Do balanced words have a short period?

Nadia Brauner[a], Yves Crama[b], Etienne Delaporte[a], Vincent Jost[a,*], Luc Libralesso[a]

[a] *Univ. Grenoble Alpes, CNRS, Grenoble INP, G-SCOP, F-38000 Grenoble, France*
[b] *QuantOM, HEC Management School, University of Liège, Belgium.*

Abstract

We conjecture that each balanced word on N letters

- either arises from a balanced word on two letters by expanding both letters with a congruence word,
- or is D-periodic with $D \leq 2^N - 1$.

Our conjecture arises from extensive numerical experiments. It implies, for any fixed N, the finiteness of the number of balanced words on N letters which do not arise from expanding a balanced word on two letters. It refines a theorem of Graham and Hubert, which states that non-periodic balanced words are congruence expansions of balanced words on two letters. It also relates to Fraenkel's conjecture, which states that for $N \geq 3$, every balanced word with distinct densities $d_1 > d_2 \ldots > d_N$ satisfies $d_i = (2^{N-i})/(2^N - 1)$, since this implies that the word is D-periodic with $D = 2^N - 1$. For $N \leq 6$, we provide a tentative list of the density vectors of balanced words which do not arise from expanding a balanced word with fewer letters. We prove that the list is complete for $N = 4$ letters.

We also prove that deleting a letter in a congruence word always produces a balanced word and this constructions allows to further reduce the list of density

*Corresponding author
Email addresses: `Nadia.Brauner@grenoble-inp.fr` (Nadia Brauner), `yves.crama@uliege.be` (Yves Crama), `etienne.delaporte@outlook.com` (Etienne Delaporte), `Vincent.Jost@grenoble-inp.fr` (Vincent Jost), `Luc.Libralesso@grenoble-inp.fr` (Luc Libralesso)

December 12, 2018

vectors that remains unexplained. Moreover, we prove that deleting a letter in a m-balanced word produces a $m+1$-balanced word, extending and simplifying a result of [11].

Keywords: Balanced words, congruence words, exact covering systems, constant gap sequences, Graham- Hubert theorem, Fraenkel's conjecture, m-balanced words.

---

1. Introduction

Balanced sequences and balanced words have attracted the attention of many researchers in discrete mathematics or number theory, but also in application fields like scheduling, maintenance, queueing, or apportionment (see, e.g., [1, 6, 3, 8, 9]). Yet, in spite of this wealth of literature, balanced words and their properties are still poorly understood. A most striking, and best known illustration of this assertion is provided by a conjecture initially formulated by Fraenkel (as mentioned in [14]) for exact covers by Beatty sequences. Fraenkel's conjecture was later extended to balanced words (Altman, Gaujal and Hordijk [1]). Although Fraenkel's conjecture has been established for words on a small number of letters (up to $N = 7$ letters [2]), its general case remains stubbornly open.

Our objective in this research was to get a better grasp of the vectors of densities associated with balanced words. For this purpose, we have conducted computational experiments in which we have generated all such vectors up to a certain size. Out of the numerical results came a conjecture that we believe to be new and of potential interest to the mathematical community.

Graham [5] made important observations about how balanced words arise out of balanced words on 2 letters and congruence words. Our study was guided by the following question "Which balanced words remain after cleaning those explained by the construction of Graham?"

This note is organized as follows: in Section 2, we review the literature required for our study. In Section 3 we explain why Fraenkel's conjecture is not sufficient, to the best of our understanding, to reduce the study of balanced words to the study of congruence words. In Section 4 we claim our main contributions: the conjecture about the short period of balanced words, as well as a simple construction for generating some balanced words by deleting a letter in a congruence word.

In Section 5 we investigate balanced words with $N \leq 6$ letters, providing a complete list of these words under the validity of our main conjecture. We also exhibit that our simple construction explains many of the balanced words that do not arise from Graham's construction.

The proof of our conjecture for $N = 4$ letters is provided in the Appendix.

2. Background

We briefly introduce the definitions that are needed in the sequel. The reader is referred to [14, 15, 11, 16] for details and additional information.

Definition 2.1. *A* sequence *is a subset of* $\mathbb{Z}$. *A sequence* S *is* D-periodic*, where* D *is a positive integer, if* $S = \{x + D : x \in S\}$. *The* period *of a sequence is the smallest* D *for which it is* D*-periodic. The* density $\delta(S)$ *of a sequence* S *is*

$$\delta(S) := \lim_{t\to\infty} \{|I \cap S|/|I| : I = \{a, ..., t\},\, a \in \mathbb{Z}\},$$

provided the limit exists.

Definition 2.2. *A* word on N letters *is a function* $W : \mathbb{Z} \to \{1, ..., N\}$*, or equivalently, a (left- and right-unbounded) string of symbols on the alphabet* $\{1, ..., N\}$. *The word* W *is* D-periodic *if* $W(k + D) = W(k)$ *for all* $k \in \mathbb{Z}$*, and the* period *of a word is the smallest positive integer* D *(if any) for which it is* D*-periodic.*

Each word is naturally associated with a partition of $\mathbb{Z}$ into a finite family of sequences $\{S_i\}_{i\in\{1,...,N\}}$, where S_i is the set of integers that W maps to i. The word is D-periodic if and only if all sequences S_i are D-periodic. Provided that all densities $\delta(S_i)$ exist, the *density (vector)* of W is $\delta(W) = (\delta(S_1), \ldots, \delta(S_N))$.

A D-periodic word is completely defined by its restriction to the interval $\{1, ..., D\}$. Therefore, when presenting examples, we usually describe a periodic word by a finite string X of integers or letters, such as $X = 1213112231$ or $X = abacaabbca$. The word itself is $W = (X)^*$, where the star operator indicates infinite repetition of the string X.

Notice that with this setting, we identify $(01)^*$ and $(10)^*$. In other words, this defines a word up to a common shift of all its sequence, which is sufficient and more convinient, for the issues addressed in this paper.

Definition 2.3. *A* factor *in a word W is a finite sequence of consecutive letters of W. Equivalently, a* factor *is the image by W of a finite interval of integers.*

Definition 2.4. *A* balanced sequence *is a sequence S such that, for every pair I_1 and I_2 of intervals of integers of the same length, the difference between the number of elements of the sequence in the two intervals is at most 1: that is, if $I_1 = \{i_1, \ldots, i_1 + t\}$ and $I_2 = \{i_2, \ldots, i_2 + t\}$, then*

$$-1 \leq |I_1 \cap S| - |I_2 \cap S| \leq 1.$$

A word is balanced *if all its associated sequences are balanced.*

Balanced sequences and words have been extensively studied [1, 14, 15]. A structural theorem about balanced sequences from [10] implies that every balanced sequence S has a density. Moreover if the density $\delta(S)$ is irrational, then the balanced sequence S is not periodic.

We now introduce an important class of balanced words which have been named in several ways: congruence words, exact covering systems, constant gap words, exact covering congruences [5].

