Asymptotic density of brick and word codes

Małgorzata Moczurad and Włodzimierz Moczurad
Institute of Computer Science, Jagiellonian University
Nawojki 11, 30-072 Kraków, Poland
e-mail: {mmoczurad,wkm}@ii.uj.edu.pl

Abstract: Bricks are polyominoes with labelled cells. The problem whether a given set of bricks is a code is undecidable in general. We consider sets consisting of square bricks only. We have shown that in this setting, the codicity of small sets (two bricks) is decidable, but 15 bricks are enough to make the problem undecidable. Thus the step from words to even simple shapes changes the algorithmic properties significantly (codicity is easily decidable for words). In the present paper we are interested whether this is reflected by quantitative properties of words and bricks. We use their combinatorial properties to show that the proportion of codes among all sets is asymptotically equal to 1 in both cases.

1 Introduction

Let A be a finite alphabet. We use the usual notation of A^* to denote the free monoid over A, and X^* to denote the submonoid generated by $X \subseteq A^*$. A set of words $X \subseteq A^*$ is a code, if X^* is free over X, i.e., every word in X^* has a unique factorization over X.

A *brick* is a partial mapping $x : \mathbb{Z}^2 \to A$, where the domain of x ($\operatorname{dom} x$) is finite and connected (in the polyomino sense, i.e., points have to be unit distance apart to be considered adjacent). It can be viewed as a polyomino with its cells labelled with the symbols of A. If $|A| = 1$, there is an obvious natural correspondence between bricks and polyominoes. The set of all bricks over A is denoted by $A^{\bowtie}$.

Given a set of bricks $X \subseteq A^{\bowtie}$, the set of all bricks tilable with (translated copies of) the elements of X is denoted by $X^{\bowtie}$. Note that we do not allow rotations of bricks. $X \subseteq A^{\bowtie}$ is a *brick code,* if every element of $X^{\bowtie}$ admits exactly one tiling with the elements of X.

The *effective alphabet* of $X \subseteq A^{\bowtie}$ is the set of all symbols that appear on bricks in X, i.e., $\bigcup_{x \in X} x(\operatorname{dom} x)$. If $x \in A^{\bowtie}$ is a square brick, then by $\operatorname{len} x$ we denote the edge length of x, i.e., $\sqrt{|\operatorname{dom} x|}$.

The problem whether a given set of bricks (even polyominoes) is a code is undecidable in general. In contrast, polynomial-time algorithms exist for (finite) sets of words, cf. [3, 14]. The general problem of counting polyominoes or bricks of a given size is hard; e.g. no exact formula or generating function is known for the sequence $p_n = |\{x \in A^{\bowtie} : \|x\| = n\}|$, cf. [16]. On the other hand, the problem is trivial for words.

The undecidability of codicity testing for bricks can be proved by reduction from the Wang tilabilty problem, see [1, 2, 12]. The problem is open for two-element sets. In this paper we consider sets consisting of square bricks only. We have shown that in this setting, the codicity of small sets (two bricks) is decidable, but 15 bricks are enough to make the problem undecidable; see [9, 10]. Thus, with bricks restricted to squares, the problem remains algorithmically hard. The combinatorial properties, however, are now much easier.

We use the following counting principle, cf. [11, 17]: Given a set of objects, choose those of size n and count those that have a desired property. We are interested in the proportion of objects with the desired property (codicity in our case) as n tends to infinity.

More formally, let $\mathcal{F}$ be the set of objects and let $\mathcal{A} \subseteq \mathcal{F}$ be the set of objects having the desired property. The *asymptotic density* (or *asymptotic probability*) $\mu_{\mathcal{F}}(\mathcal{A})$ of $\mathcal{A}$ in $\mathcal{F}$ is defined as

$$\mu_{\mathcal{F}}(\mathcal{A}) = \lim_{n\to\infty} \frac{|\{X \in \mathcal{A} : \|X\| = n\}|}{|\{X \in \mathcal{F} : \|X\| = n\}|},$$

where $\|X\|$ denotes the size of X.

The existence and value of $\mu_{\mathcal{F}}(\mathcal{A})$ depend on the choice of $\|\cdot\|$. If $\mu_{\mathcal{F}}(\mathcal{A})$ exists, it is within $[0, 1]$. Note that μ is not a probability in the classical sense since the enumerable additivity axiom does not hold.

For the sake of clarity, in the sequel we assume a two-element alphabet. All results, however, can be easily generalized to any $|A| \geq 2$. The results become trivial when $|A| = 1$.

