1] and [17, Lemma A.1], shows that this guarantees that the limit $\mathcal{S} := \lim_{n\to\infty} \frac{1}{n}\log s_n$, also called the *entropy*, exists. Bounds for the limit have been obtained in a number of investigations [5, 4, 11, 10, 24, 12, 32], which give

$$1.1184 \;\approx\; 110^{1/42} \;\leq\; \exp(\mathcal{S}) \;<\; 1.30201064\,, \tag{5}$$

but the exact value is unknown. The lower bound implies an exponential growth of s_n with n. The behaviour of the subleading corrections to the exponential growth is not understood.

One of the authors computed the numbers s_n for $n \leq 110$ [12]. Assuming an asymptotic growth of the numbers s_n of the form

$$s_n \;\sim\; A\, x_c^{-n}\, n^{\gamma-1} \qquad (n \to \infty)\,, \tag{6}$$

we used differential approximants [15] of first order to get estimates of the critical point $x_c = \exp(-\mathcal{S})$, the critical exponent γ and the critical amplitude A. We obtain

$$A \;=\; 12.72(1)\,, \qquad x_c \;=\; 0.768189(1)\,, \qquad \gamma \;=\; 1.0000(1)\,, \tag{7}$$

where the number in the bracket denotes the (estimated) uncertainty in the last digit. The value of γ, also found in [24], suggests a simple pole as dominant singularity of the generating function at $x = x_c$. Numerical analysis indicates the presence of a natural boundary, a topic which we considered further by computing approximating generating functions $S^{(\ell)}(x)$, which count the number of words which contain no squares of words of length $\leq \ell$.

3 Generating functions

We call a word $w \in \mathcal{L}$ *length-ℓ square-free* if $w = xyyz$, with $x, z \in \mathcal{L}$ and $y \in \bigcup_{n=0}^{\ell} \mathcal{L}(n)$, implies that y is the empty word λ. In other words, w does not contain the square of a word of length $\leq \ell$.

Denote the number of ternary length-ℓ square-free words of length n by $s_n^{(\ell)}$. Clearly, $\ell' > \ell$ implies $s_n^{(\ell')} \leq s_n^{(\ell)}$, because at least the same number of words are excluded. On the other hand, we have $s_n^{(\ell')} = s_n^{(\ell)} = s_n$ for $n < 2\ell$. Thus, by considering larger and larger squares ℓ, we approach the case of square-free words.

We define corresponding generating functions

$$S^{(\ell)}(x) \;=\; \sum_{n=0}^{\infty} s_n^{(\ell)}\, x^n \tag{8}$$

for the number of ternary length-ℓ square-free words. These generating functions are rational functions of the variable x which can be calculated explicitly, at least for small values of ℓ, see [24] where the computation is explained in detail. The first few generating functions are

$$
\begin{aligned}
S^{(0)}(x) &= \frac{1}{1-3x}\,,\\
S^{(1)}(x) &= \frac{1+x}{1-2x}\,,\\
S^{(2)}(x) &= \frac{1+2x+2x^2+3x^3}{1-x-x^2}\,,\\
S^{(3)}(x) &= \frac{1+3x+6x^2+11x^3+14x^4+20x^5+20x^6+21x^7+12x^8+6x^9(1-x-x^2-x^3-x^4)}{1-x^3-x^4-x^5-x^6}\,.
\end{aligned}
$$

We computed the generating functions $S^{(\ell)}(x)$ explicitly for $\ell \leq 24$. The functions are available as Mathematica code [37] at [14]. Note that some generating functions agree; for instance, $S^{(4)}(x) = S^{(5)}(x)$. The reason is that, going from $\ell = 4$ to $\ell = 5$, no "new" squares arise; in other words, all squares of square-free words of length 5 already contain a square of a word of smaller length.

The radius of convergence $x_c^{(\ell)} \leq x_c$ of the series defining the generating function $S^{(\ell)}(x)$ is determined by a pole in the complex plane located closest to the origin, thus by a zero of the denominator polynomial of smallest modulus. Due to Pringsheim's theorem [30, Sec. 7.21], a real and positive such zero exists. Note that the zeros of the numerator and denominator are mutually exclusive, because the do not contain common polynomial factors.

The values $x_c^{(\ell)}$ are given in Table 1, together with the degrees d_{num} and d_{den} of the polynomials in the numerator and in the denominator which both grow with ℓ. Thus, with growing length ℓ, the generating functions $S^{(\ell)}(x)$ have an increasing number of zeros and poles. The patterns of zeros and poles appear to accumulate in the complex plane close to the unit circle around the origin; and comparing the patterns for increasing ℓ one might be tempted to the plausible conjecture that the poles approach the unit circle in the limit as $\ell \to \infty$. However, there appear to be some oscillations in the patterns close to the real line, and at present we dot not have any argument why the poles should accumulate on the unit circle.

The values $x_c^{(\ell)}$ in Table 1 approach x_c from below, so they yield upper bounds on the exponential growth constant $\mathcal{S} = -\log(x_c)$. The upper bound quoted in equation (5) above was given in [24] on the basis of an estimate for $x_c^{(23)}$ obtained via the series expansion of $S^{(23)}(x)$. Our value for $x_c^{(23)}$, based on the complete evaluation of the generating function $S^{(23)}(x)$, is contained in Table 1; it confirms the bound of Noonan and Zeilberger [24]. The value for $\ell = 24$ slightly improves the upper bound.

Theorem 1. *The entropy $\mathcal{S}$ of ternary square-free words is bounded as $\mathcal{S} \leq -\log(x_c^{(24)})$, which gives* $\exp(\mathcal{S}) < 1.30193812 < 1/x_c^{(24)}$. □

Table 1: Degrees d_{num} and d_{den} of the numerator and denominator polynomials of the generating functions $S^{(\ell)}(x)$, respectively, and the numerical values of the radius of convergence $x_c^{(\ell)}$.

ℓ	d_{num}	d_{den}	$x_c^{(\ell)}$
0	0	1	0.333 333 333
1	1	1	0.500 000 000
2	3	2	0.618 033 989
3	5	3	0.682 327 804
4, 5	13	6	0.724 491 959
6, 7	27	15	0.750 653 202
8, 9, 10	38	19	0.757 826 433
11	81	58	0.762 463 266
12	143	106	0.765 262 611
13, 14	184	145	0.766 784 948
15	209	170	0.767 006 554
16, 17	217	178	0.767 136 379
18	441	380	0.767 542 044
19	644	594	0.767 752 831
20	968	890	0.767 887 486
21	1003	925	0.767 896 727
22	1436	1337	0.767 974 175
23	1966	1872	0.768 042 881
24	2905	2787	0.768 085 659

The complete set of poles of the generating function $S^{(24)}(x)$ is shown in Fig. 1. The pattern looks very similar for other values of ℓ. This suggests that, in the limit as ℓ becomes infinite, which corresponds to the generating function $S(x)$ of ternary square-free words, the poles accumulate close to the unit circle. This corroborates the conjecture that $S(x)$ has a natural boundary.