Definition 2.5. *A* congruence word *is a word $\{S_i\}_{i \in \{1,\ldots,N\}}$ such that all sequences S_i are* congruence sequences*, that is, sequences of the form $S_i = \{a_i n + b_i : n \in \mathbb{Z}\}$, where a_i, b_i are arbitrary integers, $a_i \neq 0$.*

Congruence sequences and words can be characterized in a way that shows that they are balanced:

Proposition 2.6. *[1] A sequence $S \subseteq \mathbb{Z}$ is a congruence sequence if and only if for every pair of intervals I_1 and I_2 of almost equal length (i.e. $||I_1| - |I_2|| \leq 1$), the balance condition holds (i.e. $||I_1 \cap S| - |I_2 \cap S|| \leq 1$).*

Of course this proposition can be used to characterize congruence words, by requiring the above condition for each of its letters.

Graham [5] observed that congruence words can be used to build balanced words from balanced words, as follows.

Definition 2.7. *Let W be a word on letters $\{1, ..., N\}$, let A be a word on letters $\{N+1, ..., M\}$, and let $j \in \{1, ..., N\}$. Consider the word $W_{A,j}$ on $M-1$ letters obtained by replacing in W the k-th occurrence of letter j by the k-th letter of the word A, for all $k \in \mathbb{Z}$ (we set the convention that the 0-th occurrence of letter j in W is the one with smallest non-negative position). If A is a congruence word, the word $W_{A,j}$ is a* congruence substitution *of the word W.*

Definition 2.8. *V is a* congruence expansion *of W if there is a finite sequence of words $W = W_1, \ldots, W_k = V$ such that for all i in $\{1, \ldots k-1\}$, the word W_{i+1} is a congruence substitution of W_i.*

Proposition 2.9. *[5] Any congruence expansion of a balanced word is also a balanced word.*

Example 2.10. *The word $W = (abacaba)^*$ is balanced and $A = (de)^*$ is a congruence word. Then, $W_{A,c} = (abadabaabaeaba)^*$ is a balanced word obtained by substituting occurrences of the letter c by d and by e, alternatively.*

Extending a theorem by Graham [5] for irrational densities, Hubert [7] established an important property of non-periodic balanced words (see also Altman et al. [1]): Proposition 2.9 provides a construction for *all* non-periodic balanced words.

Theorem 2.11. *[7] If W is a non-periodic balanced word, then W is a congruence expansion of a balanced word on two letters.*

However, not all balanced words are congruence expansions of balanced words on 2 letters. The most famous among such words are the following [1, 14]:

Definition 2.12. *The* Fraenkel word *on N letters is the periodic balanced word F^N recursively defined by $F_1 = 1$, $F_N = F_{N-1}NF_{N-1}$ if $N \geq 2$ and $F^N = (F_N)^*$ for all N.*

For example $F_3 = 1213121$. The density vector of F^N is ϕ^N, with $\phi_i^N = \frac{2^{N-i}}{2^N-1}$ for $i = 1, \ldots, N$.

Conjecture 2.13 (Fraenkel's conjecture). *For all $N \geq 3$, if W is a balanced word on N letters such that all components of its density vector are pairwise distinct, then its density vector is ϕ^N.*

Conjecture 2.13 has been proved for $N \leq 7$ (see Altman et al. [1], Barát and Varjú [2], Tijdeman [13]), but it remains open for larger values of N.

We end our literature review with the concept of projection of a word on a subset of its letters, since we will show later that removing a letter in a congruence word generates a balanced word.

Definition 2.14. *Given a word $W = \{S_i\}_{i\in\{1,\ldots,N\}}$ on N letters and $X \subseteq \{1, ..., N\}$, the projection W_{-X} of W on $\{1, ..., N\} \setminus X$ is defined by reading W, skipping letters in X.*

To refer to a position in the projected word (as it is undefined with the above definition), we will refer to the associated position in the original word.

It is known for instance that if W is balanced and letter a has density at least $1/2$, then W_{-a} is still a balanced word [1]. Moreover, if W is balanced and letter a has density at least $2/3$, then W_{-a} is a congruence word [12].

The following generalization of balancedness was proposed and studied by [11].

Definition 2.15. *For a sequence $S \subseteq \mathbb{Z}$, an interval $X = \{a, ..., b\}$ is a S-*chain *if $a-1 \in S$ and $b+1 \in S$. For a word W and a letter s, a factor X is a s-*chain *if X is directly preceeded and directly followed by an s.*

Definition 2.16. *For a non-negative integer m, a sequence $S \subseteq \mathbb{Z}$ is m-*balanced*, if for every S-chain X and every interval X' such that $|X'| = |X| + m + 1$, we have $|X' \cap S| \geq |X \cap S| + 1$.*

0-balanced sequences are exactly congruence sequences (or contain one element). 1-balanced sequences are exactly balanced sequences.

3. Fraenkel conjecture is not sufficient to reduce balanced words to congruence words

We should note that the converse of Proposition 2.9 is not valid in general: if a congruence expansion of a word W is balanced, it does not mean that W itself is balanced, as in the following example.

Example 3.1. *The word $W = (dcdedcdedcd)^*$ is not balanced. If we use the congruence word $(ab)^*$ to expand W on the letter e, we obtain the balanced word $W_{A,e} = (dcdadcdbdcd)^*$ on four letters.*

Remark 3.2. *Tijdeman [15] asks what are the balanced words on more than two letters. He goes on to observe that, for such words:*

"Obviously each letter has again a density. If the densities of two letters are equal, then they can first be identified as one letter with double density, and then the latter letter can be replaced alternately by the first and second letter. It is therefore a crucial question to determine the balanced words the letters of which have distinct densities, so-called Fraenkel words."

This comment seems to suggest that, by identifying letters of equal density in a balanced word, one obtains again a balanced word. However, this is in contradiction with Example 3.1. Indeed, for this example, the density of $W_{A,e}$ is $\left(\frac{6}{11}, \frac{3}{11}, \frac{1}{11}, \frac{1}{11}\right)$, with the last two densities (of a and b) being equal. By identifying the letters a and b, however, one obtains the density (of W) $\left(\frac{6}{11}, \frac{3}{11}, \frac{2}{11}\right)$, and there is no balanced word with these densities.

As we will see in Section 4, Example 3.1 does indeed arise from a congruence word by deleting one letter (of density $1/12$). In Section 5 however, we exhibit balanceable density vectors for which no construction seems to be known.

4. Main new statements

Building upon Theorem 2.11, we concentrate in the sequel on the case of periodic words and their (necessarily rational) densities. We tried and list all balanced words on $N = 4, 5$ or 6 letters using a computer. To this aim, we had to restrict our attention to words with a relatively small period D. Our experiments led us to the following conjecture:

Conjecture 4.1. *If a word W on N letters is balanced, then either*

(1) *W is a congruence expansion of a balanced word on two letters, or*

(2) *W is D-periodic for some $D \leq 2^N - 1$.*

Note that the set of words that satisfy condition (2) is finite for each fixed N. In the next section, we refine Conjecture 4.1 for $N \leq 6$, by listing all balanceable density vectors that do not come from congruence expansions of other balanced words. A careful study of these vectors led us to observe that several such density vectors look very much like density vectors of congruence words. Indeed, they arise just by deleting a letter of lowest density in a congruence word. These examples motivate the following observation:

Theorem 4.2. *If $\{S_a\}_{a\in\{1,\dots,N\}}$ is a congruence word, then the projected word obtained by deleting any of its letter $a \in \{1, ..., N\}$ is balanced.*

To prove this Theorem, we first claim it in the most general form allowed by our proof.

