

Arithmetics on Suffix Arrays of Fibonacci Words

Tomohiro I *Dominik Köppl*

Department of Computer Science, TU Dortmund, Germany

WORDS 2015

Fibonacci Words

$$F_n = \begin{cases} b & \text{if } n = 1, \\ a & \text{if } n = 2, \\ F_{n-1}F_{n-2} & \text{otherwise.} \end{cases}$$

- ▀ $F_3 = ab$
- ▀ $F_4 = aba$
- ▀ $F_5 = abaab$
- ▀ $F_6 = abaababa$

- ▀ $|F_n| = f_n$

Fibonacci Numbers

$$f_n = \begin{cases} 1 & \text{if } n = 1, \\ 1 & \text{if } n = 2, \\ f_{n-1} + f_{n-2} & \text{otherwise.} \end{cases}$$

- ▀ $f_3 = 2$
- ▀ $f_4 = 3$
- ▀ $f_5 = 5$
- ▀ $f_6 = 8$

suffix arrays

SA_5 : suffix array of F_5 :

- take suffixes of F_5
- sort them
- suffix array SA_5 is first column

$F_5 = \text{abaab}$

1	a	b	a	a	b
2	b	a	a	b	
3	a	a	b		
4	a	b			
5	b				

suffix arrays

SA_5 : suffix array of F_5 :

- take suffixes of F_5
- sort them
- suffix array SA_5 is first column

$F_5 = \text{abaab}$

3	a	a	b		
4	a	b			
1	a	b	a	a	b
5	b				
2	b	a	a	b	

suffix arrays

SA_5 : suffix array of F_5 :

- take suffixes of F_5
- sort them
- suffix array SA_5 is first column

$F_5 = \text{abaab}$

$SA_5 = 3\ 4\ 1\ 5\ 2$

SA_5					
3	a	a	b		
4	a	b			
1	a	b	a	a	b
5	b				
2	b	a	a	b	

BWT

BWT₅: Burrows Wheeler Transformation of $F_5 = \text{abaab}$

- same table as before
- first column: char before suffix
- first column = BWT₅ = babaa

3	a	a	b		
4	a	b			
1	a	b	a	a	b
5	b				
2	b	a	a	b	

BWT

BWT₅: Burrows Wheeler Transformation of $F_5 = \text{abaab}$

- same table as before
- first column: char before suffix
- first column = BWT₅ = babaa

b	3	a	a	b		
a	4	a	b			
b	1	a	b	a	a	b
a	5	b				
a	2	b	a	a	b	

BWT

BWT₅: Burrows Wheeler Transformation of $F_5 = \text{abaab}$

- same table as before
- first column: char before suffix
- first column = BWT₅ = babaa

b	3	a	a	b		
a	4	a	b			
b	1	a	b	a	a	b
a	5	b				
a	2	b	a	a	b	

F_6

i	1	2	3	4	5	6	7	8
F_6	a	b	a	a	b	a	b	a
SA_6	8	3	6	1	4	7	2	5
BWT_6	b	b	b	a	a	a	a	a

SA_6 : suffix array, BWT_6 : Burrows-Wheeler Transformation

F_6

i	1	2	3	4	5	6	7	8
F_6	a	b	a	a	b	a	b	a
SA_6	8	3	6	1	4	7	2	5
BWT_6	b	b	b	a	a	a	a	a

SA_6 : suffix array, BWT_6 : Burrows-Wheeler Transformation

Observations

■ $BWT_n = b^{f_4} a^{f_5}$ ($f_4 = 3, f_5 = 5$)

F_6

i	1	2	3	4	5	6	7	8
F_6	a	b	a	a	b	a	b	a
SA_6	8	3	6	1	4	7	2	5
BWT_6	b	b	b	a	a	a	a	a

SA_6 : suffix array, BWT_6 : Burrows-Wheeler Transformation

Observations

- $BWT_n = b^{f_4} a^{f_5}$ ($f_4 = 3, f_5 = 5$)
- $SA_n[i] = SA_n[i - 1] + f_4 \pmod{f_6}$

F_6

i	1	2	3	4	5	6	7	8
F_6	a	b	a	a	b	a	b	a
SA_6	8	3	6	1	4	7	2	5
BWT_6	b	b	b	a	a	a	a	a

SA_6 : suffix array, BWT_6 : Burrows-Wheeler Transformation

Observations

- ▀ $BWT_n = b^{f_4} a^{f_5}$ ($f_4 = 3, f_5 = 5$)
- ▀ $SA_n[i] = SA_n[i - 1] + f_4 \pmod{f_6}$

Definition

SA_n is called **arithmetically progressed** if $\exists k \in \mathbb{N}$:

$$SA_n[i] = SA_n[i - 1] + k \pmod{|SA_n|} \quad \forall i$$

works for *all* fibos?