4 Square-free words with fixed letter frequencies

We now consider the letter statistics of ternary square-free words. Denote the number of occurrences of the letter a in a ternary square-free word w_n of finite length n by $a(w_n)$. Clearly, the frequency of the letter a in w_n is $0 \leq a(w_n)/n \leq 1$. For an infinite ternary square-free word w, letter frequencies do not generally exist. Consider sequences $\{w_n\}$ of n-letter subwords containing arbitrarily long words. We define upper and lower frequencies $f_a^+ \geq f_a^-$ by $f_a^+ := \sup_{\{w_n\}} \limsup_{n\to\infty} a(w_n)/n$ and $f_a^- := \inf_{\{w_n\}} \liminf_{n\to\infty} a(w_n)/n$, where we take the supremum and infimum over all sequences $\{w_n\}$. We can also compute

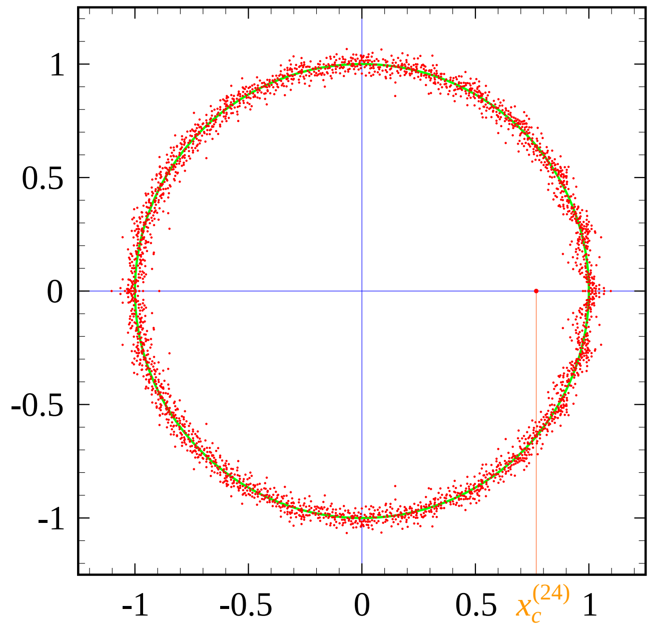

Figure 1: Pattern of poles of the generating functions $S^{(24)}(x)$ in the complex plane. The poles (red) accumulate along the unit circle (green). The isolated pole at $x_c^{(24)}$ on the real positive axis determines the radius of convergence.

these from $a_n^+ = \max_{w_n \subset w} a(w_n)$ and $a_n^- = \min_{w_n \subset w} a(w_n)$ by $f_a^\pm = \lim_{n\to\infty} a_n^\pm/n$, as these limits exist. This follows, for instance, from the subadditivity of the sequences $\{a_n^+\}$ and $\{1 - a_n^-\}$. If the infinite word w is such that $f_a^+ = f_a^- =: f_a$, we call f_a the *frequency* of the letter a in w. In general, $f_a^+ > f_a^-$, and letter frequencies do not exist, see also the discussion below.

However, we can derive bounds on the upper and lower letter frequencies f_a^+ and f_a^-. Denote the number of ternary square-free words of length n which contain the letter a exactly k times by $s_{n,k}$. Since there are no square-free words of length greater than three in two letters, a ternary square-free word contains no gaps between letters a of length greater than three. This implies $s_{n,k} = 0$ for $k < n/4$ or $k > n/2$, since the minimal number of letters b and c is, by the same argument, equal to $k = n/2$. By counting the number $s_{n,k}$ of ternary square-free words with a given number k of letters a, we can sharpen these bounds. Clearly, for fixed k, there are numbers $n_{\min}(k)$ and $n_{\max}(k)$ such that $s_{n,k} = 0$ for $n < n_{\min}(k)$ and $n > n_{\max}(k)$. This means that any ternary square-free word of length $(m+1)n_{\max}(k) \geq n > mn_{\max}(k)$, for any integer m, contains at least $mk + 1$ letters a, so the frequency of the letter a is bounded from below by $(mk+1)/(mn_{\max}(k)+1)$, which becomes $k/n_{\max}(k)$ as m tends to infinity. Similarly, any word of length $mn_{\min}(k) > n \geq (m-1)n_{\min}(k)$ contains at most $mk-1$ letters a. Thus we obtain an upper limit of $(mk-1)/(mn_{\min}(k) - 1)$, which becomes $k/n_{\min}(k)$ as m tends to infinity. We computed $n_{\max}(k)$ for $k \leq 31$ and $n_{\min}(k)$ for $k \leq 40$; the strongest bounds are derived from $n_{\max}(31) = 117$ and $n_{\min}(39) = 97$, which yield lower and upper bounds $31/117 \approx 0.265$ and $39/97 \approx 0.402$, respectively, for the frequency of a single letter in an infinite ternary square-free word. This gives

Theorem 2. *The upper and lower frequencies $f^{\pm}$ of a given letter in an infinite ternary square-free word are bounded by* $0.265 \approx 31/117 \leq f^{-} \leq f^{+} \leq 39/97 \approx 0.402$. □

Remark. In fact, there is a recent, stronger result for the lower frequency [33]. The minimum frequency $f^{-}_{\min}$ is bounded from below and above by [33]

$$0.274649 \approx 1780/6481 \leq f^{-}_{\min} \leq 64/233 \approx 0.274678\,,$$

compare also similar treatments for binary power-free words [19, 20]. The upper bound can be sharpened to $f^{+} \leq 469/1201 \approx 0.390508$ [34].

It is easy to see that the *mean letter frequency* of any given letter in the set of ternary square-free words is $1/3$. This is a consequence of symmetry under permutation of letters. Indeed, the symmetric group S_3 acts on any square-free word w by permutation of the three letters, and the set of square-free words of a given length is a disjoint union of orbits under this action. Each orbit consists of a square-free word and its images under permutation of letters, and each letter has the same mean frequency on this orbit. So, for each orbit, the mean frequency of any given letter is $1/3$, thus also for the set of all ternary square free words of any given length, or indeed for the set of all ternary square free words.