2 Counting brick codes

We use the following size measure for sets of bricks. Let $\mathcal{B}_k \subset \mathcal{P}(A^{\bowtie})$ denote the family of all sets containing k squares; $\mathcal{K}_k \subseteq \mathcal{B}_k$ will denote the family of codes containing k squares. Define the size of a set $X \in \mathcal{B}_k$ as the size of the largest square in X, i.e., $\|X\| = \max\{\operatorname{len} x : x \in X\}$.

We define the numbers $B_{k,=n}$, $B_{k,<n}$ and $B_{k,\leq n}$ as follows:

- $B_{k,=n}$ is the number of sets containing k squares, each of them of size $n \times n$
- $B_{k,<n}$ is the the number of sets containing k squares, each of them strictly smaller than $n \times n$
- $B_{k,\leq n} = |\{X \in \mathcal{B}_k : \|X\| = n\}|$ is the number of sets containing k squares, with the biggest one of size $n \times n$.

Similarly, $K_{k,=n}$, $K_{k,<n}$ and $K_{k,\leq n}$ will denote the number of codes containing k squares of respective sizes. All numbers defined above are assumed to be 1 when $k = 0$.

Obviously $B_{1,=n} = 2^{n^2}$ and for arbitrary $n \geq 1$, $k \geq 0$ we have $B_{k,=n} = \binom{B_{1,=n}}{k}$.

We are interested in finding the asymptotic density of k-element codes, i.e., the limit

$$\mu_{\mathcal{B}_k}(\mathcal{K}_k) = \lim_{n\to\infty} \frac{K_{k,\leq n}}{B_{k,\leq n}} = \lim_{n\to\infty} \frac{|\{X \in \mathcal{K}_k : \|X\| = n\}|}{|\{X \in \mathcal{B}_k : \|X\| = n\}|}.$$

The following propositions, stating basic combinatorial properties of the $B_{k,\ldots n}$ numbers, are easily proved.

Proposition 2.1 *For any $k, n \geq 1$*

$$B_{k,<n} = \sum_{i=0}^{n-1} B_{k,\leq i}$$

$$B_{k,\leq n} = \sum_{i=1}^{k} B_{i,=n} \cdot B_{k-i,<n}$$

Proof: Note that $B_{k,\leq n}$ includes sets of k squares that contain at least one $n \times n$ square. □

Proposition 2.2 *For any $k, n \geq 1$*

$$B_{k,<n} = \binom{\sum_{i=0}^{n-1} B_{1,=i}}{k}$$

$$B_{k,\leq n} = \binom{\sum_{i=0}^{n} B_{1,=i}}{k} - \binom{\sum_{i=0}^{n-1} B_{1,=i}}{k}$$

Lemma 2.3

- $\forall k \geq 1\ \exists c = c(k) : B_{k,<n} < \frac{c}{2^n} \cdot 2^{kn^2}$
- $\forall k \geq 1\ \exists c = c(k) : B_{k,\leq n} < \left(\frac{1}{k!} + \frac{c}{2^n}\right) 2^{kn^2}$

Proof: The proof is a simple induction on k. □

Recall that $K_{k,\leq n}$ is the number of codes containing k bricks, with the biggest one of size $n \times n$. Any set containing k squares of fixed size is a code, hence $K_{k,\leq n} > B_{k,=n}$. We can now compute the approximate proportion:

$$
\begin{aligned}
\frac{K_{k,\leq n}}{B_{k,\leq n}} &> \frac{B_{k,=n}}{B_{k,\leq n}} \\
&> \frac{\frac{1}{k!} B_{1,=n}(B_{1,=n} - 1)\ldots(B_{1,=n} - (k-1))}{\frac{1}{k!} 2^{kn^2} + \frac{c(k)}{2^n} 2^{kn^2}} \\
&= \frac{2^{n^2}(2^{n^2} - 1)\ldots(2^{n^2} - (k-1))}{2^{kn^2} + \frac{c(k)\cdot k!}{2^n} 2^{kn^2}} \\
&= \frac{(1 - \frac{1}{2^{n^2}})\ldots(1 - \frac{k-1}{2^{n^2}})}{1 + \frac{c(k)\cdot k!}{2^n}}
\end{aligned}
$$

Since $\mu_{\mathcal{B}_k}(\mathcal{K}_k)$ is bounded by 1, the limit is

$$
\begin{aligned}
\mu_{\mathcal{B}_k}(\mathcal{K}_k) &= \lim_{n \to \infty} \frac{(1 - \frac{1}{2^{n^2}})\ldots(1 - \frac{k-1}{2^{n^2}})}{1 + \frac{c(k)\cdot k!}{2^n}} \\
&= 1
\end{aligned}
$$

We have thus proved:

Theorem 2.4 *For any fixed k, the density of codes among sets containing k squares, $\mu_{\mathcal{B}_k}(\mathcal{K}_k)$, is equal to 1.*

Note that the codes with squares of fixed size are enough to make the density of codes equal to 1. This is not true in the one-dimensional case of words; the density of fixed-length word codes is strictly less than 1, see Remark 3.4.