Conjecture

SA_n is arithmetically progressed for every n .

works for *all* fibos?

Conjecture

SA_n is arithmetically progressed for every n .

Does not hold!

i	1	2	3	4	5
F_5	a	b	a	a	b
SA_5	3	4	1	5	2

our results

- ▀ SA_n is arithmetically progressed for *even* n
- ▀ $BWT_n = b^{f_{n-2}} a^{f_{n-1}}$ for *even* n

our results

- SA_n is arithmetically progressed for *even* n
- $BWT_n = b^{f_{n-2}} a^{f_{n-1}}$ for *even* n

Wait a sec! What about

Theorem (Mantaci et al'03)

$BWT_n = b^{f_{n-2}} a^{f_{n-1}}$ for *all* n .

our results

- SA_n is arithmetically progressed for *even* n
- $BWT_n = b^{f_{n-2}} a^{f_{n-1}}$ for *even* n

Wait a sec! What about

Theorem (Mantaci et al'03)

$BWT_n = b^{f_{n-2}} a^{f_{n-1}}$ for *all* n .

We look at suffix based BWT, not rotation based BWT!

BWT_{F₅}\$

F₅\$ = abaab\$

by rotation

a	b	a	a	b	\$
\$	a	b	a	a	b
b	\$	a	b	a	a
a	b	\$	a	b	a
a	a	b	\$	a	b
b	a	a	b	\$	a

F *L*

by suffixes

\$	a	b	a	a	b	\$
a	b	a	a	b	\$	
b	a	a	b	\$		
a	a	b	\$			
a	b	\$				
b	\$					

-1

BWT_{F₅\$}

F₅\$ = abaab\$

by rotation

\$	a	b	a	a	b
a	a	b	\$	a	b
a	b	\$	a	b	a
a	b	a	a	b	\$
b	\$	a	b	a	a
b	a	a	b	\$	a
<hr/>			<hr/>		
F			L		

by suffixes

\$	a	b	a	a	b	\$
a	b	a	a	b	\$	
b	a	a	b	\$		
a	a	b	\$			
a	b	\$				
b	\$					
<hr/>						
-1						

BWT_{abaab\$} = bba\$aa

BWT_{F₅\$}

F₅\$ = abaab\$

by rotation

\$	a	b	a	a	b
a	a	b	\$	a	b
a	b	\$	a	b	a
a	b	a	a	b	\$
b	\$	a	b	a	a
b	a	a	b	\$	a

F L

by suffixes

\$	a	b	a	a	b	\$
a	b	a	a	b	\$	
b	a	a	b	\$		
a	a	b	\$			
a	b	\$				
b	\$					

-1

BWT_{abaab\$} = bba\$aa

BWT_{F₅}\$

F₅\$ = abaab\$

by rotation

\$	a	b	a	a	b
a	a	b	\$	a	b
a	b	\$	a	b	a
a	b	a	a	b	\$
b	\$	a	b	a	a
b	a	a	b	\$	a

F *L*

BWT_{abaab\$} = bba\$aa

by suffixes

b	\$						
b	a	a	b	\$			
a	a	b	\$				
\$	a	b	a	a	b	\$	
a	b	\$					
a	b	a	a	b	\$		

-1

BWT_{abaab\$} = bba\$aa

BWT_{F₅}

$F_5 = abaab$

by rotation

a	b	a	a	b
b	a	b	a	a
a	b	a	b	a
a	a	b	a	b
b	a	a	b	a

F *L*

by suffixes

b	a	b	a	a	b
a	b	a	a	b	
b	a	a	b		
a	a	b			
a	b				

-1

BWT_{F₅}

$F_5 = abaab$

by rotation

a	a	b	a	b
a	b	a	a	b
a	b	a	b	a
b	a	a	b	a
b	a	b	a	a

F *L*

by suffixes

b	a	b	a	a	b
a	b	a	a	b	
b	a	a	b		
a	a	b			
a	b				

-1

BWT_{abaab} = bbaaa

BWT_{F₅}

F₅ = abaab

by rotation

a	a	b	a	b
a	b	a	a	b
a	b	a	b	a
b	a	a	b	a
b	a	b	a	a
<hr/>				
F				L

BWT_{abaab} = bbaaa

by suffixes

b	a	a	b		
a	a	b			
b	a	b	a	a	b
a	b				
a	b	a	a	b	
<hr/>					
-1					

BWT_{abaab} = babaa

Without \$ both variants *differ!*

let's start. . .