We now want to show that there exist ternary square-free words of infinite length with well-defined letter frequencies for the case $f_a = f_b = f_c = 1/3$ and for some cases where not all letter are equally frequent. In fact, we are going to prove not just that, but that there are exponentially many such words, so the growth rate for words of fixed frequencies, at least for the cases considered below, is positive. This can be done in a similar fashion as the proofs that the number of ternary square-free words grow exponentially [5, 4, 11, 10, 24, 12, 32]. These proofs are based on Brinkhuis triple pairs [5, 4, 11, 10, 24] and their generalisations [11, 12, 32]. We briefly sketch the argument here, see [5, 4, 11, 10, 24, 12, 32] for details.

The argument is based on square-free morphisms [6, 7]. Here, we immediately consider the generalised version of [11, 12]. Assume that we have a set of substitution rules

$$a \;\rightarrow\; \begin{cases} w_a^{(1)} \\ w_a^{(2)} \\ \vdots \\ w_a^{(k)} \end{cases} \qquad b \;\rightarrow\; \begin{cases} w_b^{(1)} \\ w_b^{(2)} \\ \vdots \\ w_b^{(k)} \end{cases} \qquad c \;\rightarrow\; \begin{cases} w_c^{(1)} \\ w_c^{(2)} \\ \vdots \\ w_c^{(k)} \end{cases} \tag{9}$$

where $w_a^{(j)}$, $w_b^{(j)}$ and $w_c^{(j)}$, $1 \leq j \leq k$, are ternary square-free words of equal length m. Starting from any ternary square-free word w of length n, consider the set of all words of length mn obtained by substituting each letter, choosing independently one of the k

words from the lists above. A *generalised Brinkhuis triple* is defined as a set of substitution rules (9) such that all these words of length mn are square-free, for any choice of w. This immediately implies that the number of square-free words grows at least as $k^{1/(m-1)}$, see [12, Lemma 2]. In the case $k=1$, this reduces to a usual substitution rule without any freedom; in this case, it only proves existence of infinite words, not exponential growth of the number of words with length.

In [12], a special class of generalised Brinkhuis triples was considered, and triples up to length $m=41$ with $k=65$ were obtained. This was recently improved to $m=43$ and $k=110$ in [32], yielding the lower bound of (5).

What about the letter frequencies? In general, the words $w_a^{(j)}$ that replace a will have different letter frequencies, and in this case it is easy to see that not all the infinite words obtained by repeated substitution will have well-defined letter frequencies. However, we can say something about letter frequencies if we consider generalised Brinkhuis triples where all words $w_a^{(j)}$, $1 \le j \le k$, have the *same* letter frequencies, and analogously for the words $w_b^{(j)}$, $1 \le j \le k$, and $w_c^{(j)}$, $1 \le j \le k$. In this case, regardless of our choice of words in the substitution process, we obtain words with well-defined letter frequencies, precisely as in the case of a standard substitution rule. Denoting the number of letters a, b and c in any of the words $w_a^{(j)}$ by n_a^a, n_a^b and n_a^c, respectively, with $n_a^a + n_a^b + n_a^c = m$, and analogously for $w_b^{(j)}$ and $w_c^{(j)}$, we can summarise the letter-counting for the generalised Brinkhuis triple in a 3×3 substitution matrix

$$M \;=\; \begin{pmatrix} n_a^a & n_b^a & n_c^a \\ n_a^b & n_b^b & n_c^b \\ n_a^c & n_b^c & n_c^c \end{pmatrix}\,. \tag{10}$$

In general, all entries of this matrix are positive integers, because there are no square-free words of length $m>3$ with only two letters. The (right) Perron-Frobenius eigenvector is thus positive, and its components encode the letter frequencies of the infinite words obtained by repeated application of the substitution rules. The Perron-Frobenius eigenvalue is m, because $(1,1,1)$ is a left eigenvector with eigenvalue m.

As mentioned previously, the generalised Brinkhuis triples considered in [12] do not have the property that the letter frequencies of the substitution words coincide. However, if we have a generalised Brinkhuis triple, any subset of substitutions also forms a triple, because all we do is restricting to a subset of words which still are square-free. So by looking at the triples of [12] and selecting suitable subsets of substitutions, we can use the same arguments to prove exponential growth of words with fixed letter frequencies.

4.1 Equal letter frequencies

Let us first consider the case of equal frequencies $f_a = f_b = f_c = 1/3$. We note that the special Brinkhuis triples of [12] had the additional property that $w_b^{(j)} = \sigma(w_a^{(j)})$ and $w_c^{(j)} = \sigma^2(w_a^{(j)})$, where σ is the permutation of letters defined by $\sigma(a)=b$ and $\sigma(b)=c$. If we select a subset of the words replacing a such that they have the same numbers of letters n_a^a, n_a^b and n_a^c, the substitution matrix for the corresponding triple consisting of those words and their images under σ is

$$M \;=\; \begin{pmatrix} n_a^a & n_a^c & n_a^b \\ n_a^b & n_a^a & n_a^c \\ n_a^c & n_a^b & n_a^a \end{pmatrix} \tag{11}$$

which is symmetric. Hence the right Perron-Frobenius eigenvector is $(1,1,1)^t$, and the letter frequencies are given by $f_a = f_b = f_c = 1/3$.

The simplest example is a Brinkhuis triple with $m = 18$ [12] (see also [24]) which explicitly given by

$$\begin{aligned} w_a^{(1)} &= abcacbacabacbcacba\,, \\ w_a^{(2)} &= abcacbcabacabcacba \;=\; \overline{w}_a^{(1)}\,, \end{aligned} \tag{12}$$

where $\overline{w}_a^{(1)}$ denotes $w_a^{(1)}$ read back-to-front, which thus has the same letter numbers $n_a^a = 7$, $n_a^b = 5$ and $n_a^c = 6$. So the number of ternary square-free words with letter frequencies $f_a = f_b = f_c = 1/3$ grows at least as $2^{1/17}$. By looking for the largest subsets of words with equal letter frequencies in the special Brinkhuis triples of [12], we can improve this bound. For $m = 41$, we find 30 words $w_a^{(j)}$ with letter numbers $n_a^a = 14$, $n_a^b = 13$ and $n_a^c = 14$, yielding a lower bound of $30^{1/40} \approx 1.08875$ for the exponential of the entropy. One of the two triples for $m = 43$ of [32] contains 39 words with $n_a^a = 14$, $n_a^b = 14$ and $n_a^c = 15$. This gives the following result.

