Pamela Fleischmann, Joel D. Day, Florin Manea, and Dirk Nowotka

fpa@informatik.uni-kiel.de

August, 9<sup>th</sup>

Kiel University Dependable Systems Group



# Information Loss

k-Spectra



# Information Loss



### informal definition:

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

$$\exists u_0 \dots u_n \in \Sigma^* : w = u_0 v_1 u_1 v_2 \dots v_{n-1} u_{n-1} v_n u_n.$$



 $\bigcirc$  set of all scattered factors of *w* is the spectrum ScatFact(*w*)



- $\bigcirc$  set of all scattered factors of *w* is the spectrum ScatFact(*w*)
- set of all scattered factors of w of length k is the k-spectrum ScatFact<sub>k</sub>(w)



- $\bigcirc$  set of all scattered factors of *w* is the spectrum ScatFact(*w*)
- $\bigcirc$  set of all scattered factors of *w* of length *k* is the *k*-spectrum ScatFact<sub>k</sub>(*w*)

Example: abba

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



- $\bigcirc$  set of all scattered factors of *w* is the spectrum ScatFact(*w*)
- $\bigcirc$  set of all scattered factors of *w* of length *k* is the *k*-spectrum ScatFact<sub>k</sub>(*w*)

### Example: abba

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

We are not considering multisets.



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



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.



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



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



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

### or more restricted:

### Problem

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



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

or more restricted:

### Problem

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

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

$$\cap$$
 *n* = 3, *k* = 2: *w* = aabb



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



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



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

 $\bigcirc$  *n* square number at least 4,  $k := 2(\sqrt{n} - 1)$ :  $w = a^{\frac{\kappa}{2}}b^k a^{\frac{\kappa}{2}}$ 



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

$$||w|_{\mathsf{a}} - |w|_{\mathsf{b}}| = c.$$



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

$$||w|_{\mathsf{a}} - |w|_{\mathsf{b}}| = c.$$

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



○ ScatFact<sub>k</sub>(
$$\overline{w}$$
) = { $\overline{u}$  |  $u \in$  ScatFact<sub>k</sub>( $w$ )}  
○ ScatFact<sub>k</sub>( $w^R$ ) = { $u^R$  |  $u \in$  ScatFact<sub>k</sub>( $w$ )}



○ ScatFact<sub>k</sub>(
$$\overline{w}$$
) = { $\overline{u}$  |  $u \in$  ScatFact<sub>k</sub>( $w$ )}  
○ ScatFact<sub>k</sub>( $w^R$ ) = { $u^R$  |  $u \in$  ScatFact<sub>k</sub>( $w$ )}

$$|\operatorname{ScatFact}_k(w)| = |\operatorname{ScatFact}_k(w^R)| = |\operatorname{ScatFact}_k(\overline{w})|.$$



○ ScatFact<sub>k</sub>(
$$\overline{w}$$
) = { $\overline{u}$  |  $u \in$  ScatFact<sub>k</sub>( $w$ )}  
○ ScatFact<sub>k</sub>( $w^R$ ) = { $u^R$  |  $u \in$  ScatFact<sub>k</sub>( $w$ )}

 $|\operatorname{ScatFact}_k(w)| = |\operatorname{ScatFact}_k(w^R)| = |\operatorname{ScatFact}_k(\overline{w})|.$ 

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



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



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



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



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

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



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.



Given  $k, n \in \mathbb{N}$  with  $n - 1 \le k$  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:

- $\bigcirc$  *i* = 0: a<sup>*r*</sup>b<sup>*s*</sup> 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



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



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



- 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



- 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  $n \in \mathbb{N}$



- 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*



- 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
- $\bigcirc$  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

 $w \in \Sigma^*, k, c \in \mathbb{N}_0$  with  $c \leq k$ :

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

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



#### Lemma

 $w \in \Sigma^*, k, c \in \mathbb{N}_0$  with  $c \leq k$ :

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

*Moreover*  $|\text{ScatFact}_{k-i}(w)| \ge k - c + 1$  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)

 $\bigcirc \Rightarrow w = w_1 a b a w_2$ 

 $\bigcirc$  induction on word-length



# Max. Card. for Weakly-O-Balanced Words of Length 2k

### Theorem

 $w \in \Sigma^*$ :

# $\operatorname{ScatFact}_k(w) = \Sigma^k \quad iff \quad \{\operatorname{ab}, \operatorname{ba}\}^k \cap \operatorname{ScatFact}_{2k}(w) \neq \emptyset$



# Max. Card. for Weakly-O-Balanced Words of Length 2k

# Theorem

 $w \in \Sigma^*$ :

 $\operatorname{ScatFact}_k(w) = \Sigma^k \quad iff \quad \{\operatorname{ab}, \operatorname{ba}\}^k \cap \operatorname{ScatFact}_{2k}(w) \neq \emptyset$ 

**Conclusion:** for  $w \in \Sigma^{2k}$  weakly-0-balanced

 $\operatorname{ScatFact}_k(w) = \Sigma^k \quad \text{iff} \quad w \in \{\operatorname{ab}, \operatorname{ba}\}^k,$ 