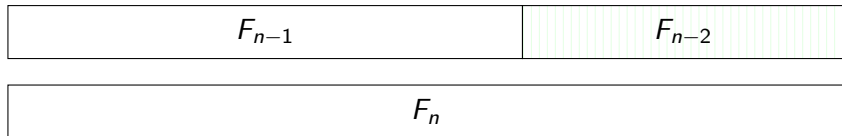
splits for n even



F_n

splits for n even

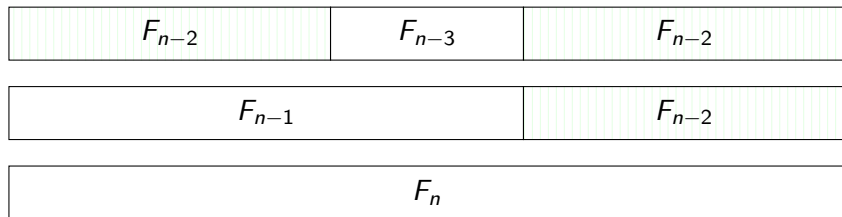
▀ $F_n = F_{n-1}F_{n-2}$



splits for n even

▀ $F_n = F_{n-1}F_{n-2}$

▀ $F_{n-1} = F_{n-2}F_{n-3}$



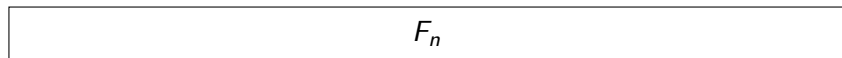
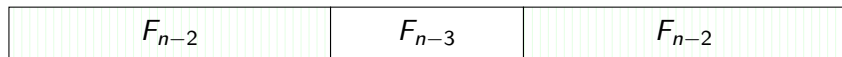
splits for n even

▀ $F_n = F_{n-1}F_{n-2}$

▀ $F_{n-1} = F_{n-2}F_{n-3}$

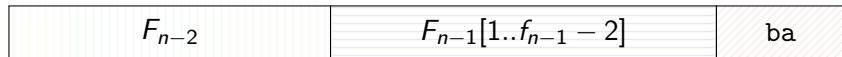
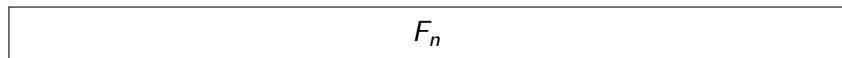
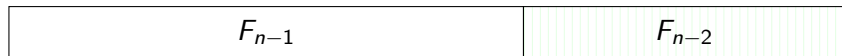
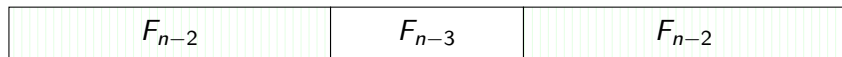
▀ $F_n = \dots ba$ and $F_{n-1} = \dots ab$

[induction]



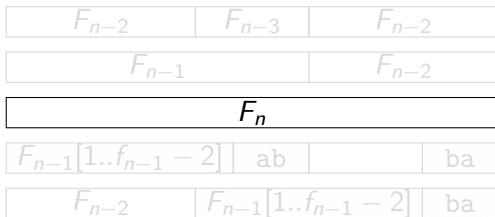
splits for n even

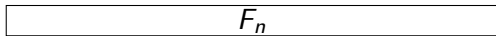
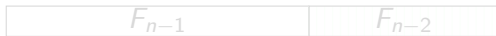
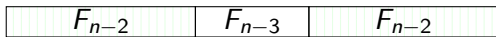
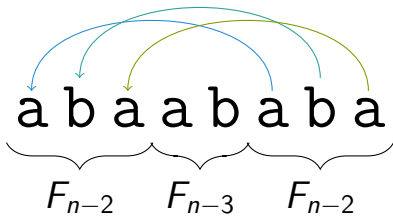
- $F_n = F_{n-1}F_{n-2}$
- $F_{n-1} = F_{n-2}F_{n-3}$
- $F_n = \dots ba$ and $F_{n-1} = \dots ab$ [induction]
- $F_n = F_{n-2}F_{n-1}[1..f_{n-1} - 2]ba$ [Christodoulakis et al'03]

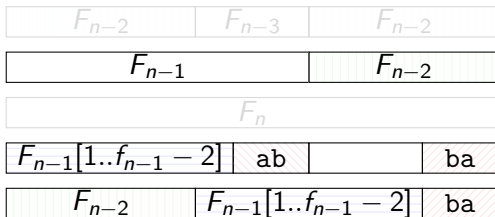
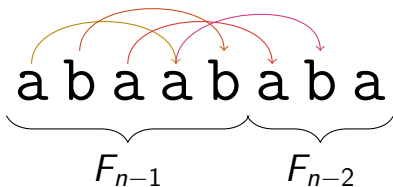