Lemma 1. *The entropy $\mathcal{S}(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ of ternary square-free words with letter frequencies $f_a = f_b = f_c = 1/3$ is bounded from below via* $\exp[\mathcal{S}(\frac{1}{3},\frac{1}{3},\frac{1}{3})] \geq 39^{1/42} \approx 1.09115$. □

Remark. This bound can without doubt be improved, because the triples of [12] and [32] where not optimised to contain the largest number of words of equal frequency.

4.2 Unequal letter frequencies

What about words with non-equal letter frequencies? The following square-free substitution rule [38]

$$\begin{aligned} a &\longrightarrow cacbcabacbab \\ b &\longrightarrow cabacbcacbab \\ c &\longrightarrow cbacbcabcbab \end{aligned} \tag{13}$$

already shows that infinite words with unequal letter frequencies exist. In this case, the substitution matrix is

$$M = \begin{pmatrix} 4 & 4 & 3 \\ 4 & 4 & 5 \\ 4 & 4 & 4 \end{pmatrix}\,, \tag{14}$$

and the right Perron-Frobenius eigenvector with eigenvalues 12 is $(11, 13, 12)^t$. Thus this substitution leads to a ternary square-free word with letter frequencies $f_a = 11/36$, $f_b = 13/36$ and $f_c = 1/3$.

Can we show that, for some frequencies, there are exponentially many words? Indeed, for some examples we can find generalised Brinkhuis triples by choosing subsets of those given in [12]. Here, we restrict ourselves to a few examples.

Consider the two generating words

$$\begin{aligned} w_1 &= abcbacabacbcabacabcbacbcabcba \qquad (n_a = 10,\, n_b = 10,\, n_c = 9)\,, \\ w_2 &= abcbacabacbcacbacabcacbcabcba \qquad (n_a = 10,\, n_b = 9,\, n_c = 10)\,, \end{aligned} \tag{15}$$

of a Brinkhuis triple with $m = 29$ [12]. Choosing $w_a^{(1)} = w_1$, $w_a^{(2)} = \overline{w}_1$, $w_b^{(1)} = \sigma(w_1)$, $w_b^{(2)} = \sigma(\overline{w}_1)$, $w_c^{(1)} = \sigma^2(w_2)$ and $w_c^{(2)} = \sigma^2(\overline{w}_2)$, where again $\overline{w}$ denotes the words obtained by reversing w, we obtain a Brinkhuis triple with substitution matrix

$$M = \begin{pmatrix} 10 & 9 & 9 \\ 10 & 10 & 10 \\ 9 & 10 & 10 \end{pmatrix}. \tag{16}$$

The corresponding frequencies are $f = (f_a, f_b, f_c) = (\frac{9}{28}, \frac{10}{29}, \frac{271}{812})$, and the growth rate for this case is at least $2^{1/28}$.

Consider now two generating words

$$\begin{aligned} w_1 &= abcbacabacbabcabacabcacbcabcba \qquad (n_a = 11,\, n_b = 10,\, n_c = 9)\,, \\ w_2 &= abcbacabacbcabcbacabcacbcabcba \qquad (n_a = 10,\, n_b = 10,\, n_c = 10)\,, \end{aligned} \tag{17}$$

of a Brinkhuis triple with $m = 30$ [12]. Choosing $w_a^{(1)} = w_1$, $w_a^{(2)} = \overline{w}_1$, $w_b^{(1)} = \sigma(w_2)$, $w_b^{(2)} = \sigma(\overline{w}_2)$, $w_c^{(1)} = \sigma^2(w_\alpha)$ and $w_c^{(2)} = \sigma^2(\overline{w}_\alpha)$, where $\alpha \in \{1, 2\}$, we obtain two Brinkhuis triples with substitution matrices M_α given by

$$M_1 = \begin{pmatrix} 11 & 10 & 10 \\ 10 & 10 & 9 \\ 9 & 10 & 11 \end{pmatrix}, \qquad M_2 = \begin{pmatrix} 11 & 10 & 10 \\ 10 & 10 & 10 \\ 9 & 10 & 10 \end{pmatrix}. \tag{18}$$

The corresponding frequencies now are $f_1 = (\frac{10}{29}, \frac{271}{841}, \frac{280}{841})$ and $f_2 = (\frac{10}{29}, \frac{1}{3}, \frac{28}{87})$, and the growth rates for these examples are at least $2^{1/29}$.

Our next examples use the generating words

$$\begin{aligned} w_1 &= abcacbacabcbabcabacbcabcbacbcacba \qquad (n_a = 11,\, n_b = 11,\, n_c = 11)\,, \\ w_2 &= abcacbcabacabcacbabcbacabacbcacba \qquad (n_a = 12,\, n_b = 10,\, n_c = 11)\,, \end{aligned} \tag{19}$$

of a Brinkhuis triple with $m = 33$ [12]. Choosing as above $w_a^{(1)} = w_1$, $w_a^{(2)} = \overline{w}_1$, $w_b^{(1)} = \sigma(w_2)$, $w_b^{(2)} = \sigma(\overline{w}_2)$, $w_c^{(1)} = \sigma^2(w_\alpha)$ and $w_c^{(2)} = \sigma^2(\overline{w}_\alpha)$, where $\alpha \in \{1, 2\}$, we obtain two Brinkhuis triples, this time with substitution matrices M_α given by

$$M_1 = \begin{pmatrix} 11 & 11 & 11 \\ 11 & 12 & 11 \\ 11 & 10 & 11 \end{pmatrix}, \qquad M_2 = \begin{pmatrix} 11 & 11 & 10 \\ 11 & 12 & 11 \\ 11 & 10 & 12 \end{pmatrix}. \tag{20}$$

The corresponding frequencies now are $f_1 = (\frac{1}{3}, \frac{11}{32}, \frac{31}{96})$ and $f_2 = (\frac{331}{1024}, \frac{11}{32}, \frac{341}{1024})$. Here, the growth rate is at least $2^{1/32}$.