3 Counting word codes

We now consider ordinary word codes. Although they differ from brick codes in that codicity testing is decidable, their probabilistic behaviour is similar in that the density of codes is equal to 1.

By $\mathcal{S}_k \subset \mathcal{P}(A^*)$ we denote the family of all sets containing k words; $\mathcal{C}_k \subseteq \mathcal{S}_k$ is the family of codes containing k words. The size $\|X\|$ of a set $X \in \mathcal{S}_k$ is the length of the longest word in X.

Similarly to the brick case, $S_{k,=n}$, $S_{k,<n}$ and $S_{k,\leq n}$ denote the number of sets containing k words with respective lengths (all of length n, all shorter than n, the longest one of length n) and $C_{k,=n}$, $C_{k,<n}$ and $C_{k,\leq n}$ denote the number of codes containing k words with respective lengths.

Once again, we are interested in finding the asymptotic density of k-element codes, i.e., the limit

$$\mu_{\mathcal{S}_k}(\mathcal{C}_k) = \lim_{n\to\infty} \frac{C_{k,\leq n}}{S_{k,\leq n}} = \lim_{n\to\infty} \frac{|\{X \in \mathcal{C}_k : \|X\| = n\}|}{|\{X \in \mathcal{S}_k : \|X\| = n\}|}.$$

Clearly $S_{1,=n} = 2^n$ and basic combinatorial properties follow those of the $B_{k,\ldots n}$ numbers.

Proposition 3.1 *For any* $k, n \geq 1$

$$\begin{aligned} S_{k,<n} &= \sum_{i=0}^{n-1} S_{k,\leq i} \\ S_{k,\leq n} &= \sum_{i=1}^{k} S_{i,=n} \cdot S_{k-i,<n} \end{aligned}$$

Proposition 3.2 *For any* $k, n \geq 1$

$$\begin{aligned} S_{k,<n} &= \binom{\sum_{i=0}^{n-1} S_{1,=i}}{k} \\ S_{k,\leq n} &= \binom{\sum_{i=0}^{n} S_{1,=i}}{k} - \binom{\sum_{i=0}^{n-1} S_{1,=i}}{k} \end{aligned}$$

Lemma 3.3

$$\lim_{n\to\infty} \frac{S_{k,\leq n}}{2^{kn}} = \frac{2^k - 1}{k!}$$

Proof: By Proposition 3.2

$$\begin{aligned}
S_{k,\leq n} &= \binom{\sum_{i=0}^{n} S_{1,=i}}{k} - \binom{\sum_{i=0}^{n-1} S_{1,=i}}{k} \\
&= \binom{2^{n+1}-1}{k} - \binom{2^n-1}{k} \\
&= \frac{(2^{n+1}-k)...(2^{n+1}-1)}{k!} - \frac{(2^n-k)...(2^n-1)}{k!}
\end{aligned}$$

Hence

$$\begin{aligned}
\lim_{n\to\infty} \frac{S_{k,\leq n}}{2^{nk}} &= \frac{1}{k!}\lim_{n\to\infty}[(2-\frac{k}{2^n})...(2-\frac{1}{2^n})-(1-\frac{k}{2^n})...(1-\frac{1}{2^n})] \\
&= \frac{2^k-1}{k!}
\end{aligned}$$

□

Remark 3.4 *As noted earlier, the density of fixed-length word codes is strictly less than 1, hence different estimates will have to be used. Note that for* $k, n \geq 1$

$$\begin{aligned}
\frac{C_{k,=n}}{S_{k,\leq n}} &= \frac{\binom{2^n}{k}}{\binom{2^{n+1}-1}{k} - \binom{2^n-1}{k}} \\
&= \frac{(1-\frac{k-1}{2^n})...(1-\frac{1}{2^n})}{(2-\frac{k}{2^n})...(2-\frac{1}{2^n})-(1-\frac{k}{2^n})...(1-\frac{1}{2^n})} \\
&\xrightarrow[n\to\infty]{} \frac{1}{2^k-1} \\
&\xrightarrow[k\to\infty]{} 0
\end{aligned}$$

We now estimate $C_{k,\leq n}$, the number of codes containing k words, with the longest one of length n.