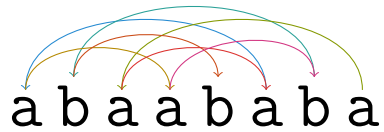
SA_n is arithmetically progressed

a b a a b a b a



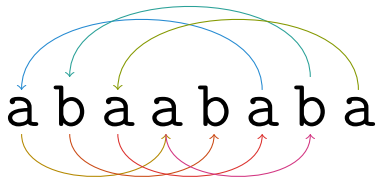






moves:


$$SA_6 = 8$$



moves: -5

$SA_6 = 8\ 3$


a b a a b a b a



moves: $-5 +3$

$SA_6 = 8\ 3\ 6$


a b a a b a b a



moves: -5 $+3$ -5

$SA_6 = 8\ 3\ 6\ 1$

a b a a b a b a



moves: $-5 +3 -5 +3$

$SA_6 = 8 3 6 1 4$

a b a a b a b a

A diagram showing the string "a b a a b a b a" with a yellow curved arrow pointing from the first 'a' to the second 'a'.

moves: $-5 +3 -5 +3 +3$

$SA_6 = 8 3 6 1 4 7$

a b a a b a b a



moves: -5 +3 -5 +3 +3 -5

$SA_6 = 8\ 3\ 6\ 1\ 4\ 7\ 2$


a b a a b a b a



moves: $-5 +3 -5 +3 +3 -5 +3$

$SA_6 = 8 3 6 1 4 7 2 5$

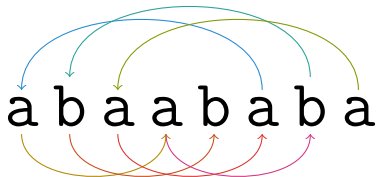
a b a a b a b a



moves: $-5 +3 -5 +3 +3 -5 +3$

$SA_6 = 8 3 6 1 4 7 2 5$

We move always $3 \pmod{f_6}$ right!



main result

for n even:

Theorem

$$SA_n[i] = \begin{cases} f_n & \text{if } i = 1, \\ (SA_n[i - 1] + f_{n-2}) \bmod f_n & \text{otherwise.} \end{cases}$$

$$\text{BWT}_n = \mathbf{b}^{f_{n-2}} \mathbf{a}^{f_{n-1}}$$

2-letter substrings of F_n

Running example F_6 :

i	1	2	3	4	5	6	7	8
F_6	a	b	a	a	b	a	b	a
SA_6	8	3	6	1	4	7	2	5
$F_6[SA_n[i]]$	a	a	a	a	a	b	b	b
BWT_6	b	b	b	a	a	a	a	a

2-letter substrings of F_n

Running example F_6 :

i	1	2	3	4	5	6	7	8
F_6	a	b	a	a	b	a	b	a
SA_6	8	3	6	1	4	7	2	5
$F_6[SA_n[i]]$	a	a	a	a	a	b	b	b
BWT_6	b	b	b	a	a	a	a	a

Observation

$F_n[SA_n[i] - 1]$	b ... b	a ... a	a ... a
$F_n[SA_n[i]]$	a ... a	a ... a	b ... b
Blocks	ba-type	aa-type	ab-type

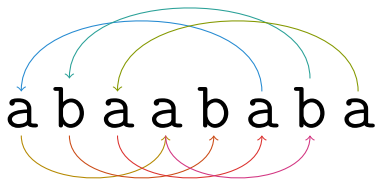
BWT₆

▀ $\text{BWT}_n[i] = F_n[\text{SA}_n[i] - 1]$

a b a a b a b a

BWT₆

- ▀ $BWT_n[i] = F_n[SA_n[i] - 1]$
- ▀ follow SA_n -arrows



$BWT_6 = b$

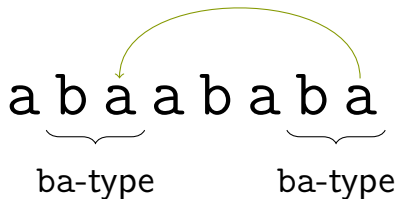
- ▀ $BWT_n[i] = F_n[SA_n[i] - 1]$
- ▀ follow SA_n -arrows
- ▀ start at $SA_n[1] = f_n$

a b a a b a b a



$BWT_6 = b b$

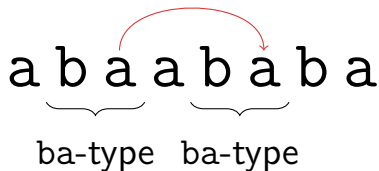
- ▶ $BWT_n[i] = F_n[SA_n[i] - 1]$
- ▶ follow SA_n -arrows
- ▶ start at $SA_n[1] = f_n$