Lemma 4.3. *Let $m \in \mathbb{Z}_+$ and S, T be two disjoint sequences, such that S is m-balanced and T is $(m + 3)$-balanced. Let R be the sequence $\mathbb{Z} \setminus (S \cup T)$, and*

W be the word formed with the three sequences R, S, T. Then, in the word W_{-T} (that is W projected on R and S), the sequence S is $(m+1)$-balanced.

A weaker version of Lemma 4.3 already appears in [11] (as Theorem 4.2: “If we remove a sequence in a m-balanced word W and if W is also billiard, then we obtain a $m+1$-balanced word”; most interestingly, Lemma 4.3 claims that the billiard assumption can be relaxed).

Let us deduce Theorem 4.2 from Lemma 4.3. In a congruence word, all sequences are disjoint and 0-balanced. Hence, removing any sequence in a congruence word leaves the other sequences 1-balanced.

Proof. of Lemma 4.3. Let $\bar{S}$ be the sequence of integers corresponding to S in W_{-T}. Assume that $\bar{S}$ is not $(m+1)$-balanced in W_{-T}. Denote by s and t the letters corresponding to sequences S and T.

Since $\bar{S}$ is not $(m+1)$-balanced, there exists a s-chain $\bar{X}$ in W_{-T} and a factor $\bar{X}'$ such that $|\bar{X}'| = |\bar{X}| + m + 2$ and $|\bar{X}' \cap \bar{S}| \leq |\bar{X} \cap \bar{S}|$.

There exists a s-chain $\bar{X}'' \supseteq \bar{X}'$ of W_{-T} such that $|\bar{X}''| \geq |\bar{X}| + m + 2$ and $|\bar{X}'' \cap \bar{S}| = |\bar{X} \cap \bar{S}|$.

$\bar{X}$ (resp. $\bar{X}''$) is the projection of a s-chain X (resp. X'') of W. Both are uniquely defined by the fact that they are s-chain in W. We have $|\bar{X}''| \geq |\bar{X}| + m + 2$ and $|\bar{X}'' \cap \bar{S}| = |\bar{X} \cap \bar{S}|$ and $|\bar{X}| = |\bar{X} \cap \bar{S}| + |\bar{X} \cap \bar{R}|$ and $|\bar{X}''| = |\bar{X}'' \cap \bar{S}| + |\bar{X}'' \cap \bar{R}|$. Hence $|X'' \cap R| - |X \cap R| \geq m + 2$.

Also we know that X is a s-chain and X'' verifies $|X'' \cap S| = |X \cap S|$. Since S is m-balanced, it implies that $|X''| \leq |X| + m$ (since otherwise, one would have $|X'' \cap S| \geq |X \cap S| + 1$). We have $|X| = |X \cap S| + |X \cap R| + |X \cap T|$ and $|X''| = |X'' \cap S| + |X'' \cap R| + |X'' \cap T|$ and $|X \cap S| = |X'' \cap S|$ and $|X'' \cap R| - |X \cap R| \geq m+2$. Then, $|X''| \leq |X| + m$ implies that $|X'' \cap T| + m + 2 \leq |X \cap T| + m$ and hence $|X \cap T| - |X'' \cap T| \geq 2$.

So there exists a t-chain $Y \subseteq X$ such that $|Y \cap T| = |X'' \cap T|$.

We also have that $|Y \cap S| \leq |X'' \cap S|$ and $|Y \cap R| + (m+2) \leq |X'' \cap R|$.

Finally, adding up the counts of the three sequences to compare their length, the t-chain Y and the factor $sX''s$ serve as a certificate for non $(m+3)$-balancedness of the sequence T: Y is a t-chain and $|Y| + m + 4 \leq |sX''s|$ and $|sX''s \cap T| = |Y \cap T|$. □

5. Small values of N

For simplicity, we call *density vector* (or *density*, for short) any vector $\delta \in \mathbb{Q}^N$ such that $\sum_{i=1}^{N} \delta_i = 1$, and we assume that density vectors are defined up to a permutation of their components.

Definition 5.1. *A density vector $\delta = (\delta_1, \dots, \delta_N) \in \mathbb{Q}^N$ is* balanceable *if there exists a balanced word W on N letters such that $\delta(W) = \delta$. For a balanceable density δ, the* period *of δ, denoted $D(\delta)$, is the smallest period of a balanced word with density δ.*

For a density vector with pairwise distinct components, Conjecture 2.13 implies Conjecture 4.1. If the components take at most two distinct values, the density vector is of the form $(\alpha/k_1, ..., \alpha/k_1, (1-\alpha)/k_2, ..., (1-\alpha)/k_2)$ for some $\alpha \in (0, 1)$ and two integers k_1, k_2. Hence the vector is balanceable and validates case (1) in Conjecture 4.1.

5.1. The case $N \leq 3$

The following results are well-known and can be found for instance in Altman et al. [1] or Tijdeman [13].

Proposition 5.2. *For $N = 2$ letters, the balanceable density vectors are exactly the vectors of the form: $(\alpha, 1-\alpha)$, for all $0 < \alpha < 1$.*

Proposition 5.3. *For $N = 3$ letters, the balanceable density vectors are exactly the vectors of the form: $(\alpha/2, \alpha/2, 1-\alpha)$, for all $0 < \alpha < 1$. The only balanceable vector not in this infinite list is $\phi^3 = (4/7, 2/7, 1/7)$.*

These results immediately imply that Conjecture 4.1 holds when $N \leq 3$.

5.2. The case $N = 4$

Altman et al. [1] establish several results on balanceable vectors of four densities, including a proof of Fraenkel's conjecture. The Appendix of [1] also provides a list of balanceable vectors for $N = 4$. This list is actually complete:

Theorem 5.4. *For $N = 4$ letters, the balanceable density vectors are exactly the vectors in the following classes:*
1) All vectors of the form

$$(\frac{\alpha}{3}, \frac{\alpha}{3}, \frac{\alpha}{3}, 1-\alpha), (\frac{\alpha}{2}, \frac{\alpha}{4}, \frac{\alpha}{4}, 1-\alpha), (\frac{\alpha}{2}, \frac{\alpha}{2}, \frac{1-\alpha}{2}, \frac{1-\alpha}{2}),$$

for all $0 < \alpha < 1$.
2) Five balanceable vectors which are not in the previous infinite classes, namely:
$\left(\frac{6}{11}, \frac{3}{11}, \frac{1}{11}, \frac{1}{11}\right)$ $\left(\frac{6}{11}, \frac{2}{11}, \frac{2}{11}, \frac{1}{11}\right)$ $\left(\frac{4}{11}, \frac{4}{11}, \frac{2}{11}, \frac{1}{11}\right)$ $\left(\frac{8}{14}, \frac{4}{14}, \frac{1}{14}, \frac{1}{14}\right)$ $\left(\frac{8}{15}, \frac{4}{15}, \frac{2}{15}, \frac{1}{15}\right)$.