Lemma 3.5

$$S_{k,\leq n} - C_{k,\leq n} < kn \cdot S_{k-1,\leq n}$$

Proof: The number on the left-hand side is the number of all non-codes (with given cardinality and size). This is less than e.g. the number of all non-prefix sets. □

Lemma 3.6

$$\lim_{n\to\infty} \frac{S_{k,\le n} - C_{k,\le n}}{2^{kn}} = 0$$

Proof: By Lemma 3.3 and 3.5 we have

$$\begin{aligned}
\lim_{n\to\infty} \frac{S_{k,\le n} - C_{k,\le n}}{2^{kn}} &\le \lim_{n\to\infty} \frac{kn \cdot S_{k-1,\le n}}{2^{kn}} \\
&= k \cdot \lim_{n\to\infty} \frac{S_{k-1,\le n}}{2^{(k-1)n}} \lim_{n\to\infty} \frac{n}{2^n} \\
&= k \frac{2^{k-1}-1}{(k-1)!} \lim_{n\to\infty} \frac{n}{2^n} \\
&= 0
\end{aligned}$$

□

The approximate proportion of codes among words can now be computed as

$$\begin{aligned}
\frac{C_{k,\le n}}{S_{k,\le n}} &= \frac{S_{k,\le n} - (S_{k,\le n} - C_{k,\le n})}{S_{k,\le n}} \\
&= 1 - \frac{S_{k,\le n} - C_{k,\le n}}{2^{kn}} \cdot \frac{2^{kn}}{S_{k,\le n}}
\end{aligned}$$

Thus, by Lemma 3.3 and 3.6

$$\begin{aligned}
\lim_{n\to\infty} \frac{C_{k,\le n}}{S_{k,\le n}} &= 1 - 0 \cdot \frac{k!}{2^k - 1} \\
&= 1
\end{aligned}$$

We have now proved:

Theorem 3.7 *For any fixed k, the density of codes among sets containing k words, $\mu_{S_k}(C_k)$, is equal to 1.*

Note that the estimate used to prove Theorem 3.7 is based on prefix codes alone. Consequently, the density of prefix codes among all word codes is equal to 1.

Corollary 3.8

$$\lim_{k\to\infty} \mu_{\mathcal{B}_k}(\mathcal{K}_k) = \lim_{k\to\infty} \mu_{S_k}(C_k) = 1$$

4 Conclusions

We have proved that for both square bricks and words the asymptotic densities are equal to 1. Thus, what we know about the decidability and density of brick codes with square bricks amounts to the following:

- If the effective alphabet is non-trivial, we have decidability for two-element sets and undecidability for sets with at least 15 elements.
- The proportion of codes among all sets of square bricks is asymptotically equal to 1.

The obvious direction for further research is to generalize these results to wider classes of brick codes.

References

[1] P. Aigrain, D. Beauquier: Polyomino tilings, cellular automata and codicity. *Theoret. Comp. Sci.* 147 (1995) 165–180.

[2] D. Beauquier, M. Nivat: A codicity undecidable problem in the plane. *Theoret. Comp. Sci.* 303 (2003) 417–430.

[3] J. Berstel, D. Perrin: Theory of Codes. Academic Press (1985).

[4] J. Karhumäki: Some open problems in combinatorics of words and related areas. *TUCS Technical Report* 359 (2000).

[5] M. Margenstern: Frontiers between decidability and undecidability: a survey. *Theoret. Comp. Sci.* 231 (2000) 217–251.

[6] Yu. Matiyasevich: Simple examples of undecidable associative calculi. *Soviet Math. Dokladi* 8 (1967) 555–557.

[7] Yu. Matiyasevich: Word problem for Thue systems with a few relations. LNCS 909 (1995) 39–53.

[8] Yu. Matiyasevich, G. Sénizergues: Decision problems for semi-Thue systems with a few rules. Proceedings LICS'96 (1996) 523–531.

[9] M. Moczurad, W. Moczurad: Decidability of simple brick codes. In: Mathematics and Computer Science III (Algorithms, Trees, Combinatorics and Probabilities), Trends in Mathematics, Birkhäuser (2004).

[10] M. Moczurad, W. Moczurad: Some open problems in decidability of brick (labelled polyomino) codes. Cocoon 2004 Proceedings, LNCS 3106 (2004) 72–81.

[11] M. Moczurad, J. Tyszkiewicz, M. Zaionc: Statistical properties of simple types. *Math. Struct. in Comp. Science* 10 (2000) 575–594.

[12] W. Moczurad: Algebraic and algorithmic properties of brick codes. Doctoral thesis, Jagiellonian University (2000).

[13] W. Moczurad: Brick codes: families, properties, relations. *Intern. J. Comp. Math.* 74 (2000) 133–150.

[14] W. Rytter: The Space Complexity of the Unique Decipherability Problem. *Inf. Proc. Letters* 23 (1986) 1-3.

[15] E. Szczypka: Object complexity. Doctoral thesis, Jagiellonian University (2004).

[16] H. Wilf: generatingfunctionology. Academic Press (1994).

[17] K. Yeats: Asymptotic Density in Combined Number Systems. *New York J. Math.* 8 (2002) 63–83.