$BWT_6 = b b b$

- ▀ $BWT_n[i] = F_n[SA_n[i] - 1]$
- ▀ follow SA_n -arrows
- ▀ start at $SA_n[1] = f_n$

a b a a b a b a
ba-type ba-type



$\text{BWT}_6 = \text{b b b a}$

- $\text{BWT}_n[i] = F_n[\text{SA}_n[i] - 1]$
- follow SA_n -arrows
- start at $\text{SA}_n[1] = f_n$

a b a a b a b a

↑

$F_n[f_{n-1} + 1]$

$\text{BWT}_6 = \text{b b b a} \cdots \text{a}$

- ▀ $\text{BWT}_n[i] = F_n[\text{SA}_n[i] - 1]$
- ▀ follow SA_n -arrows
- ▀ start at $\text{SA}_n[1] = f_n$

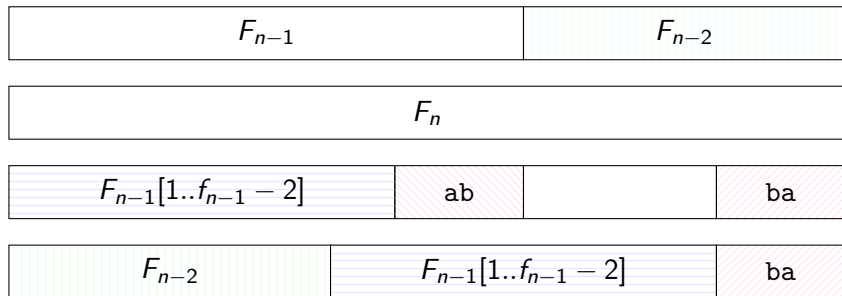
a b a a b a b a
 ↑
 $F_n[f_{n-1} + 1]$

Proposition

The run of ba-types ends at suffix $F_n[f_{n-1} + 1..]$

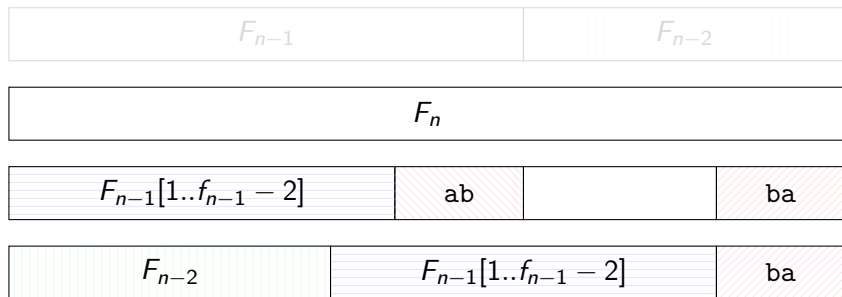
($f_5 = 5$)

splits



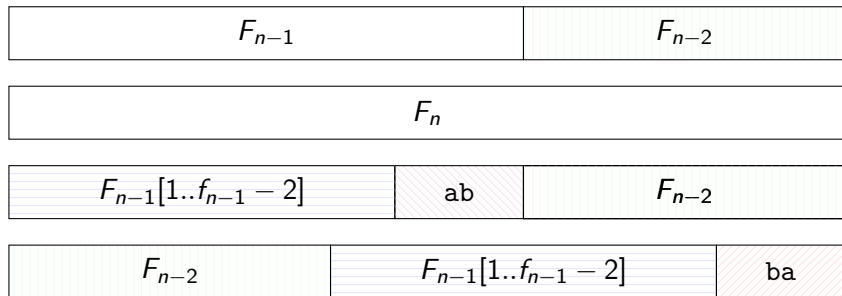
- $F_n[i] = F_n[i + f_{n-2}] \forall i = 1, \dots, f_{n-1} - 2$
- $F_n[i] = F_n[i + f_{n-1}] \forall i = 1, \dots, f_{n-2}$
- Move $f_{n-2} \bmod f_n$ chars right \Rightarrow find same 2-letter substring!
(not true at $F_n[f_{n-1} - 1..f_{n-1}]$.)

splits



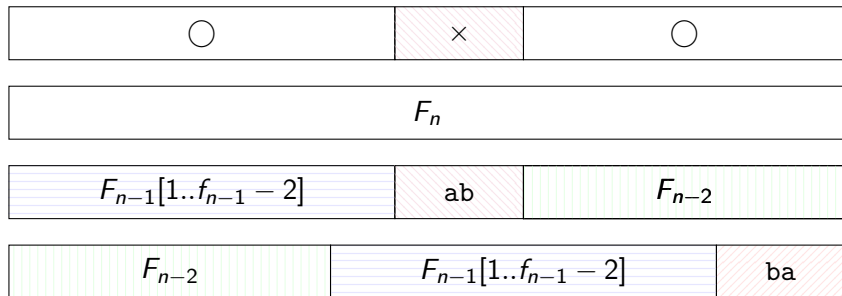
- ▀ $F_n[i] = F_n[i + f_{n-2}] \forall i = 1, \dots, f_{n-1} - 2$
- ▀ $F_n[i] = F_n[i + f_{n-1}] \forall i = 1, \dots, f_{n-2}$
- ▀ Move $f_{n-2} \bmod f_n$ chars right \Rightarrow find same 2-letter substring!
(not true at $F_n[f_{n-1} - 1..f_{n-1}]$.)

