

# $k$ -SPECTRA OF $c$ -BALANCED WORDS

## $k$ -SPECTRA

---

Pamela Fleischmann, Joel D. Day

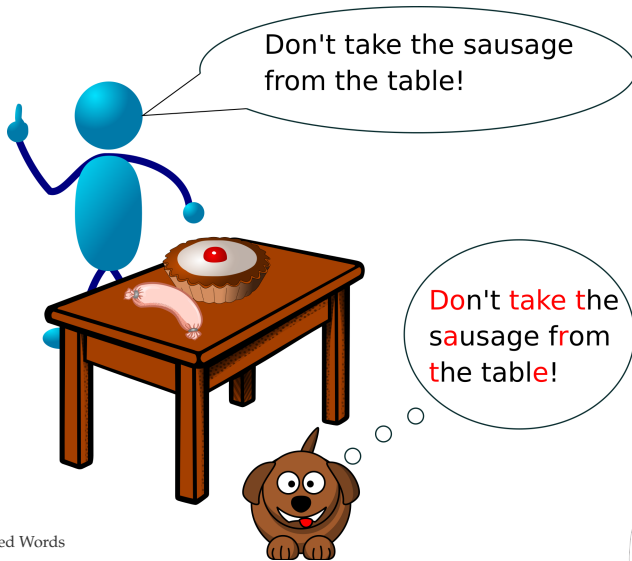
{fpa,jda}@informatik.uni-kiel.de

September, 13<sup>th</sup>

Kiel University  
Dependable Systems Group



# Information Loss



# Scattered Factors

**informal:** deleting arbitrary letters from a word (preserving the order) results in a scattered factor of this word



# Scattered Factors

**informal:** deleting arbitrary letters from a word (preserving the order) results in a scattered factor of this word

EXAMPLE

b a n a n a



# Scattered Factors

**informal:** deleting arbitrary letters from a word (preserving the order) results in a scattered factor of this word

EXAMPLE

a a a



# Scattered Factors

**informal:** deleting arbitrary letters from a word (preserving the order) results in a scattered factor of this word

EXAMPLE

b a n a n a



# Scattered Factors

**informal:** deleting arbitrary letters from a word (preserving the order) results in a scattered factor of this word

EXAMPLE

b a n a



# Scattered Factors

**informal:** deleting arbitrary letters from a word (preserving the order) results in a scattered factor of this word

## Definition (Scattered Factor, (Scattered) Subword)

$v = v_1 \dots v_n \in \Sigma^*$  scattered factor of  $w$  iff

$$\exists u_0 \dots u_n \in \Sigma^* : w = u_0 v_1 u_1 \dots u_{n-1} v_n u_n.$$





## Definition

- set of all scattered factors of  $w$  is the **spectrum**  $\text{ScatFact}(w)$



## Definition

- set of all scattered factors of  $w$  is the **spectrum**  $\text{ScatFact}(w)$
- set of all scattered factors of  $w$  of length  $k$  is the  **$k$ -spectrum**  $\text{ScatFact}_k(w)$



## Definition

- set of all scattered factors of  $w$  is the **spectrum**  $\text{ScatFact}(w)$
- set of all scattered factors of  $w$  of length  $k$  is the  **$k$ -spectrum**  $\text{ScatFact}_k(w)$

Example: abba

{ abba }	4-spectrum
{ aba, bba, abb }	3-spectrum
{ aa, ab, bb, ba }	2-spectrum
{ a, b }	1-spectrum



## Definition

- set of all scattered factors of  $w$  is the **spectrum**  $\text{ScatFact}(w)$
- set of all scattered factors of  $w$  of length  $k$  is the  **$k$ -spectrum**  $\text{ScatFact}_k(w)$

We are not considering multisets.