The proof of Theorem 5.4 is provided in Appendix A. The first three infinite classes in Theorem 5.4 satisfy condition (1) of Conjecture 4.1: they can be obtained from a two-letter word on $\{a, b\}$ with density $(\alpha, 1-\alpha)$, by expanding either the first letter a using one of the congruence words cde or $cdce$, or each of the two letters a and b using the congruence words cd and ef, respectively. Note also that $\left(\frac{8}{14}, \frac{4}{14}, \frac{1}{14}, \frac{1}{14}\right)$ arises from a congruence expansion of the Fraenkel word with density $\left(\frac{4}{7}, \frac{2}{7}, \frac{1}{7}\right)$. Hence, we have the following statement:

Corollary 5.5. *For $N = 4$ letters, the balanceable density vectors which are not density vectors of a congruence expansion of some balanced word on $N \leq 3$ letters, are:*
$\left(\frac{6}{11}, \frac{3}{11}, \frac{1}{11}, \frac{1}{11}\right)$ $\left(\frac{6}{11}, \frac{2}{11}, \frac{2}{11}, \frac{1}{11}\right)$ $\left(\frac{4}{11}, \frac{4}{11}, \frac{2}{11}, \frac{1}{11}\right)$ $\left(\frac{8}{15}, \frac{4}{15}, \frac{2}{15}, \frac{1}{15}\right)$

All of them can be constructed from Theorem 4.2 (increasing both the number of letters and the period by 1 yields the density vector of a congruence word, e.g. $(6/12, 3/12, 1/12, 1/12, 1/12)$ in the first case).

5.3. The case $N = 5$

For $N = 5$, Fraenkel's conjecture has been proved by Tijdeman [13]. We have verified by computer the following conjecture for all vectors with a period $D \leq 130$.

Conjecture 5.6. *For $N = 5$, the balanceable vectors which are not density vectors of a congruence expansion of some balanced word on $N \leq 4$ letters are:*

$\boxed{\left(\frac{8}{13}, \frac{2}{13}, \frac{1}{13}, \frac{1}{13}, \frac{1}{13}\right)}$	$\left(\frac{6}{17}, \frac{6}{17}, \frac{2}{17}, \frac{2}{17}, \frac{1}{17}\right)$	$\left(\frac{12}{23}, \frac{6}{23}, \frac{3}{23}, \frac{1}{23}, \frac{1}{23}\right)$
$\boxed{\left(\frac{6}{13}, \frac{3}{13}, \frac{2}{13}, \frac{1}{13}, \frac{1}{13}\right)}$	$\left(\frac{9}{17}, \frac{3}{17}, \frac{3}{17}, \frac{1}{17}, \frac{1}{17}\right)$	$\left(\frac{12}{23}, \frac{6}{23}, \frac{2}{23}, \frac{2}{23}, \frac{1}{23}\right)$
$\boxed{\left(\frac{4}{13}, \frac{3}{13}, \frac{2}{13}, \frac{2}{13}, \frac{2}{13}\right)}$	$\left(\frac{6}{17}, \frac{6}{17}, \frac{3}{17}, \frac{1}{17}, \frac{1}{17}\right)$	$\left(\frac{12}{23}, \frac{4}{23}, \frac{4}{23}, \frac{2}{23}, \frac{1}{23}\right)$
	$\left(\frac{8}{23}, \frac{8}{23}, \frac{4}{23}, \frac{2}{23}, \frac{1}{23}\right)$	$\left(\frac{16}{31}, \frac{8}{31}, \frac{4}{31}, \frac{2}{31}, \frac{1}{31}\right)$

Those we cannot construct from Theorem 4.2 are boxed.

5.4. The case $N = 6$

The case $N = 6$ is similar to the previous ones. The following conjecture has been tested by computer for all density vectors with a period $D \leq 80$.

Conjecture 5.7. *For $N = 6$, the balanceable vectors which are not density vectors of a congruence expansion of some balanced word on $N \leq 5$ letters are:*

$\boxed{\left(\frac{5}{13},\frac{3}{13},\frac{2}{13},\frac{1}{13},\frac{1}{13},\frac{1}{13}\right)}$	$\boxed{\left(\frac{6}{19},\frac{4}{19},\frac{4}{19},\frac{2}{19},\frac{2}{19},\frac{1}{19}\right)}$	$\left(\frac{12}{35},\frac{12}{35},\frac{6}{35},\frac{2}{35},\frac{2}{35},\frac{1}{35}\right)$
$\boxed{\left(\frac{9}{16},\frac{3}{16},\frac{1}{16},\frac{1}{16},\frac{1}{16},\frac{1}{16}\right)}$	$\left(\frac{10}{19},\frac{5}{19},\frac{1}{19},\frac{1}{19},\frac{1}{19},\frac{1}{19}\right)$	$\left(\frac{12}{35},\frac{12}{35},\frac{4}{35},\frac{4}{35},\frac{2}{35},\frac{1}{35}\right)$
$\boxed{\left(\frac{10}{17},\frac{2}{17},\frac{2}{17},\frac{1}{17},\frac{1}{17},\frac{1}{17}\right)}$	$\left(\frac{10}{19},\frac{2}{19},\frac{2}{19},\frac{2}{19},\frac{2}{19},\frac{1}{19}\right)$	$\left(\frac{24}{47},\frac{12}{47},\frac{6}{47},\frac{3}{47},\frac{1}{47},\frac{1}{47}\right)$
$\boxed{\left(\frac{9}{17},\frac{3}{17},\frac{2}{17},\frac{1}{17},\frac{1}{17},\frac{1}{17}\right)}$	$\left(\frac{4}{19},\frac{4}{19},\frac{4}{19},\frac{4}{19},\frac{2}{19},\frac{1}{19}\right)$	$\left(\frac{24}{47},\frac{12}{47},\frac{6}{47},\frac{2}{47},\frac{2}{47},\frac{1}{47}\right)$
$\boxed{\left(\frac{8}{17},\frac{3}{17},\frac{2}{17},\frac{2}{17},\frac{1}{17},\frac{1}{17}\right)}$	$\boxed{\left(\frac{8}{21},\frac{8}{21},\frac{2}{21},\frac{1}{21},\frac{1}{21},\frac{1}{21}\right)}$	$\left(\frac{24}{47},\frac{12}{47},\frac{4}{47},\frac{4}{47},\frac{2}{47},\frac{1}{47}\right)$
$\boxed{\left(\frac{6}{17},\frac{4}{17},\frac{3}{17},\frac{2}{17},\frac{1}{17},\frac{1}{17}\right)}$	$\boxed{\left(\frac{12}{25},\frac{6}{25},\frac{3}{25},\frac{2}{25},\frac{1}{25},\frac{1}{25}\right)}$	$\left(\frac{24}{47},\frac{8}{47},\frac{8}{47},\frac{4}{47},\frac{2}{47},\frac{1}{47}\right)$
$\boxed{\left(\frac{6}{17},\frac{4}{17},\frac{2}{17},\frac{2}{17},\frac{2}{17},\frac{1}{17}\right)}$	$\left(\frac{9}{26},\frac{9}{26},\frac{3}{26},\frac{3}{26},\frac{1}{26},\frac{1}{26}\right)$	$\left(\frac{16}{47},\frac{16}{47},\frac{8}{47},\frac{4}{47},\frac{2}{47},\frac{1}{47}\right)$
$\boxed{\left(\frac{4}{17},\frac{4}{17},\frac{3}{17},\frac{2}{17},\frac{2}{17},\frac{2}{17}\right)}$	$\left(\frac{12}{35},\frac{12}{35},\frac{6}{35},\frac{3}{35},\frac{1}{35},\frac{1}{35}\right)$	$\left(\frac{32}{63},\frac{16}{63},\frac{8}{63},\frac{4}{63},\frac{2}{63},\frac{1}{63}\right)$
$\boxed{\left(\frac{10}{19},\frac{3}{19},\frac{2}{19},\frac{2}{19},\frac{1}{19},\frac{1}{19}\right)}$	$\left(\frac{18}{35},\frac{9}{35},\frac{3}{35},\frac{3}{35},\frac{1}{35},\frac{1}{35}\right)$	
$\boxed{\left(\frac{6}{19},\frac{6}{19},\frac{3}{19},\frac{2}{19},\frac{1}{19},\frac{1}{19}\right)}$	$\left(\frac{18}{35},\frac{6}{35},\frac{6}{35},\frac{3}{35},\frac{1}{35},\frac{1}{35}\right)$	
$\boxed{\left(\frac{9}{19},\frac{3}{19},\frac{3}{19},\frac{2}{19},\frac{1}{19},\frac{1}{19}\right)}$	$\left(\frac{18}{35},\frac{6}{35},\frac{6}{35},\frac{2}{35},\frac{2}{35},\frac{1}{35}\right)$	