splits



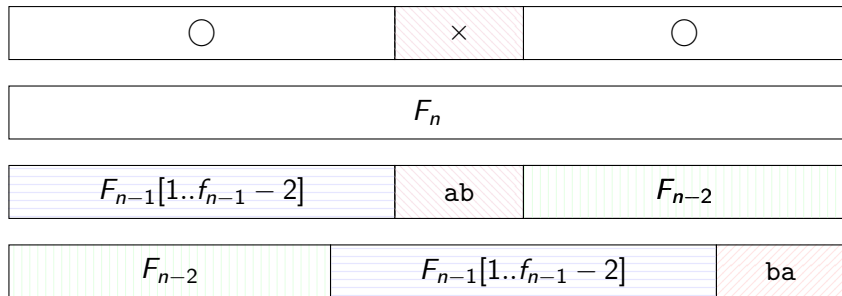
- ▀ $F_n[i] = F_n[i + f_{n-2}] \forall i = 1, \dots, f_{n-1} - 2$
- ▀ $F_n[i] = F_n[i + f_{n-1}] \forall i = 1, \dots, f_{n-2}$
- ▀ Move $f_{n-2} \bmod f_n$ chars right \Rightarrow find same 2-letter substring!
(not true at $F_n[f_{n-1} - 1..f_{n-1}]$.)

splits



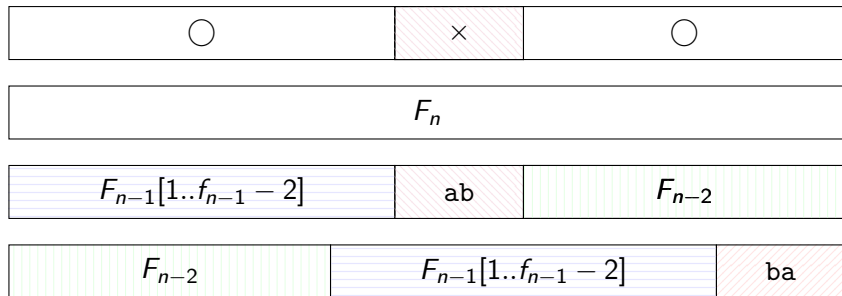
- ▶ $F_n[i] = F_n[i + f_{n-2}] \forall i = 1, \dots, f_{n-1} - 2$
- ▶ $F_n[i] = F_n[i + f_{n-1}] \forall i = 1, \dots, f_{n-2}$
- ▶ Move $f_{n-2} \bmod f_n$ chars right \Rightarrow find same 2-letter substring!
(not true at $F_n[f_{n-1} - 1..f_{n-1}]$.)

hunting ba-types



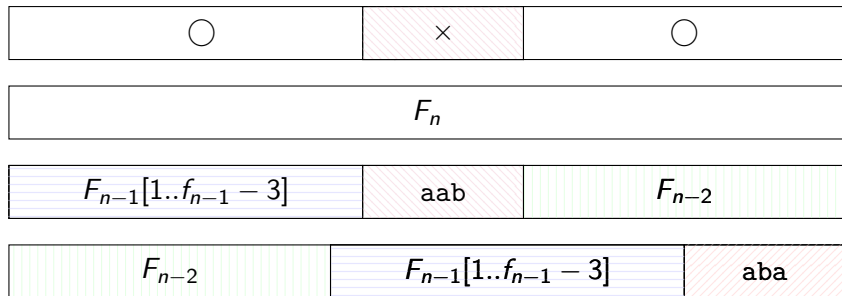
- ▀ $SA_n[1] = f_n$ and $F_n[f_n - 1..] = ba$
- ▀ find ba-types until reaching ×
- ▀ SA_n arithmetically progressed, step f_{n-2} [proved]
- ▀ $SA_n[f_{n-2}] = f_{n-1} + 1$ [calculate]