# Open Problems

## Problem

*Given  $S \subseteq \Sigma^*$  decide whether  $S$  is the spectrum of some word  $w$ .*



# Open Problems

## Problem

*Given  $S \subseteq \Sigma^*$  decide whether  $S$  is the spectrum of some word  $w$ .*

## Problem

*Given a  $k$ -spectrum decide whether it is independent, e.g.  $\{ab, ba, aa\}$  is not independent since  $aa$  can be deduced from  $ab$  and  $ba$ .*



# Open Problems

## Problem

*Given  $S \subseteq \Sigma^*$  decide whether  $S$  is the spectrum of some word  $w$ .*

## Problem

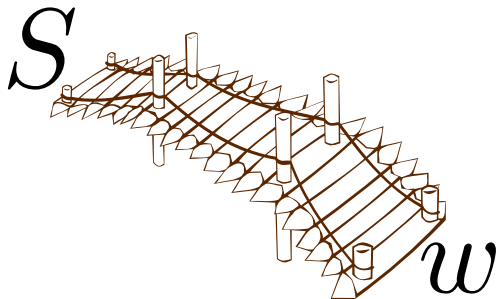
*Given a  $k$ -spectrum decide whether it is independent, e.g.  $\{ab, ba, aa\}$  is not independent since  $aa$  can be deduced from  $ab$  and  $ba$ .*

## Problem

*Determine the index of the equivalence relation that relates word with the same spectrum.*

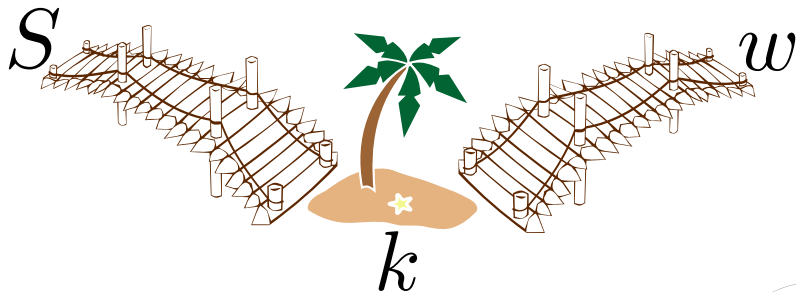


# Middle Step Between $S$ and $w$





# Middle Step Between $S$ and $w$



# Reformulated Problem

## Problem

*Decide for a given  $n \in \mathbb{N}$  whether there exists  $w \in \Sigma^*$  and  $k \in \mathbb{N}$  with  $|\text{ScatFact}_k(w)| = n$ .*



# Reformulated Problem

## Problem

*Decide for a given  $n \in \mathbb{N}$  whether there exists  $w \in \Sigma^*$  and  $k \in \mathbb{N}$  with  $|\text{ScatFact}_k(w)| = n$ .*

or more restricted:

## Problem

*Decide for given  $n, k \in \mathbb{N}$  whether there exists  $w \in \Sigma^*$  with  $|\text{ScatFact}_k(w)| = n$ .*



# Reformulated Problem

## Problem

*Decide for a given  $n \in \mathbb{N}$  whether there exists  $w \in \Sigma^*$  and  $k \in \mathbb{N}$  with  $|\text{ScatFact}_k(w)| = n$ .*

or more restricted:

## Problem

*Decide for given  $n, k \in \mathbb{N}$  whether there exists  $w \in \Sigma^*$  with  $|\text{ScatFact}_k(w)| = n$ .*

To start with we only consider a binary alphabet  $\Sigma = \{a, b\}$ .



# Examples

- $n = 3, k = 2: w = aabb$



# Examples

- $n = 3, k = 2: w = aabb$
- $n = k + 2, k \in \mathbb{N}_{>2}, |w|_a = |w|_b$  does not have a solution



# Examples

- $n = 3, k = 2: w = aabb$
- $n = k + 2, k \in \mathbb{N}_{>2}, |w|_a = |w|_b$  does not have a solution
- $n = 2^k, k \in \mathbb{N}: w = (ab)^k$