Finally, we give one example with a rather large deviation from equidistribution of letters. This uses three generating words

$$\begin{aligned} w_1 &= abcacbacabacbcabacabcacbcabacbcacba & (n_a = 13,\, n_b = 10,\, n_c = 12)\,, \\ w_2 &= abcacbcabacbabcbacabcbabcabacbcacba & (n_a = 12,\, n_b = 12,\, n_c = 11)\,, \\ w_3 &= abcacbacabacbcabacabcbabcabacbcacba & (n_a = 13,\, n_b = 11,\, n_c = 11)\,, \end{aligned} \tag{21}$$

of a Brinkhuis triple with $m = 35$ [12]. Choosing $w_a^{(1)} = w_1$, $w_a^{(2)} = \overline{w}_1$, $w_b^{(1)} = \sigma(w_2)$, $w_b^{(2)} = \sigma(\overline{w}_2)$, $w_c^{(1)} = \sigma^2(w_3)$ and $w_c^{(2)} = \sigma^2(\overline{w}_3)$, we obtain a Brinkhuis triple with substitution matrix

$$M = \begin{pmatrix} 13 & 11 & 11 \\ 10 & 12 & 11 \\ 12 & 12 & 13 \end{pmatrix}, \tag{22}$$

which yields frequencies $f = (\frac{1}{3}, \frac{16}{51}, \frac{6}{17})$. The growth rate is at least $2^{1/34}$.

To summarise, we proved the following.

Lemma 2. *The entropy of ternary square-free words with fixed letter frequency f_a is strictly positive for $f_a \in \{\frac{16}{51}, \frac{9}{28}, \frac{28}{87}, \frac{271}{841}, \frac{31}{96}, \frac{331}{1024}, \frac{280}{841}, \frac{341}{1024}, \frac{1}{3}, \frac{271}{812}, \frac{11}{32}, \frac{10}{29}, \frac{6}{17}\}$.* □

One should expect that the entropy is strictly positive for all letter frequencies f_a in an interval. However, it is not straightforward to show that by using substitutions of Brinkhuis triples with different letter frequencies. The reason is that, in general, the infinite words obtained by such substitutions do not have well-defined letter frequencies.

In the following sections, we are going to use methods from the theory of generating functions and convex analysis [31] which are often applied in the context of statistical mechanics [17]. The free energy of square-free words, which we will define below, is related to the entropy function of square-free words with fixed letter density, as follows from Proposition 2. An immediate consequence of the concavity of the entropy function is that the entropy is strictly positive for all frequencies $f_a \in (16/51, 6/17) \approx (0.3137, 0.3529)$, see below.

5 Free energy

Since the language of square-free words is subword closed, the numbers $s_{n,k}$ satisfy the submultiplicative inequality

$$s_{n+m,k} \;\leq\; \sum_{l=0}^{k} s_{n,l}\, s_{m,k-l}\,. \tag{23}$$

Consider the functions $s_n(q)$ defined by $s_n(q) \;=\; \sum_{k=0}^{n} s_{n,k}\, q^k$. These are polynomials in q of degree not larger than n. The submultiplicative inequality (23) implies for the functions

$s_n(q)$ that $s_{n+m}(q) \leq s_n(q)\, s_m(q)$ for $0 < q < \infty$. We are interested in the exponential growth rate of $s_n(q)$. To this end, define $F_n(q) := \frac{1}{n} \log s_n(q)$. The submultiplicative inequality yields [17, Lemma A.1] that the limit $F(q) := \lim_{n\to\infty} F_n(q)$ exists, and that $F(q) < \infty$ for $0 < q < \infty$. The function $F(q)$ is called the *free energy* of the model. More can be said about the properties of the free energy by using convexity arguments. These are largely independent of the underlying combinatorial model and are discussed in detail in [17, Sec. 2.1, App. B]. We obtain

Proposition 1. *The functions $F_n(q) = \frac{1}{n} \log s_n(q)$ of ternary square-free words are continuous, analytic and convex in $\log q$ in $(0, \infty)$. The free energy $F(q)$ of ternary square-free words*

$$F(q) \;=\; \lim_{n\to\infty} F_n(q) \tag{24}$$

exists and satisfies $F(q) < \infty$ for $q \in (0, \infty)$. Moreover, it is a convex function of $\log q$ for $q \in (0, \infty)$. If $F(q)$ is finite, its right- and left-derivatives exist everywhere in $(0, \infty)$, and they are non-decreasing functions of q. The function $F(q)$ is differentiable almost everywhere, and wherever the derivative $dF(q)/dq$ exists, it is given by $\lim_{n\to\infty} dF_n(q)/dq$. □

In the following, we will apply the results of the preceding section in order to derive bounds on the free energy. This will show that the free energy $F(q)$ is finite for $0 < q < \infty$. Using the above substitution rule (13) and the substitution rule given in [33], we first derive a lower bound on the free energy.

Lemma 3. *The free energy $F(q)$ is bounded from below by*

$$F(q) \;\geq\; \max\left\{ \frac{64}{233} \log q, \frac{13}{36} \log q \right\} . \tag{25}$$

Proof. Consider ternary square-free words w_n of length $n = 12k$, where $k \in \mathbb{N}$, generated by the substitution rule (13), with $w_1 = c$. Define $k_+(n) = 13n/36 + \delta_+(n)$, which denotes the number of letters of type a in w_n. Note that $\delta_+(n) = o(n)$. We have $s_n(q) \geq s_{n,k_+(n)} q^{k_+(n)}$. Taking the logarithm, dividing by n and performing the limit leads to $F(q) \geq \frac{13}{36} \log q$. The second part of the statement follows by the same argument with the substitution rule given in [33]. □

Remark. A weaker bound with $64/233$ replaced by $11/36 > 64/233$ may be derived using the substitution (13), where the role of a and b are interchanged.

We now turn to the question of an upper bound, which can be analysed using the bounds for letter frequencies obtained in [33, 34] or in Theorem 2.

Lemma 4. *The free energy $F(q)$ of ternary square-free words is bounded from above by*

$$F(q) \;\leq\; -\log x_c \,+\, \max\left\{ \frac{1780}{6481} \log q, \frac{469}{1201} \log q \right\} \tag{26}$$

where $x_c = \lim_{n\to\infty} s_n^{1/n} \approx 0.768189$ is the critical point of ternary square-free words.