hunting ba-types



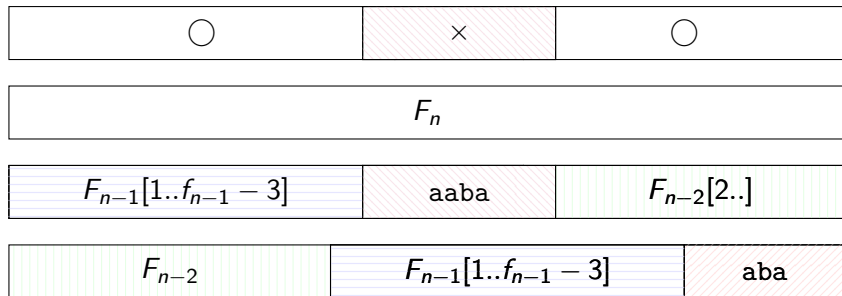
- $SA_n[1] = f_n$ and $F_n[f_n - 1..] = ba$
- find ba-types until reaching ×
- SA_n arithmetically progressed, step f_{n-2} [proved]
- $SA_n[f_{n-2}] = f_{n-1} + 1$ [calculate]

hunting ba-types



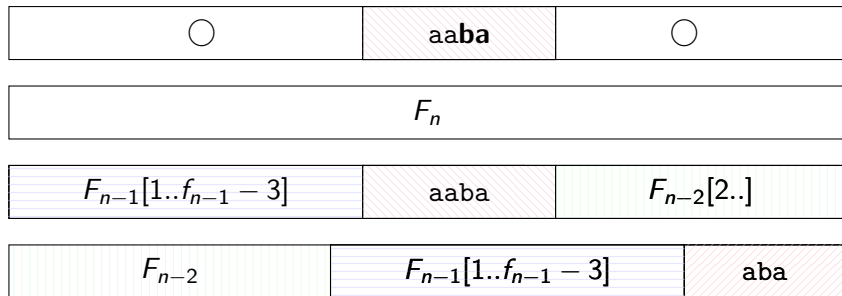
- $SA_n[1] = f_n$ and $F_n[f_n - 1..] = ba$
- find ba-types until reaching ×
- SA_n arithmetically progressed, step f_{n-2} [proved]
- $SA_n[f_{n-2}] = f_{n-1} + 1$ [calculate]

hunting ba-types



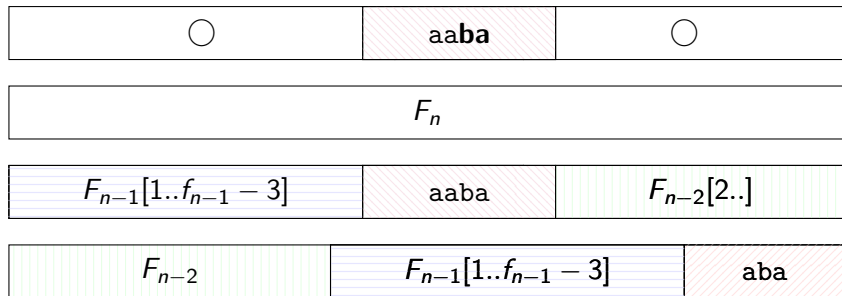
- $SA_n[1] = f_n$ and $F_n[f_n - 1..] = ba$
- find ba-types until reaching ×
- SA_n arithmetically progressed, step f_{n-2} [proved]
- $SA_n[f_{n-2}] = f_{n-1} + 1$ [calculate]

hunting ba-types



- $SA_n[1] = f_n$ and $F_n[f_n - 1..] = ba$
- find ba-types until reaching \times
- SA_n arithmetically progressed, step f_{n-2} [proved]
- $SA_n[f_{n-2}] = f_{n-1} + 1$ [calculate]

hunting ba-types



- $SA_n[1] = f_n$ and $F_n[f_n - 1..] = ba$
- find ba-types until reaching \times
- SA_n arithmetically progressed, step f_{n-2} [proved]
- $SA_n[f_{n-2}] = f_{n-1} + 1$ [calculate]

Theorem

$\text{BWT}_n = b^{f_{n-2}} a^{f_{n-1}}$ for even n .

- ▀ $\#b = f_{n-2}$
- ▀ $\text{SA}_n[f_{n-2}] = f_{n-1} + 1$
- ▀ Found all ba-substrings!

[folklore]

[previous slide]



summary

results

- SA_n arithmetically progressed for *even* n
- BWT_n based on suffixes = $b^{f_{n-2}}a^{f_{n-1}}$ for *even* n

further results (read the paper!)

- $SA_n[i] = ISA_{F_n}[i + f_{n-2} + 1 \pmod{f_n}]$ for *even* n
- any LZ77 factor of F_n has same properties
- pre-/appending chars \Rightarrow similar characteristics

open problems

- class of strings that has the same properties?
- what about odd n ?