# Examples

- $n = 3, k = 2: w = aabb$
- $n = k + 2, k \in \mathbb{N}_{>2}, |w|_a = |w|_b$  does not have a solution
- $n = 2^k, k \in \mathbb{N}: w = (ab)^k$
- $n$  square number at least 4:  $k := 2(\sqrt{n} - 1), w = a^{\frac{k}{2}}b^ka^{\frac{k}{2}}$





## Definition

Binary word  $w \in \{a, b\}^*$   **$c$ -balanced** for a  $c \in \mathbb{N}_0$  iff

$$||w|_a - |w|_b| = c.$$



## Definition

Binary word  $w \in \{a, b\}^*$   $c$ -balanced for a  $c \in \mathbb{N}_0$  iff

$$||w|_a - |w|_b| = c.$$

- $c = 1$ :  $w$  balanced



## Definition

Binary word  $w \in \{a, b\}^*$   $c$ -balanced for a  $c \in \mathbb{N}_0$  iff

$$||w|_a - |w|_b| = c.$$

- $c = 1$ :  $w$  balanced
- $c = 0$ :  $w$  strictly balanced



## Definition

Binary word  $w \in \{a, b\}^*$   $c$ -balanced for a  $c \in \mathbb{N}_0$  iff

$$||w|_a - |w|_b| = c.$$

- $c = 1$ :  $w$  balanced
- $c = 0$ :  $w$  strictly balanced

Obviously for every  $w \in \{a, b\}^*$  exists  $c \in \mathbb{N}_0$  such that  $w$  is  $c$ -balanced.



# Peculiarities of Restriction to Cardinalities

## Example: 3-spectrum

abbab	baaba	babba	abaab
	aaa		aaa
aab	aab		aab
aba	aba	aba	aba
abb		abb	abb
	baa	baa	baa
bab	bab	bab	bab
bba	bba	bba	
bbb		bbb	
6	6	6	6



# Peculiarities of Restriction to Cardinalities

## Example: 3-spectrum

abbab	baaba	babba	abaab
aab	aaa		aaa
aba	aab		aab
abb	aba	aba	aba
		abb	abb
	baa	baa	baa
bab	bab	bab	bab
bba	bba	bba	
bbb		bbb	
6	6	6	6

	renaming	reverse	both
abbab	baaba	babba	abaab
aab	bba	baa	abb
aba	bab	aba	bab
abb	baa	bba	aab
bab	aba	bab	aba
bba	aab	abb	baa
bbb	aaa	bbb	aaa
6	6	6	6



# Renaming and Reversing

## Definition

- $\bar{\cdot} : \Sigma \rightarrow \Sigma$  with  $\bar{a} = b$  and  $\bar{b} = a$  **renaming morphism**



# Renaming and Reversing

## Definition

- $\bar{\cdot} : \Sigma \rightarrow \Sigma$  with  $\bar{a} = b$  and  $\bar{b} = a$  **renaming morphism**
- $\cdot^R : \Sigma^* \rightarrow \Sigma^*$  with  $w^R = w[|w|] \dots w[1]$  with the  $i^{\text{th}}$  letter  $w[i]$  of  $w$





# Renaming and Reversing

## Definition

- $\bar{\cdot} : \Sigma \rightarrow \Sigma$  with  $\bar{a} = b$  and  $\bar{b} = a$  a **renaming morphism**
- $\cdot^R : \Sigma^* \rightarrow \Sigma^*$  with  $w^R = w[|w|] \dots w[1]$  with the  $i^{\text{th}}$  letter  $w[i]$  of  $w$

## Lemma

- $\text{ScatFact}(\bar{w}) = \{\bar{u} \mid u \in \text{ScatFact}(w)\}$



# Renaming and Reversing

## Definition

- $\bar{\cdot} : \Sigma \rightarrow \Sigma$  with  $\bar{a} = b$  and  $\bar{b} = a$  a **renaming morphism**
- $\cdot^R : \Sigma^* \rightarrow \Sigma^*$  with  $w^R = w[|w|] \dots w[1]$  with the  $i^{\text{th}}$  letter  $w[i]$  of  $w$