# Max. Card. for Weakly-O-Balanced Words of Length 2k

### Theorem

 $w\in \Sigma^*:$ 

$$\operatorname{ScatFact}_k(w) = \Sigma^k \quad iff \quad \{\operatorname{ab}, \operatorname{ba}\}^k \cap \operatorname{ScatFact}_{2k}(w) \neq \emptyset$$

**Conclusion:** for  $w \in \Sigma^{2k}$  weakly-0-balanced

ScatFact<sub>k</sub>(
$$w$$
) =  $\Sigma^k$  iff  $w \in \{ab, ba\}^k$ ,

i.e.  $w \in {ab, ba}^k$  iff  $|ScatFact_k(w)| = 2^k$ 



unfortunately the result for the maximal cardinality is not generalisable in the same way as for the minimal one

k	ababa	abbaa	
1	Σ	Σ	prev. result
2	$\Sigma^2$	$\Sigma^2$	prev. result
3	$\Sigma^3 \setminus \{b^3\}$	$\{a^3, aba, ab^2, ba^2, b^2a\}$	



unfortunately the result for the maximal cardinality is not generalisable in the same way as for the minimal one

k	ababa	abbaa	
1	Σ	Σ	prev. result
2	$\Sigma^2$	$\Sigma^2$	prev. result
3	$\Sigma^3 \setminus \{b^3\}$	$\{a^3, aba, ab^2, ba^2, b^2a\}$	

for c > 0, the switches from a to b and v.v. matter!



For all 
$$i \le k - c$$
,  $c \in [k]_0$ ,  $k \in \mathbb{N}$   
 $|\operatorname{ScatFact}_i((\operatorname{ab})^{k-c} \operatorname{a}^c)| =$   
 $1 + 2^{k-c} + \sum_{j \in [(i+c)-k-1]_0} |\operatorname{ScatFact}_{i-j-1}((\operatorname{ab})^{k-c-1} \operatorname{a})|$ 

with  $|\operatorname{ScatFact}_{\ell}(\operatorname{Pref}_{n}(\operatorname{ab})^{\omega})| = \sum_{j \in [n-\ell]_{0}} {\ell \choose n-\ell-j}$ 



For all 
$$i \le k - c$$
,  $c \in [k]_0$ ,  $k \in \mathbb{N}$   
 $|\text{ScatFact}_i((ab)^{k-c}a^c)| =$   
 $1 + 2^{k-c} + \sum_{j \in [(i+c)-k-1]_0} |\text{ScatFact}_{i-j-1}((ab)^{k-c-1}a)|$ 

with  $|\operatorname{ScatFact}_{\ell}(\operatorname{Pref}_{n}(\operatorname{ab})^{\omega})| = \sum_{j \in [n-\ell]_{0}} {\ell \choose n-\ell-j}$ 

## Proof-Idea.

- $\bigcirc$  (*i*<sub>1</sub>,...,*i*<sub>m</sub>) deleting sequence  $\rightsquigarrow$  scattered factor
- $\odot$  several deleting sequences lead to the same scattered factor
- $\bigcirc$  count only one of these sequences

# k-spectra for weakly-Obalanced words of length 2k

# Properties of weakly-O-balanced words

#### $\bigcirc$ same amount of as and bs



- $\bigcirc$  same amount of as and bs
- always even length



- $\bigcirc$  same amount of as and bs
- always even length
- $\bigcirc$  the *k*-spectra has at most 2<sup>*k*</sup> elements



# 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!





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



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



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

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

A LING

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

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

### 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







 $a^{k-1}b^k$  a generalisable to  $a^{k-i}b^ka^i$  for  $i \in \left\lfloor \lfloor \frac{k}{2} \rfloor \right]$ :

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



# Spectrum of *k*-spectra



 $a^{k-1}b^ka$  generalisable to  $a^{k-i}b^ka^i$  for  $i \in \left\lfloor \lfloor \frac{k}{2} \rfloor \right\rfloor$ :

$$|$$
ScatFact<sub>k</sub> $(a^{k-i}b^ka^i)| = k(i+1) - i^2 + 1$ 



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



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





but (unfortunately)



but (unfortunately)

#### Lemma

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



but (unfortunately)

### Lemma

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

and this result is generalisable



### Lemma

For 
$$k \ge 5$$
 and  $i \in [k-1]$   
 $\bigcirc$  |ScatFact<sub>k</sub>( $a^{k-2}b^{i}ab^{k-i}a$ )| =  $k(2i+2) - 6i+2$   
 $\bigcirc$  |ScatFact<sub>k</sub>( $a^{k-2}b^{i}a^{2}b^{k-i}$ )| =  $k(2i+1) - 4i+2$ 

ALL ONLY STORE

$$k \ge 38$$

$$3k - 3 - 3 - 4k - 8 - 5k - 15 - 6k - 24 - 7k - 35 - 8k - 48 - 10 - 7k - 10 - 8k - 16$$



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



# 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  $i \in [k-2]$ .



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:

 $\bigcirc$  "="  $\checkmark$ 

○ "⇒" 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 weakly-O-balanced words

k-Spectra