Thank you for listening. Any questions are welcome!

summary

results

- SA_n arithmetically progressed for *even* n
- BWT_n based on suffixes = $b^{f_{n-2}}a^{f_{n-1}}$ for *even* n

further results (read the paper!)

- $SA_n[i] = ISA_{F_n}[i + f_{n-2} + 1 \pmod{f_n}]$ for *even* n
- any LZ77 factor of F_n has same properties
- pre-/appending chars \Rightarrow similar characteristics

open problems

- class of strings that has the same properties?
- what about odd n ?

Thank you for listening. Any questions are welcome!

F_6

i	1	2	3	4	5	6	7	8
F_6	a	b	a	a	b	a	b	a
SA_6	8	3	6	1	4	7	2	5
ISA_{F_6}	4	7	2	5	8	3	6	1

SA_n : suffix array, ISA: inverse suffix array

F_6

i	1	2	3	4	5	6	7	8
F_6	a	b	a	a	b	a	b	a
SA_6	8	3	6	1	4	7	2	5
ISA_{F_6}	4	7	2	5	8	3	6	1

SA_n : suffix array, ISA: inverse suffix array

Observation

▀ $SA_6[i + f_4 \bmod f_6] = ISA_{F_6}[i]$ ($f_4 = 3$)

F_6

i	1	2	3	4	5	6	7	8
F_6	a	b	a	a	b	a	b	a
SA_6	8	3	6	1	4	7	2	5
ISA_{F_6}	4	7	2	5	8	3	6	1

SA_n : suffix array, ISA: inverse suffix array

Observation

$$\blacksquare SA_6[i + f_4 \bmod f_6] = ISA_{F_6}[i] \quad (f_4 = 3)$$

Definition

ISA is called a rotation of $SA_n \Leftrightarrow \exists k \in \mathbb{N}$ (called shift):

$$SA_n[i] = ISA[(k + i) \bmod |SA_n|]$$

rotation theorem

for n even:

$$SA_n[i] = \begin{cases} f_n & \text{if } i = 1, \\ (SA_n[i - 1] + f_{n-2}) \bmod f_n & \text{otherwise.} \end{cases}$$

rotation theorem

for n even:

$$SA_n[i] = \begin{cases} f_n & \text{if } i = 1, \\ (SA_n[i-1] + f_{n-2}) \bmod f_n & \text{otherwise.} \end{cases}$$

■ $f_{n-2}^2 \bmod f_n = 1$

[Folklore]

rotation theorem

for n even:

$$SA_n[i] = \begin{cases} f_n & \text{if } i = 1, \\ (SA_n[i-1] + f_{n-2}) \bmod f_n & \text{otherwise.} \end{cases}$$

■ $f_{n-2}^2 \bmod f_n = 1$ [Folklore]

ISA is a rotation of SA_n :

$$\begin{aligned} SA_n[i + f_{n-2} \bmod f_n] &= SA_n[i] + f_{n-2}^2 \bmod f_n \\ &= SA_n[i] + 1 \bmod f_n \end{aligned}$$

Hence:

$$ISA[SA_n[i] + 1 \bmod f_n] = i + f_{n-2} \bmod f_n$$

ISA is arithmetically progressed like SA_n , so its a rotation of SA_n .

rotation theorem

for n even:

$$SA_n[i] = \begin{cases} f_n & \text{if } i = 1, \\ (SA_n[i - 1] + f_{n-2}) \bmod f_n & \text{otherwise.} \end{cases}$$

■ $SA_n[1] = f_n$

rotation theorem

for n even:

$$SA_n[i] = \begin{cases} f_n & \text{if } i = 1, \\ (SA_n[i-1] + f_{n-2}) \bmod f_n & \text{otherwise.} \end{cases}$$

- ▀ $SA_n[1] = f_n$
- ▀ $SA_n[f_n] = SA_n[1] + (f_n - 1)f_{n-2} \bmod f_n$ [by recursion]

rotation theorem

for n even:

$$SA_n[i] = \begin{cases} f_n & \text{if } i = 1, \\ (SA_n[i-1] + f_{n-2}) \bmod f_n & \text{otherwise.} \end{cases}$$

- ▀ $SA_n[1] = f_n$
- ▀ $SA_n[f_n] = SA_n[1] + (f_n - 1)f_{n-2} \bmod f_n$ [by recursion]

shift is

$$\begin{aligned} SA_n[f_n] - ISA[f_n] \bmod f_n &\equiv SA_n[f_n] - ISA[SA_n[1]] \\ &\equiv SA_n[f_n] - 1 \\ &\equiv f_n + (f_n - 1)f_{n-2} - 1 \\ &\equiv -f_{n-2} - 1 \\ &\quad (\text{treat } f_n \hat{=} 0) \end{aligned}$$