## Lemma

- $\text{ScatFact}(\bar{w}) = \{\bar{u} \mid u \in \text{ScatFact}(w)\}$
- $\text{ScatFact}(w^R) = \{u^R \mid u \in \text{ScatFact}(w)\}$



# Peculiarities of Restriction to Cardinalities

## Corollary

*The cardinalities of the spectra (and  $k$ -spectra) of  $w$ ,  $w^R$ , and  $\bar{w}$  are the same:*

$$|\text{ScatFact}_k(w)| = |\text{ScatFact}_k(w^R)| = |\text{ScatFact}_k(\bar{w})|.$$



# Peculiarities of Restriction to Cardinalities

## Corollary

*The cardinalities of the spectra (and  $k$ -spectra) of  $w$ ,  $w^R$ , and  $\bar{w}$  are the same:*

$$|\text{ScatFact}_k(w)| = |\text{ScatFact}_k(w^R)| = |\text{ScatFact}_k(\bar{w})|.$$

$a < b$  assumed: only consider the lexicographically smallest element in such a equivalence class



# Solving the first problem

## Theorem

*For all  $n \in \mathbb{N}$  the  $k$ -spectrum of  $w = a^k b^k$  for  $k = n - 1$  has  $n$  elements, i.e.  $|\text{ScatFact}_{n-1}(a^{n-1} b^{n-1})| = n$ .*



# Solving the first problem

## Theorem

For all  $n \in \mathbb{N}$  the  $k$ -spectrum of  $w = a^k b^k$  for  $k = n - 1$  has  $n$  elements, i.e.  $|\text{ScatFact}_{n-1}(a^{n-1}b^{n-1})| = n$ .

Proof:

- all  $a^r b^s$  for  $r + s = n - 1$  are the scattered factors of length  $n - 1$



# Solving the first problem

## Theorem

For all  $n \in \mathbb{N}$  the  $k$ -spectrum of  $w = a^k b^k$  for  $k = n - 1$  has  $n$  elements, i.e.  $|\text{ScatFact}_{n-1}(a^{n-1}b^{n-1})| = n$ .

Proof:

- all  $a^r b^s$  for  $r + s = n - 1$  are the scattered factors of length  $n - 1$
- $n$  possibilities

□



# Solving the first problem

## Theorem

For all  $n \in \mathbb{N}$  the  $k$ -spectrum of  $w = a^k b^k$  for  $k = n - 1$  has  $n$  elements, i.e.  $|\text{ScatFact}_{n-1}(a^{n-1}b^{n-1})| = n$ .

## Corollary

$S_n = \{a^r b^s \mid r + s = n \in \mathbb{N}\}$  is a scattered factor set for all  $n \in \mathbb{N}$ .





# Partly Solving the Second Problem

## Theorem

*Given  $k, n \in \mathbb{N}$  with  $n - 1 \leq k$  set  $c = k - n + 1$  and consider  $w = a^k b^{k-c}$ . Then for all  $i \in [c]_0$  the  $(k - i)$ -spectrum of  $w$  has cardinality  $n$ .*



# Partly Solving the Second Problem

## Theorem

Given  $k, n \in \mathbb{N}$  with  $n - 1 \leq k$  set  $c = k - n + 1$  and consider  $w = a^k b^{k-c}$ . Then for all  $i \in [c]_0$  the  $(k - i)$ -spectrum of  $w$  has cardinality  $n$ .

Proof:

- $i = 0$ :  $a^r b^s$  with  $r + s = k \rightsquigarrow k - c + 1 = n$  possibilities
- $i \neq 0$ : all the scattered factor are just *shortened* for the  $(k - i)$ -spectra



# Are we happy now?

- Given  $n \in \mathbb{N}$  for each  $c$  we have  $c + 1$  different sets being a spectrum of cardinality  $n$ .