Those we cannot construct from Theorem 4.2 are boxed.

6. Conclusion

Most balanced words seem to arise from balanced words on two letters by congruence expansion. Those which don't come from a word on two letters seem to have a period D satisfying $D \leq 2^N - 1$. Among them, several come from deleting a letter in a congruence word. However there remain balanced words on 5 and 6 letters for which no particular structure seems to be known.

In the last 2 decades, attention on balanced words focused on Fraenkel's conjecture, providing proofs for up to 7 letters. We argue however, that even proving this conjecture might not be sufficient to understand balanced words, and that new constructions are needed to obtain a satisfying structural description of balanced words.

Acknowledgments.

This work has been partially supported by the LabEx PERSYVAL-Lab (ANR–11-LABX-0025).

Yves Crama has been supported by Grenoble Institute of Technology and by the Interuniversity Attraction Poles Program initiated by the Belgian Science Policy Office (grant P7/36).

[1] E. Altman, B. Gaujal, and A. Hordijk. Balanced sequences and optimal routing. *Journal of the ACM* 47 (1999) 752–775.

[2] J. Barát and P.P. Varjú, Partitioning the positive integers to seven Beatty sequences, *Indagationes Mathematicae* 14 (2003) 149–161.

[3] N. Brauner and V. Jost, Small deviations, JIT sequencing and symmetric case of Fraenkel's conjecture, *Discrete Mathematics* 308 (2008) 2319–2324.

[4] A.S. Fraenkel, Complementing and exactly covering sequences, *Journal of Combinatorial Theory (Ser. A)* 14 (1973) 8–20.

[5] R.L. Graham, Covering the positive integers by disjoint sets of the form $\{[n\alpha + \beta] : n = 1, 2, \ldots\}$, *Journal of Combinatorial Theory (Ser. A)* 15 (1973) 354–358.

[6] J.W. Herrmann, Finding optimally balanced words for production planning and maintenance scheduling, *IIE Transactions* 44 (2012), 215–229.

[7] P. Hubert, Suites équilibrées, *Theoretical Computer Science* 242 (2000), 91–108.

[8] W. Kubiak, Fair sequences, in: J.Y-T. Leung, ed., *Handbook of Scheduling: Algorithms, Models and Performance Analysis*, Chapman & Hall/CRC, Boca Raton, FL, 2004, pp. 19-1–19–21.

[9] W. Kubiak, *Proportional Optimization and Fairness*, Springer, New York, NY, 2009.

[10] M. Morse and G. A. Hedlund, Symbolic Dynamics, *American Journal of Mathematics* 60:4 (1938), 815–866.

[11] S. Sano, N. Miyoshi and Ryohei Kataoka, m-Balanced words: A generalization of balanced words, *Theoretical Computer Science* 314:1 (2004), 97–120.

[12] R. J. Simpson, Disjoint covering systems of rational Beatty sequences, *Discrete Mathematics* 92:1-3 (1991), 361–369.

[13] R. Tijdeman, Fraenkel's conjecture for six sequences, *Discrete Mathematics* 222 (2000) 223–234.

[14] R. Tijdeman, Exact covers of balanced sequences and Fraenkel's conjecture, in: F. Halter-Koch and R.F. Tichy, eds, *Algebraic Number Theory and Diophantine Analysis*, Walter de Gruyter, Berlin, New York, 2000, pp. 467–483.

[15] R. Tijdeman, Periodicity and almost-periodicity, in: E. Győry, G.O.H. Katona and L. Lovász, eds., *More Sets, Graphs and Numbers,* Springer and Bolyai Mathematical Society, Berlin Heidelberg New York Budapest, 2006, pp. 381-405.

[16] L. Vuillon, Balanced words, *Bulletin of the Belgian Mathematical Society Simon Stevin* 10 (2003) 787–805.

A. Proof of Theorem 5.2

If a balanced word on 4 letters is not periodic, it is a congruence expansion of a word on 2 letters (see Theorem 2.11) and hence, it is of the form

$$\left(\frac{\alpha}{3}, \frac{\alpha}{3}, \frac{\alpha}{3}, 1-\alpha\right), \left(\frac{\alpha}{2}, \frac{\alpha}{4}, \frac{\alpha}{4}, 1-\alpha\right), \left(\frac{\alpha}{2}, \frac{\alpha}{2}, \frac{1-\alpha}{2}, \frac{1-\alpha}{2}\right).$$

Notice that the density vector of a (periodic balanced) word on 4 letters with exactly 2 distinct values is of the form $\left(\frac{\alpha}{2}, \frac{\alpha}{2}, \frac{1-\alpha}{2}, \frac{1-\alpha}{2}\right)$ or $\left(1-\alpha, \frac{\alpha}{3}, \frac{\alpha}{3}, \frac{\alpha}{3}\right)$. If a balanced word on 4 letters has a density vector with 4 distinct values then, by [1], this vector is $\left(\frac{8}{15}, \frac{4}{15}, \frac{2}{15}, \frac{1}{15}\right)$.

We therefore consider the remaining case, *i.e.* a periodic balanced word W on 4 letters with exactly 3 distinct densities $\delta_a \geq \delta_b \geq \delta_c \geq \delta_d$. One has $\delta_a > \delta_c$ and $\delta_b > \delta_d$.