Proof. Assume that $q \neq 1$. (The case $q = 1$ has been discussed in Section 2, where $F(1) = -\log x_c$ was proven.) Assume that B_n and A_n are numbers such that $s_{n,k} = 0$ for $k > B_n$ or $k < A_n$, $s_{n,B_n} > 0$, and $s_{n,A_n} > 0$. For $1 \neq q \in (0, \infty)$ we have the estimate

$$s_n(q) \;\leq\; s_n \sum_{A_n}^{B_n} q^k \;=\; s_n \, \frac{q^{B_n+1} - q^{A_n}}{q-1} \, . \tag{27}$$

Assume that $q > 1$. Taking the logarithm, dividing by n and performing the limit $n \to \infty$, this implies $F(q) \leq \log x_c + \epsilon_+ \log q$, where $\epsilon_+ = \limsup_{n\to\infty} B_n/n$. Note that $\epsilon_+ \leq 469/1201$, as follows from the bound given in [34]. A similar argument holds for $q < 1$, involving the lower bound A_n. From [33], we get the bound 1780/6481. Combining the two results, we get the inequality (26). ☐

Remark. A weaker bound with $(1780/6481, 469/1201)$ replaced by $(31/117, 39/97)$ follows from Theorem 2.

Define the two-variable generating function $S(x, q)$

$$S(x,q) \;=\; \sum_{n=0}^{\infty} \sum_{k=0}^{n} s_{n,k} \, x^n \, q^k \;=\; \sum_{n=0}^{\infty} s_n(q) \, x^n \, . \tag{28}$$

Denote the radius of convergence of $S(x, q)$ by $x_c(q)$. The curve $x_c(q)$ is called *critical curve*, and the plot of $x_c(q)$ in the xq-plane is called the *phase diagram* of the model. The free energy is related to the critical curve by

$$x_c(q)^{-1} \;=\; \lim_{n\to\infty} s_n(q)^{1/n} \;=\; e^{F(q)} \, . \tag{29}$$

We set $x_c = x_c(1)$ for the critical point of ternary square-free words. Bounds on the curve $x_c(q)$ can be derived from bounds on the free energy $F(q)$ as given above. This yields

$$x_c \, \min\{q^{-1780/6481}, q^{-469/1201}\} \;\leq\; x_c(q) \;\leq\; \min\{q^{-64/233}, q^{-13/36}\} \, . \tag{30}$$

The phase diagram is shown in Fig. 2. Using the series data from exact enumeration for length $n \leq 100$, we extrapolated the values of $x_c(q)$ for different values of q, using first order differential approximants [15]. The critical curve $x_c(q)$ is, within the analysed range of q, very close to the curve $x_c \, q^{-1/3}$, reflecting the fact that the values $k = k(n)$ where $s_{n,k} \neq 0$ are sharply concentrated around $k = \lfloor n/3 \rfloor$. For large values of q, such a form is, however, not compatible with the derived bounds on $x_c(q)$. Numerical analysis suggests that the leading divergence of $S(x, q)$ is a simple pole, which is approached uniformly in x and q. Thus, there is no indication that the nature of the singularity changes, in contrast to other examples from statistical mechanics, where such a change indicates a phase transition [17].

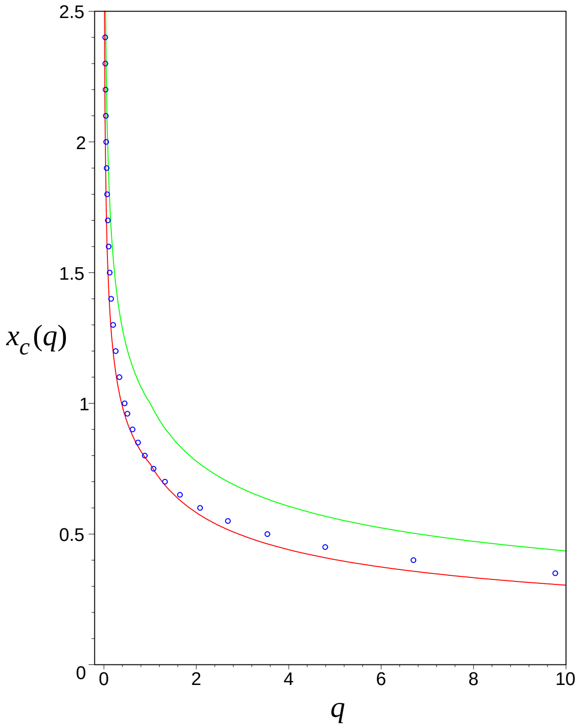

Figure 2: Phase diagram of ternary square-free words, as extrapolated from exact enumeration data (circles). Upper and lower bounds on $x_c(q)$ are drawn for comparison.

6 Entropy and symmetry

We now address the question of the number of ternary square-free words, where we fix the frequency of letters of type a. We consider the number of square-free words $s_{n,\lfloor \epsilon n \rfloor}$ in n letters with $\lfloor \epsilon n \rfloor$ occurrences of the letter a. The number ϵ may thus be regarded as the frequency of the letter a. We are interested in the exponential growth rate of $s_{n,\lfloor \epsilon n \rfloor}$. This leads to the question whether sequences of the form $\frac{1}{n} \log s_{n,\lfloor \epsilon n \rfloor}$ have a limit as $n \to \infty$, which we then call *entropy function* $P(\epsilon)$. It is related to the free energy $F(q)$ by a Legendre-Fenchel transform, as we will now show.

Note that there is a constant $K > 0$ such that $0 \le s_{n,k} \le K^n$ for each value of n and k. This follows from the existence of the entropy s of ternary square-free words. Note also that there exists a finite constant $C > 0$, and numbers A_n and B_n such that $s_{n,A_n} > 0$ and $s_{n,B_n} > 0$, and $s_{n,k} \ge 0$, when $0 \le A_n < k < B_n \le Cn$. This follows from the substitution rule (13). Take A_n and B_n such that $s_{n,k} = 0$ if $k < A_n$ or $k > B_n$. Define the numbers

$$\epsilon_+ \;=\; \limsup_{n\to\infty} \frac{B_n}{n}\,, \qquad \epsilon_- \;=\; \liminf_{n\to\infty} \frac{A_n}{n}\,. \tag{31}$$

From [33, 34] and the substitution rule (13), we have $0.361 \approx 13/36 \le \epsilon_+ \le 469/1201 \approx 0.391$ and $0.274649 \approx 1780/6481 \le \epsilon_- \le 64/233 \approx 0.274678$. Thus, the assumptions in [17, Thm. 3.19] are satisfied, and we obtain

Proposition 2. *The entropy function $P(\epsilon)$ of ternary square-free words exists in (ϵ_-, ϵ_+) and is defined by*

$$P(\epsilon) \;=\; \inf_{0<q<\infty} \{F(q) - \epsilon \log q\}\,. \tag{32}$$

Moreover, there is a sequence of integers $\{\sigma_n\}_{n=0}^{\infty}$ such that $\sigma_n = o(n)$ and the limit

$$P(\epsilon) \;=\; \lim_{n\to\infty} \frac{1}{n} \log s_{n,\lfloor \epsilon n\rfloor+\sigma_n} \tag{33}$$

exists and is finite and concave in (ϵ_-, ϵ_+). Lastly, note also that $\delta_n = \lfloor \epsilon n\rfloor + \sigma_n$ is the least value of k that maximises $s_{n,k}\,\tilde{q}^k$, where $\tilde{q}$ is that value of q where the infimum is taken in (32). $\square$

Remark. Together with Lemma 2, an immediate consequence of the concavity of the entropy function is that the entropy is strictly positive for all frequencies $\epsilon \in (16/51, 6/17) \approx (0.3137, 0.3529)$.