# Are we happy now?

- Given  $n \in \mathbb{N}$  for each  $c$  we have  $c + 1$  different sets being a spectrum of cardinality  $n$ .
- $k$ -spectrum of  $w = a^2b^4a^2$  has cardinality 9



# Are we happy now?

- Given  $n \in \mathbb{N}$  for each  $c$  we have  $c + 1$  different sets being a spectrum of cardinality  $n$ .
- $k$ -spectrum of  $w = a^2b^4a^2$  has cardinality 9
- $abba$  is a scattered factor of  $w$  and not in the aforementioned sets



# Are we happy now?

- Given  $n \in \mathbb{N}$  for each  $c$  we have  $c + 1$  different sets being a spectrum of cardinality  $n$ .
- $k$ -spectrum of  $w = a^2b^4a^2$  has cardinality 9
- $abba$  is a scattered factor of  $w$  and not in the aforementioned sets
- which scattered factor sets have cardinality  $n \in \mathbb{N}$



# Are we happy now?

- Given  $n \in \mathbb{N}$  for each  $c$  we have  $c + 1$  different sets being a spectrum of cardinality  $n$ .
- $k$ -spectrum of  $w = a^2b^4a^2$  has cardinality 9
- $abba$  is a scattered factor of  $w$  and not in the aforementioned sets
- which scattered factor sets have cardinality  $n \in \mathbb{N}$
- for a fixed  $c \in \mathbb{N}$  and  $c$ -balanced words: which cardinalities are *reachable*



# Are we happy now?

- Given  $n \in \mathbb{N}$  for each  $c$  we have  $c + 1$  different sets being a spectrum of cardinality  $n$ .
- $k$ -spectrum of  $w = a^2b^4a^2$  has cardinality 9
- $abba$  is a scattered factor of  $w$  and not in the aforementioned sets
- which scattered factor sets have cardinality  $n \in \mathbb{N}$
- for a fixed  $c \in \mathbb{N}$  and  $c$ -balanced words: which cardinalities are *reachable*

We were not happy! We would like to fully characterise for given  $c$  and word-length which cardinalities are reachable





## Lemma

For  $w \in \Sigma^*$  and  $k, c \in \mathbb{N}_0$  with  $c \leq k$  we have

$$\forall i \in [c]_0 : |\text{ScatFact}_{k-i}(w)| = k - c + 1 \quad \text{iff} \quad w = a^k b^{k-c}.$$

Moreover  $|\text{ScatFact}_{k-i}(w)| \geq k - c + 1$  for all  $i \in [c]_0$



# Minimal Cardinality

## Lemma

For  $w \in \Sigma^*$  and  $k, c \in \mathbb{N}_0$  with  $c \leq k$  we have

$$\forall i \in [c]_0 : |\text{ScatFact}_{k-i}(w)| = k - c + 1 \quad \text{iff} \quad w = a^k b^{k-c}.$$

Moreover  $|\text{ScatFact}_{k-i}(w)| \geq k - c + 1$  for all  $i \in [c]_0$

Proof idea for remaining part:

- suppose  $w \neq a^k b^{k-c}$  (neither one of the symmetric cases)
- $\Rightarrow w = w_1 a b a w_2$
- induction on word-length



$k$ -SPECTRA FOR STRICTLY BALANCED  
WORDS OF LENGTH  $2k$

---

# Properties of strictly balanced words

- same amount of as and bs



# Properties of strictly balanced words

- same amount of as and bs
- always even length



# Properties of strictly balanced words

- same amount of as and bs
- always even length
- the  $k$ -spectra has at most  $2^k$  elements



# Reaching the Maximal Cardinality

## Lemma

*The  $k$ -spectrum of a strictly balanced word  $w \in \Sigma^*$  has cardinality  $2^k$  iff  $w \in \{ab, ba\}^k$ , i.e.*

$$|\text{ScatFact}_k(w)| = 2^k \Leftrightarrow w \in \{ab, ba\}^k.$$



# Reaching the Maximal Cardinality

## Lemma

*The  $k$ -spectrum of a strictly balanced word  $w \in \Sigma^*$  has cardinality  $2^k$  iff  $w \in \{ab, ba\}^k$ , i.e.*

$$|\text{ScatFact}_k(w)| = 2^k \Leftrightarrow w \in \{ab, ba\}^k.$$

Sketch of Proof:





# Reaching the Maximal Cardinality

## Lemma

The  $k$ -spectrum of a strictly balanced word  $w \in \Sigma^*$  has cardinality  $2^k$  iff  $w \in \{ab, ba\}^k$ , i.e.

$$|\text{ScatFact}_k(w)| = 2^k \Leftrightarrow w \in \{ab, ba\}^k.$$

Sketch of Proof:

- " $\Rightarrow$ " contraposition; if  $aa$  is a factor of  $w$  then one  $b^r a^{k-r}$  is not a scattered factor



# Reaching the Maximal Cardinality

## Lemma

The  $k$ -spectrum of a strictly balanced word  $w \in \Sigma^*$  has cardinality  $2^k$  iff  $w \in \{ab, ba\}^k$ , i.e.

$$|\text{ScatFact}_k(w)| = 2^k \Leftrightarrow w \in \{ab, ba\}^k.$$

Sketch of Proof:

- " $\Rightarrow$ " contraposition; if  $aa$  is a factor of  $w$  then one  $b^r a^{k-r}$  is not a scattered factor
- " $\Leftarrow$ " induction



# Spectrum of $k$ -spectra



# Spectrum of $k$ -spectra



Proof for " $|\text{ScatFact}_k(w)| = k + 1$  iff  $w = a^k b^k$ " gives also that  $k + 2$  is not reachable!



# Spectrum of $k$ -spectra



Proof for " $|\text{ScatFact}_k(w)| = k + 1$  iff  $w = a^k b^k$ " gives also that  $k + 2$  is not reachable!



Jumbling a and b:  $a^k b^k \rightarrow (ab)^k$

if the as and bs are not nicely ordered we have at least one switch from b to a



# Jumbling a and b: $a^k b^k \rightarrow (ab)^k$

if the as and bs are not nicely ordered we have at least one switch from b to a

## Lemma

*The  $k$ -spectrum of a strictly balanced word  $w \in \Sigma^*$  has cardinality  $2k$  iff  $w$  is either  $a^{k-1}bab^{k-1}$  or  $a^{k-1}b^k a$ , i.e.*

$$|\text{ScatFact}_k(w)| = 2k \Leftrightarrow w \in \{a^{k-1}bab^{k-1}, a^{k-1}b^k a\}.$$



# Jumbling a and b: $a^k b^k \rightarrow (ab)^k$

if the as and bs are not nicely ordered we have at least one switch from b to a

## Lemma

*The  $k$ -spectrum of a strictly balanced word  $w \in \Sigma^*$  has cardinality  $2k$  iff  $w$  is either  $a^{k-1}bab^{k-1}$  or  $a^{k-1}b^k a$ , i.e.*

$$|\text{ScatFact}_k(w)| = 2k \Leftrightarrow w \in \{a^{k-1}bab^{k-1}, a^{k-1}b^k a\}.$$

Our proof also shows

- If  $w$  is neither  $a^k b^k$  nor  $a^{k-1}bab^{k-1}$  nor  $a^{k-1}b^k a$ , then the cardinality is greater than  $2k$





# Spectrum of $k$ -spectra



# Spectrum of $k$ -spectra

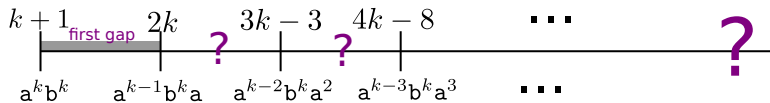


$a^{k-1} b^k a$  generalisable to  $a^{k-i} b^k a^i$  for  $i \in \llbracket \lfloor \frac{k}{2} \rrbracket \rrbracket$ :