The sketch of the proof of Theorem 5.2 follows from the claims hereunder:

- If W contains aaa, then its density vector is of the form $\left(1-\alpha, \frac{\alpha}{2}, \frac{\alpha}{4}, \frac{\alpha}{4}\right)$ (Claim 1)
- if W does not contain aaa,
  - If $\delta_a = \delta_b$ then the densities are $(\frac{4}{11}, \frac{4}{11}, \frac{2}{11}, \frac{1}{11})$ (Claim 2)
  - If $\delta_a > \delta_b$, then
    - W cannot contain bab as a factor (Claim 3)
    - If W contains $baab$ then the densities are $(\frac{8}{14}, \frac{4}{14}, \frac{1}{14}, \frac{1}{14})$ or $(\frac{6}{11}, \frac{3}{11}, \frac{1}{11}, \frac{1}{11})$ (Claim 4)
    - $\delta_a < 2\delta_c$ is impossible (Claim 5)
    - If $\delta_a = 2\delta_c$ then the densities are $(\frac{\alpha}{2}, 1-\alpha, \frac{\alpha}{4}, \frac{\alpha}{4})$ or $(\frac{\alpha}{2}, \frac{\alpha}{4}, \frac{\alpha}{4}, 1-\alpha)$ (Claim 6)
    - We then consider that $\delta_a > 2\delta_c$ (Claim 7)
      - $\delta_b > 2\delta_c$ is impossible
      - If $\delta_b = 2\delta_c$ then the densities are $(1-\alpha, \frac{\alpha}{2}, \frac{\alpha}{4}, \frac{\alpha}{4})$
      - If $\delta_b < 2\delta_c$ then the densities are $(\frac{6}{11}, \frac{2}{11}, \frac{2}{11}, \frac{1}{11})$

Before proving the claims, we state some preliminary general lemmas on balanced words.

Lemma A.1. *Let s and t be two integers. If a balanced sequence S has density $\delta_S > t/s$ then there exists an interval of s integers with at least $t+1$ elements from S.*

Proof. Suppose that any interval of s integers contains at most t elements from S. By partitionning the integers into a sequence of intervals of size s, we see that the density of S is at most t/s. □

Moreover, the converse is true if S is periodic:

Lemma A.2. *Let s and t be two integers. Let S be a periodic balanced sequence. S has density $\delta_S > t/s$ if and only if there exists an interval of s integers with at least $t+1$ elements from S.*

Proof. Suppose the last condition holds. By balancedness, each interval of size s has at least t elements. Partition the integers into a sequence of intervals of size s. Let D be a period of S. In every set of D consecutive intervals of size s, there is at least one interval with at least $t+1$ elements. Hence the density is at least $t/s + 1/Ds$. □

Lemma A.3. *Let W be a periodic balanced word, and x and y be two letters of W and k be an integer. Then $\delta_x > k\delta_y$ if and only if there exists a factor containing $k+1$ x's and no y.*

Proof. Assume that W contains a factor I with $k+1$ x's and no y. Partition W into factors of size $|I|$. By balancedness, in each such factor, there are at least k x's and at most one y. By periodicity, a factor with $k+1$ x's and no y appears sufficiently often so that $\delta_x > \frac{k}{|I|}$ and $\delta_y < \frac{1}{|I|}$. Hence, $\delta_x > k\delta_y$.

Assume that between two consecutive y, there are at most k x's. Partition W in factors starting with a y and with no other y inside. Each factor contains at most k x's. Approximating the densities (which exist either by periodicity or by balancedness) on any number of consecutive such factors, one obtains $\delta_x \leq k\delta_y$. □

Lemma A.4. *Let W be a periodic balanced word, and x and y be two letters of W and k be an integer. Then $\delta_x < k\delta_y$ if and only if there exists a factor starting and ending with a y and no other y and strictly less than k x's.*

Proof. Assume there exists a factor I starting and ending with a y and no other y and strictly less than k x's. Partition W into factors of size $|I|$. By balancedness, in each such factor, there are at most k x's and at least one y. By periodicity, a factor with strictly less than k x's between 2 y's appears sufficiently often so that $\delta_x < \frac{k}{|I|}$ and $\delta_y > \frac{1}{|I|}$. Hence, $\delta_x < k\delta_y$.

Assume that between two consecutive y, there are at least k x's. Partition W in factors starting with a y and with no other y inside. Each factor contains at least k x's. We thus obtain that $\delta_x \geq k\delta_y$. □

In particular, the case $k = 1$ means that there is a factor spanned by 2 y and containing no x.

Notice that Lemmas A.2 to A.4 also apply to Beatty sequences but they cannot be generalized to non balanced sequences. The following lemma is an immediate consequence of the last two lemmas.

Lemma A.5. *Let W be a periodic balanced word, and x and y be two letters of W with densities $\delta_x = \delta_y$. Then, letters x and y alternate in W.*

Let X be a factor. The notation $\underline{X}$ means that we do not make any assumption on the order of the letters of X. For instance $a\underline{bc}b$ represents either $abcb$ or $acbb$.

Claim 1. *A periodic balanced word on 4 letters, with exactly 3 distinct densities and containing aaa as a factor, has densities of the form $(1-\alpha, \frac{\alpha}{2}, \frac{\alpha}{4}, \frac{\alpha}{4})$.*

Proof. If a periodic balanced word W on 4 letters contains the factor aaa, then by Lemma A.2, $\delta_a \geq 2/3$. In this case, Simpson's Theorem [12] indicates that the word W induced by removing a is a congruence word on 3 letters. The unique possible density vectors of a congruence word on 3 letters are $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, see [1]. Therefore, W has a density vector of the form $(1-\alpha, \frac{\alpha}{3}, \frac{\alpha}{3}, \frac{\alpha}{3})$ or $(1-\alpha, \frac{\alpha}{2}, \frac{\alpha}{4}, \frac{\alpha}{4})$. Assuming 3 distinct densities, only remains the latter. □

We therefore consider in the sequel that W contains no aaa.

Claim 2. *A balanced word on 4 letters with $\delta_a = \delta_b > \delta_c > \delta_d$ and no aaa as a factor has densities of the form $(\frac{4}{11}, \frac{4}{11}, \frac{2}{11}, \frac{1}{11})$.*

Proof. Condition $\delta_a = \delta_b$ implies that between two a, there is a b (Lemma A.5) and conversely. Therefore, aa is not a factor of W. In this proof, the factors are given up to a renaming of a and b.

By Lemma A.2, W contains aba if and only if $\frac{1}{3} < \delta_a = \delta_b$ which is in turn equivalent to W contains a factor of 3 letters with 2 b which alternate with an a. Therefore, W contains aba if and only if it contains bab.

If W contains no aba, then it has no bab. Since $\delta_c > \delta_d$, W contains $c\underline{ab}c$ and since $\delta_a > \delta_c$, it also contains $a\underline{bd}a$ (Lemma A.4 with $k = 1$). This contradicts balancedness for c on 4 letters.

Therefore W contains aba (and hence bab). It implies that each factor of 3 letters contains at least an a and a b.

Since $\delta_c > \delta_d$, the word W contains a factor of the form $c(ab)^k a^l c$ (or $c(ba)^k b^l c$, but with the renaming of a and b we won't consider this second case) for some value $k \geq 1$ and $l = 0$ or 1. It implies that every factor of $2k-1$ letters contains at least $(k-1)$ times a and $(k-1)$ times b and every factor of $2k+l+2$ letters contains at least a c. This implies that around each d the factor is $(ab)^{k-1}d(ab)^{k-1}$ with $4k-3$ letters without any c. Hence $4k-3 < 2k+l+2$ with $l \leq 1$ which implies $k \leq 2$.