We consider now the question where the entropy function takes its maximum. To this end, we assume a special regularity condition on the free energy, whose validity is supported by the numerical analysis of the preceding section, see also the discussion in the conclusion.

Lemma 5. *Let $\epsilon \in (\epsilon_-, \epsilon_+)$. If $F(q) \in C^2(0,\infty)$, and if $F(q)$ is strictly convex in $\log q$, we have $P(\epsilon) \in C^2(\epsilon_-, \epsilon_+)$ for the entropy function, and it is given by*

$$P(\epsilon) \;=\; F\big(q(\epsilon)\big) \;-\; \epsilon\, \log q(\epsilon)\,, \tag{34}$$

where $q(\epsilon)$ is the unique positive solution of

$$\epsilon \;=\; q\,\frac{d}{dq} F(q)\,. \tag{35}$$

The entropy function $P(\epsilon)$ attains its global maximum at $q = 1$.

Proof. Since $F(q)$ is convex in $\log q$ and continuous, and $F(q) \geq \max\{\epsilon_- \log q, \epsilon_+ \log q\}$, the infimum in (32) occurs at a unique value $q = q(\epsilon) \in (0,\infty)$. Since $F(q) \in C^1(0,\infty)$, we obtain $\epsilon = qF'(q) = \frac{d}{d(\log q)} F(q)$ as an implicit equation for $q(\epsilon)$. This uniquely defines a positive function $q = q(\epsilon) \in C^1(\epsilon_-, \epsilon_+)$, since strict convexity of $F(q)$ and $F(q) \in C^2(0,\infty)$ implies $\frac{d^2}{d(\log q)^2} F(q) \neq 0$. We have explicitly $P'(\epsilon) = -\log q(\epsilon)$, which shows that $P(\epsilon) \in C^2(\epsilon_-, \epsilon_+)$, and $-\infty < P''(\epsilon) = -(\frac{d^2}{d(\log q)^2} F(q))^{-1} < 0$. This implies that $q = 1$ is a local maximum of $P(\epsilon)$. Due to the concavity of $P(\epsilon)$, it is the global maximum. $\square$

We note that at $q = 1$, the letter density $\epsilon = F'(1)$ is the mean letter density, which was determined above to be $\epsilon = 1/3$ by a symmetry argument. Thus, under the above regularity assumption, maximum entropy occurs at equal (mean) letter density $\epsilon_a = \epsilon_b = \epsilon_c = 1/3$. This is an example of the more general result that maximum entropy in occurs at points of maximum symmetry, see [26] for the concept of symmetry and its implications for the free energy and entropy of the combinatorial problem of random tilings, which is applicable in this case.

7 Conclusions

In this article, we considered the growth rate, or the entropy, of the set of ternary square-free words. By computing generating functions $S^{(\ell)}(x)$ for length-ℓ square-free words, where the condition of square-freeness is truncated at length ℓ, we verified an upper bound proposed in [24] and slightly improved it. The pattern of poles of these generating functions, and their behaviour as ℓ increases, points towards a natural boundary for the generating function $S(x)$.

The presence of a natural boundary in a model indicates that it cannot be solved exactly in terms of standard functions of mathematical physics, which obey linear differential equations with polynomial coefficients [16]. This would exclude, for ternary square-free words, an exact value for the entropy and the functional form of the free energy. It may even be difficult to prove the existence of a critical exponent, compare the related self-avoiding walk problem [17].

In the ternary alphabet, no letter is preferred by the condition of square-freeness. Thus, averaging over the entire sets of ternary square-free words, all letters appear equally often. However, in a single infinite word this need not be the case, indeed, the letter frequency may not be well-defined. However, one can derive limits on the minimum or maximum frequency of a given letter in an infinite ternary square-free words, and by explicitly constructing infinite words with given well-defined frequencies by means of substitution rules the minimum and maximum frequency can be bounded from above and below. We obtained limits from counting square-free words up to a certain length, sharper limits were given recently in [33, 34]. The bounds for the maximum frequency can certainly be further improved employing the approach of [19, 20, 33].

Lower bounds on the entropy are based on Brinkhuis triples and their generalisations. We used these to prove that, for a list of rational values, the entropy of the set of square-free words with a fixed letter frequency is strictly positive. Together with the concavity of the entropy function, obtained by methods of convex analysis and statistical mechanics, this led to the result that the entropy is strictly positive on an entire interval.

Concerning the entropy function, it would be interesting to extend the interval of strict positivity by providing sharper bounds from suitable substitution rules. This might be achievable by following and suitably modifying the approach taken in [19, 20, 33]. It is conceivable, albeit not necessary, that there exists a region of frequencies for which infinite square-free words exist, but the entropy vanishes, because the number of square-free words with that given letter frequency grows sub-exponentially. Such behaviour has been reported for kth-power-free binary square-free words with rational powers in the range $2 < k \leq 7/3$ [13].

Further, it is necessary to prove the validity of the regularity assumption on the free energy in Theorem 5. In contrast to other problems in statistical mechanics [17], there is no indication of a phase transition in the model of ternary square-free words, wherefore an analytic free energy is expected.

It would also be interesting to analyse the letter distribution using probabilistic methods. Similar examples lead, in an appropriate scaling limit, to Gaussian distribution functions [23].

8 Acknowledgements

We thank the Erwin Schrödinger International Institute for Mathematical Physics in Vienna for support during a stay in winter 2002/2003, where part of this work was done. CR would like to acknowledge financial support by the German Research Council (DFG). We are grateful to Jeffrey O. Shallit and to Bernd Sing for making us aware of references [33] and [38], respectively. Furthermore, we thank R. Kolpakov and Y. Tarannikov for providing us with their recent (partly unpublished) results. UG also wishes to acknowledge useful discussions with X. Y. Sun and D. Zeilberger.

References

[1] M. Baake, V. Elser and U. Grimm, The entropy of square-free words, *Mathematical and Computer Modelling* **26** (1997) 13–26.