$$|\text{ScatFact}_k(a^{k-i} b^k a^i)| = k(i+1) - i^2 + 1$$



# Spectrum of $k$ -spectra

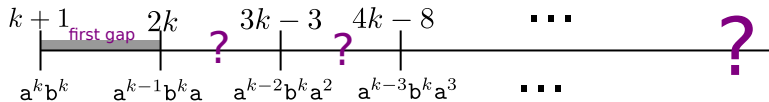


$a^{k-1} b^k a$  generalisable to  $a^{k-i} b^k a^i$  for  $i \in \llbracket \lfloor \frac{k}{2} \rrbracket \rrbracket$ :

$$|\text{ScatFact}_k(a^{k-i} b^k a^i)| = k(i+1) - i^2 + 1$$



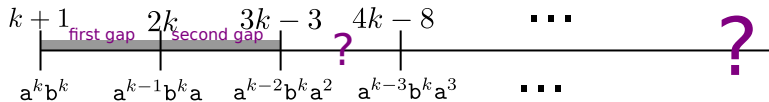
# Spectrum of $k$ -spectra



**Promising news:** the  $k$ -spectra of strictly balanced words cannot have cardinality  $2k + i$  for  $i \in [k - 4]$



# Spectrum of $k$ -spectra



**Promising news:** the  $k$ -spectra of strictly balanced words cannot have cardinality  $2k + i$  for  $i \in [k - 4]$



# The third gap . . .

Raised hope: gap between  $(i + 1)k - i^2 + 1$  and  $(i + 2)k - (i + 1)^2 + 1$



# The third gap . . .

Raised hope: gap between  $(i + 1)k - i^2 + 1$  and  $(i + 2)k - (i + 1)^2 + 1$

but (unfortunately)



# The third gap . . .

Raised hope: gap between  $(i + 1)k - i^2 + 1$  and  $(i + 2)k - (i + 1)^2 + 1$

but (unfortunately)

## Lemma

*The  $k$ -spectrum of  $a^{k-1}b^2ab^{k-2}$  has exactly  $3k - 2$  elements.*





# The third gap . . .

Raised hope: gap between  $(i + 1)k - i^2 + 1$  and  $(i + 2)k - (i + 1)^2 + 1$

but (unfortunately)

## Lemma

*The  $k$ -spectrum of  $a^{k-1}b^2ab^{k-2}$  has exactly  $3k - 2$  elements.*

and this result is generalisable



# The Thing in the "Gap"

## Lemma

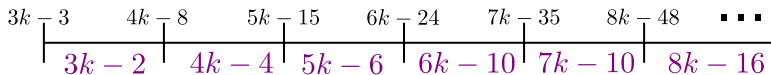
For  $k \geq 5$  and  $i \in [k - 1]$

- $|\text{ScatFact}_k(a^{k-2}b^i ab^{k-i}a)| = k(2i + 2) - 6i + 2$
- $|\text{ScatFact}_k(a^{k-2}b^i a^2 b^{k-i})| = k(2i + 1) - 4i + 2$



# Spectrum of $k$ -spectra

$$k \geq 38$$



# The other end of the spectrum

We saw already that the cardinality  $2^k$  is reached iff  $w = (ab)^k$ .



# The other end of the spectrum

We saw already that the cardinality  $2^k$  is reached iff  $w = (ab)^k$ .

## Lemma

*The  $k$ -spectrum of  $w$  has cardinality  $2^k - 1$  iff  $w = (ab)^i a^2 b^2 (ab)^{k-i-2}$  for some  $i \in [k - 2]$ .*



# The other end of the spectrum

We saw already that the cardinality  $2^k$  is reached iff  $w = (ab)^k$ .

## Lemma

*The  $k$ -spectrum of  $w$  has cardinality  $2^k - 1$  iff  $w = (ab)^i a^2 b^2 (ab)^{k-i-2}$  for some  $i \in [k - 2]$ .*

Proof:

- " $\Leftarrow$ "  $\checkmark$
- " $\Rightarrow$ " if there is a scattered factor not of the form  $b^{i+1} a^{k-i-1}$  then less than  $2^k - 1$  element are in the  $k$ -spectrum



# Overview for strictly balanced words