Word W contains $abdab$ (it contains a d and in 3 consecutive letters there is at least an a and a b). Therefore, there are at least 4 letters between two consecutive c which implies $k = 2$.

If $l = 1$, then W contains $cababac$ and every 5 consecutive letters contain at least 2 a and also 2 b by Lemma A.2 together with $\delta_a = \delta_b$. Therefore, letter d is contained in $abdabab$ which contradicts the balancedness of c on 7 letters.

Therefore, every factor spanned by 2 consecutive c not containing a d is $cababc$. It implies that around a d, one has $cabdabc$ (3 letters have at least 1 a and 1 b and 6 letters have at least one c). Let $C = cabab$ and $D = cabdab$ and

W can be viewed as a word on C and D where one can exchange a and b in C or D.

Extending C with the conditions on a and b leads to $abcabab$. Therefore, 7 letters contain at most one d and hence DD contradicts balancedness of d on 7 letters.

Extending D leads to $bcabdabca$ and hence 8 letters contain at most 3 a and 3 b. But $CC = cababcabab$ contradicts the balancedness of a and b on 8 letters.

Therefore, C and D alternate in W and hence the density vector is $(\frac{4}{11}, \frac{4}{11}, \frac{2}{11}, \frac{1}{11})$.

□

We assume in the sequel that $\delta_a > \delta_b$.

Claim 3. *A balanced word W on 4 letters with $\delta_a > \delta_b$ and exactly 3 distinct densities cannot contain bab as a factor.*

Proof. Suppose that W contains bab. Since $\delta_a > \delta_b$, there exists a factor with two a and no b. Consider a longest factor X with this property. If $|X| = 2$ then $X = aa$, and every 2 letters contains an a. Hence around c, one has aca which contradicts the maximality of X. Hence $|X| \geq 3$ which, together with bab contradicts balancedness of b on 3 letters. □

We assume in the sequel that W does not contain bab as a factor.

Claim 4. *A balanced word W on 4 letters with $\delta_a > \delta_b$ and exactly 3 distinct densities and containing baab as a factor has density vector either $(\frac{8}{14}, \frac{4}{14}, \frac{1}{14}, \frac{1}{14})$ or $(\frac{6}{11}, \frac{3}{11}, \frac{1}{11}, \frac{1}{11})$.*

Proof. The word W contains $baab$ implies $\delta_b > \delta_c$ by Lemma A.3 with $k = 1$. Hence we have $\delta_c = \delta_d$.

Because of $baab$, between 2 b's, there are at most 3 letters and every two consecutive letters contain an a. Let $A = baa$ and $C = baca$ and $D = bada$ Word W can be viewed as a word on A, C and D each of them appearing at least once. $AA = baabaa$ is impossible since it contradicts the balancedness of a on 5 letters because of the existence of $bacab$.

Since $\delta_c = \delta_d$, C and D alternate by Lemma A.5 and hence W contains CAD or DAC around an A. In both cases, W contains a factors of 9 letters with no d and a factor of 9 letters with no c. Therefore W can contain neither DCD nor CDC (9 letters with 2 d and 9 letters with 2 c which contradicts the balancedness of c and d).

ACA (and ADA) is incompatible with CD and DC. Indeed, $ACA = baabacabaa$ contains 9 letters with 6 a. But both CD and DC (necessarily followed by a b) imply 9 letters with 4 a. This contradicts balancedness for a on 9 letters.

Assume that W contains CAD. If CAD is followed by A, then $W = (CADA)^*$ with density vector $(\frac{8}{14}, \frac{4}{14}, \frac{1}{14}, \frac{1}{14})$. If CAD is followed by C then $W = (DCA)^*$ with density vector $(\frac{6}{11}, \frac{3}{11}, \frac{1}{11}, \frac{1}{11})$.

If W contains no CAD, then it contains DAC and the same reasonning applies. □

We assume in the sequel that W does not contain $baab$ as a factor.

Claim 5. *There does not exist a balanced word W on 4 letters, with exactly 3 distinct densities, such that $\delta_a > \delta_b$ and $\delta_a < 2\delta_c$.*

Proof.

In W, let X be a factor spanned by 2 consecutive c and containing exactly one a. X exists because of the condition $\delta_a < 2\delta_c$ (Lemma A.4 with $k = 2$). Moreover, since $\delta_a > \delta_c$, there also exists a word with two a and no c. Because of the conditions on the density, X contains one a, zero or one d and one or two b. Since W does not contain bab (Claim 3), $X \neq cbabc$.

- Case 1. $X = c\underline{ab}c$

  Then, between 2 consecutive c there are at most 3 letters. Moreover, since $\delta_a > \delta_c$, there are 2 consecutive c with 2 a in between. Hence, every 3 letters contain at least an a.

A factor containing *d* must be of the form *cdabc* or *cbadc* (always an *a* and a *b* and at most 3 letters between two *c* and every 3 letters contains an *a*).

Since $\delta_b \geq \delta_c$, there are at most 3 letters between two *b* and hence the previous factors expand to *bcdabc* or *cbadcb* which contradicts balancedness of *a* on 3 letters.

- Case 2. $X = c\underline{abbd}c$ where $\underline{abbd}$ means that those letters can be in any order. In this case, 4 letters contain at least one *b*.

  Since $\delta_a > \delta_b$, there exists a factor with 2 *a* and no *b*. Because of the previous condition, this factor is of length at most 3. It implies that W contains *aa* or *aca* or *ada*. Hence, every 3 letters contain an *a*. This is in contradiction with $X = c\underline{abbd}c$ which is a 6 letters word with only one *a*.

- Case 3. $X = c\underline{abd}c$ and $\delta_b > \delta_c$. Hence, there are at most 4 letters between two consecutive *c*. Since $\delta_b > \delta_c$, there exists 2 *b* between 2 consecutive *c*. Moreover, between 2 *b*, there is at least one *a* and *bab* is forbidden. Therefore, W contains $c\underline{abbd}c$ and Case 2 applies.

- Case 4. $c\underline{abd}c$ and $\delta_b = \delta_c$.

  Word W contains *bacab* (2 *b* with no *d*, no *aa* because of *bc* or *dc*, no *bacb* because already done in Case 1).

  Therefore, W contains *bcadb* or *bdacb* (only one *a* between two consecutive *b*). Hence, between 2 *a* there are one or two letters.

  Extending *bcadb* leads to *abcadba* and then to *dabcadbac*: the first letter is a *d* because *aa* is forbidden and there are at least 3 letters between two consecutive *b* or *c*. The last letter is a *c* because there are at most 4 letters between 2 *c*.

  Therefore, we know that there are 4 or 5 letters between 2 *d* (*bacab* and *dabcad*) and 3 or 4 letters between 2 *c* and between 2 *b*.