[2] K. Baker, G. McNulty and W. Taylor, Growth problems for avoidable words, *Theoret. Comput. Sci.* **69** (1989) 319–345.

[3] D. R. Bean. A. Ehrenfeucht and G. F. McNulty, Avoidable patterns in strings of symbols, *Pacific J. Math.* **85** (1979) 261–294.

[4] F.-J. Brandenburg, Uniformly growing k^{th} power-free homomorphisms, *Theoret. Comput. Sci.* **23** (1983 69–82.

[5] J. Brinkhuis, Non-repetitive sequences on three symbols, *Quart. J. Math. Oxford* **34** (1983) 145–149.

[6] M. Crochemore, Sharp characterizations of squarefree morphisms, *Theoret. Comput. Sci.* **18** (1982) 221–226.

[7] M. Crochemore, Tests sur les morphismes faiblement sans carré, in *Combinatorics on Words*, ed. L. J. Cummings, Academic Press, Toronto (1983), pp. 63–89.

[8] J. D. Currie, Open problems in pattern avoidance, *Amer. Math. Monthly* **100** (1993) 790–793.

[9] J. D. Currie, There are circular square-free words of length n for $n \geq 18$, *The Electronic J. Comb.* **9** (2002) #N10.

[10] S. B. Ekhad and D. Zeilberger, There are more than $2^{n/17}$ n-letter ternary square-free words, *J. Integer Sequences* **1** (1998) 98.1.9.

[11] V. Elser, Repeat-free sequences, Lawrence Berkeley Laboratory report LBL-16632 (1983).

[12] U. Grimm, Improved bounds on the number of ternary square-free words, *J. Integer Sequences* **4** (2001) 01.2.7.

[13] U. Grimm, Counting power-free words in two letters, Poster presented at the workshop on *Aperiodic Order* in Oberwolfach, May 2001 (unpublished).

[14] U. Grimm, `http://mcs.open.ac.uk/ugg2/words/TernarySquareFreeGenFun.m` (2003).

[15] A. J. Guttmann, Asymptotic analysis of power series expansions, in *Phase Transitions and Critical Phenomena*, vol. 13, eds. C. Domb and J. Lebowitz, Academic Press, New York (1989), pp. 1–234.

[16] A. J. Guttmann, Indicators of solvability for lattice models, *Discr. Math.* **217** (2000) 167–189.

[17] E. J. Janse van Rensburg, *The Statistical Mechanics of Interacting Walks, Polygons, Animals and Vesicles*, Oxford University Press, New York (2000).

[18] Y. Kobayashi, Repetition-free words, *Theoret. Comput. Sci.* **44** (1986) 175–197.

[19] R. Kolpakov and G. Kucherov, Minimal letter frequencies in n-power-free binary words, in *Mathematical Foundations of Computer Science 1997*, Lecture Notes in Computer Science 1295, eds. I. Privara and P. Ružička, Springer, Berlin (1997), pp. 347–357.

[20] R. Kolpakov, G. Kucherov and Yu. Tarannikov, On repetition-free binary words of minimal density, *Theoret. Comput. Sci.* **218** (1999) 161–175.

[21] M. Leconte, k-th power-free codes, in *Automata on Infinite Words*, Lecture Notes in Computer Science 192, eds. M. Nivat and D. Perrin, Springer, Berlin (1985), pp. 172–187.

[22] M. Lothaire, *Combinatorics on Words*, Cambridge University Press, Cambridge (1983).

[23] G. Louchard, Probabilistic analysis of column-convex and directed diagonally-convex animals, *Random Struct. Algorithms* **11** (1996) 151–178.

[24] J. Noonan and D. Zeilberger, The Goulden-Jackson cluster method: Extensions, applications, and implementations, *J. Difference Eq. Appl.* **5** (1999) 355–377.

[25] P. A. B. Pleasants, Nonrepetitive sequences, *Proc. Cambr. Philos. Soc.* **68** (1970) 267–274.

[26] C. Richard, M. Höffe, J. Hermisson and M. Baake, Random tilings: concepts and examples, *J. Phys. A: Math. Gen.* **31** (1998) 6385–6408.

[27] P. Séébold, Overlap-free sequences, in *Automata on Infinite Words*, Lecture Notes in Computer Science 192, eds. M. Nivat and D. Perrin, Springer, Berlin (1985), pp. 207–215.

[28] R. O. Shelton, On the structure and extendability of square-free words, in *Combinatorics on Words*, ed. L. J. Cummings, Academic Press, Toronto (1983), pp. 101–118.

[29] N. J. A. Sloane and S. Plouffe, *The Encyclopedia of Integer Sequences*, Academic Press, San Diego (1995); see also the *On-Line Encyclopedia of Integer Sequences*, `http://www.research.att.com/~njas/sequences/`.

[30] R. P. Stanley, Generating functions, in *Studies in Combinatorics*, M.A.A. Studies in Mathematics, vol. 17 (1978), ed. G.-C. Rota, The Mathematical Association of America, pp. 100-141.

[31] R. P. Stanley, *Enumerative Combinatorics*, vol. 1, Cambridge University Press, Cambridge (1997); vol. 2, Cambridge University Press, Cambridge (1999).

[32] X. Y. Sun, New lower bound on the number of ternary square free words, *preprint* (2002); see also `http://www.math.temple.edu/~xysun/`.

[33] Yu. Tarannikov, The minimal density of a letter in an infinite ternary square-free words is $0.2746\ldots$, *J. Integer Sequences* **5** (2002) 02.2.2.

[34] Yu. Tarannikov, private communication.

[35] A. Thue, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. I, Mat. Nat. Kl. Christiana* **7** (1906) 1–22; reprinted in *Selected Mathematical Papers of Axel Thue*, eds. T. Nagell, A. Selberg, S. Selberg and K. Thalberg, Universitetsforlaget, Oslo (1977), pp. 139–158.

[36] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske Vid. Selsk. Skr. I, Mat. Nat. Kl. Christiana* **1** (1912) 1–67; reprinted in *Selected Mathematical Papers of Axel Thue*, eds. T. Nagell, A. Selberg, S. Selberg and K. Thalberg, Universitetsforlaget, Oslo (1977), pp. 413–477.

[37] S. Wolfram, *Mathematica book*, 4th edition, Cambridge University Press, Cambridge (1999).

[38] T. Zech, Wiederholungsfreie Folgen, *Z. angew. Math. Mech.* **38** (1958) 206–209.