  We now prove that W cannot contain both *badc* and *bc*. Suppose W contains a factor the form $badcXbc$ or $bcXbadc$ for some factor X. Since

b and c alternate in X, there are the same number x of b and c in X. In $cXbc$ (resp. $bcXb$), there is $x+2$ times the the letter c (resp. b) and in $adcS$ (resp. $Xbad$) there are x times the letter b (resp. c) and both words have the same length. This is in contradiction with $\delta_b = \delta_c$. The same reasonning applies for proving that W cannot contain both *dcab* and *cb*.

Extending *cabac* leads to *badcabacdab* (at most 5 letters between two d, every three letters an a, at most 4 letters between two b). But W contains *cbadc* or *cdabc* (there exists a factor beggining and ending with c and only one a and every such factor contains at least 3 letters), hence W contains *cb* or *bc*. But *badcabacdab* contains *badc* and *cdab* This is in contradiction with the previous result.

□

Claim 6. *A balanced word W on 4 letters, with exactly 3 distinct densities, such that $\delta_a > \delta_b$ and $\delta_a = 2\delta_c$ has densities of the form $\delta = (\frac{\alpha}{2}, 1-\alpha, \frac{\alpha}{4}, \frac{\alpha}{4})$ or $\delta = (\frac{\alpha}{2}, \frac{\alpha}{4}, \frac{\alpha}{4}, 1-\alpha)$.*

Proof. With $\delta_a = 2\delta_c$ and $\delta_a > \delta_b$, if $\delta_b = \delta_c$ then, the density is of the form $\delta = (\frac{\alpha}{2}, \frac{\alpha}{4}, \frac{\alpha}{4}, 1-\alpha)$ and if $\delta_c = \delta_d$, then it is of the form $\delta = (\frac{\alpha}{2}, 1-\alpha, \frac{\alpha}{4}, \frac{\alpha}{4})$. □

Claim 7. *A balanced word W on 4 letters, with exactly 3 distinct densities, such that $\delta_a > \delta_b$ and $\delta_a > 2\delta_c$ satisfies $\delta = (1-\alpha, \frac{\alpha}{2}, \frac{\alpha}{4}, \frac{\alpha}{4})$ or $\delta = (\frac{6}{11}, \frac{2}{11}, \frac{2}{11}, \frac{1}{11})$.*

Proof. The condition $\delta_a > 2\delta_c$ implies that there exists a factor with at least 3 a between 2 c (Lemma A.3 with $k=2$). Let X be such a factor. Then, X is of the form $X = cX'c$ where X' contains no c and contains at least 3 a.

If $\delta_b > 2\delta_c$, then there exists 3 b between two consecutive c. Hence W contains ba^kb with $k \geq 1$ which case has already been eliminated.

If $\delta_b = 2\delta_c$, then we are done: the density is of the form $(1-\alpha, \frac{\alpha}{2}, \frac{\alpha}{4}, \frac{\alpha}{4})$.

Hence, we consider the remaining case $\delta_b < 2\delta_c$. It implies that there exists a factor with only one b between two consecutive c (Lemma A.4 with $k=2$).

- Case 1. $X = caabaac$ and $\delta_b = \delta_c$.

  Every 5 letters contain at least 3 a and there are at most 6 letters between 2 c. Therefore, every d is included in the factor $aacabaadacaa$ up to a symmetry. But $\delta_b = \delta_c$ implies that every 7 letters contain at least a b which is not the case in the previous factor.

- Case 2. $X = cabaac$ and $\delta_b = \delta_c$.

  Then the only other possibility between two c up to a symmetry is $cabadac$ (at most 5 letters between 2 c and every 4 letters contain at least 2 a). Denote $A = caaba$ and $\overline{A} = cabaa$ and $B = cabada$ and $\overline{B} = cadaba$. W can be viewed as a word on letters A, $\overline{A}$, B and $\overline{B}$.

  The factor $AA = c[aabacaa]bac$ contradicts balancedness for a on 7 letters with B or $\overline{B}$ followed by a c by definition (the same for $\overline{A}\,\overline{A}$).

  $BB = caba[dacabad]ac$ or $B\overline{B}$ or $\overline{B}\,\overline{B}$ contradict balancedness for d on 7 letters with A or $\overline{A}$ followed by ca by definition.

  $A\overline{A} = caa[bacab]aa$ or $\overline{B}B$ contradict balancedness for b on 5 letters with $cab[adaca]$.

  $\overline{A}A = cabaacaaba$ has already been considered in Case 1 (exchanging the role of b and c since $\delta_b = \delta_c$).

  Therefore, A or $\overline{A}$ alternate with B or $\overline{B}$ which implies densities verifying $(\frac{6}{11}, \frac{2}{11}, \frac{2}{11}, \frac{1}{11})$.

- Case 3. $X = cabadac$ and $\delta_b = \delta_c$.

  Then W contains $caabac$ or $caabaac$ (balancedness of c when there is no d between two consecutive c). This has already been considered in the previous cases.

  If X contains 4 a, the reasoning is similar. It cannot contain more a's because of $caabac$ or $caabaac$.

- Case 4. W contains aa and $\delta_c = \delta_d$.

There exists 2 consecutive c with only one b in between ($\delta_c > 2\delta_b$ and there also exists 2 consecutive c with 2 b in between ($\delta_c < \delta_b$). Letters c and d alternate. Therefore W contains $A = ca^*ba^xdbac$ and $B = ca^lba^ydca$ with $l \leq 2$. in A there are at least $5 + x$ letters between two c. Hence, $4 + x \leq l + 2 + y$ for balancedness of c. Since $l \leq 2$, one obtains $y \geq x$. In B, there are $y + 3 \geq x + 3$ letters without a b and in A, there are $3 + x$ letters with two b contradicting balancedness on b.

- Case 5. W does not contain aa and $\delta_c = \delta_d$.

  There exists 3 a between 2 consecutive c. Therefore W contains $c\underline{aaabd}c$ or $c\underline{aaabbd}c$ and hence, every 3 letters contain at least an a.

  There exists a factor with one b between 2 c. Therefore, if W contains $c\underline{aaabbd}c$ then, it contains $cabadac$ or $cadabac$ (at least 5 letters between 2 c).

  Therefore, there exists 4 consecutive letters with no b. But there exists 2 b between consecutive c therefore, W contains $cbadabc$ or $cabadabc$ or $cbadabac$.

  We now prove that W cannot contain both $cabad$ and cad. Suppose W is of the form $cabadXcad$ for some word X. Since c and d alternate, there are the same number μ of c and d in X. In $dXcad$ there are $\mu + 2$ times the letter d and in $abadS$, there are μ times letter c and both words have the same length. This is in contradiction with $\delta_c = \delta_d$. Similarly, one can prove that W cannot contain both $dabac$ and dac (exchange the roles of c and d in the proof).

  If W contains $cbadabc$, then it is extended to $dacbadabcad$. But W also contains $cabadac$ or $cadabac$. In the former case, it contains both $cabad$ and cad and in the latter case, it contains $dabac$ and dac, which contradicts the previous paragraph.

  If W contains $cabadabc$, then it is extended to $cabadabcad$ which contains both $cabad$ and cad again leading to a contradiction. Similarly, $cbadabac$

extends to *dacbadabac* which contains both *dac* and *dabac*.

$\square$