Net occurrences in Fibonacci and Thue-Morse Words

Kaisei Kishi[1],

Peaker Guo[2,3]

1 Kyushu University

2 Institute of Science Tokyo

3 The University of Melbourne

Fibonacci words

Definition

The k-th Fibonacci words F_k over $\Sigma = \{a, b\}$ is defined as follows:

- $F_1 = b$, $F_2 = a$,
- $F_k = F_{k-1} \cdot F_{k-2}$ $(k \geq 3)$

Let $f_k = |F_k|$.   (k-th Fibonacci number)

e.g. $F_3 = a \ b$,

$F_4 = a \ b \ a$,

$F_5 = a \ b \ a \ a \ b$,

$F_6 = a \ b \ a \ a \ b \ a \ b \ a$

Fibonacci words

Definition

The k-th Fibonacci words F_k over $\Sigma = \{a, b\}$ is defined as follows:

- $F_1 = b, \quad F_2 = a,$
- $F_k = F_{k-1} \cdot F_{k-2} \; (k \geq 3)$

Let $f_k = |F_k|$. (k-th Fibonacci number)

e.g. $F_3 = a \; b,$

$F_4 = a \; b \; a,$

$F_5 = a \; b \; a \; a \; b,$

$F_6 = a \; b \; a \; a \; b \; a \; b \; a$

Thue-Morse words

Definition

The k-th Thue-Morse word TM_k over $\Sigma = \{a, b\}$ is defined as follows:

- $\mathrm{TM}_0 = a$,
- $\mathrm{TM}_k = \mathrm{TM}_{k-1} \cdot \overline{\mathrm{TM}_{k-1}}$ $(k \geq 1)$

Let $\tau_k = |\mathrm{TM}_k|$.

We get $\overline{\mathrm{TM}_j}$ by replacing all a and b in TM_j

e.g. $\mathrm{TM}_1 = a\ b$,

$\mathrm{TM}_2 = a\ b\ b\ a$,

$\mathrm{TM}_3 = a\ b\ b\ a\ b\ a\ a\ b$,

$\mathrm{TM}_4 = a\ b\ a\ a\ b\ a\ a\ b\ b\ a\ a\ b\ a\ b\ b\ a$

Thue-Morse words

Definition

The k-th Thue-Morse word TM_k over $\Sigma = \{a, b\}$ is defined as follows:

- $\mathrm{TM}_0 = a$,
- $\mathrm{TM}_k = \mathrm{TM}_{k-1} \cdot \overline{\mathrm{TM}_{k-1}}$ $(k \geq 1)$

Let $\tau_k = |\mathrm{TM}_k|$.

We get $\overline{\mathrm{TM}_j}$ by replacing all a and b in TM_j

e.g. $\mathrm{TM}_1 = a\ b$,

$\mathrm{TM}_2 = a\ b\ b\ a$, $\overline{\mathrm{TM}_2} = b\ a\ a\ b$,

$\mathrm{TM}_3 = a\ b\ b\ a\ b\ a\ a\ b$,

$\mathrm{TM}_4 = a\ b\ a\ a\ b\ a\ a\ b\ b\ a\ a\ b\ a\ b\ b\ a$

Net occurrence

Definition
In a string T, an occurrence of a substring $S = T[i \ldots j]$ is a net occurrence if S is repeated, while both left extension $T[i-1 \ldots j]$ and right extension $T[i \ldots j+1]$ are unique in T.

e.g. T = the_theoretical_theme

There are three occurrences of "the" in T.

Net occurrence

Definition

In a string T, an occurrence of a substring S = T[i … j] is a net occurrence if S is repeated, while both left extension T[i − 1 … j] and right extension T[i … j + 1] are unique in T.

e.g. T = the_theoretical_theme

the left extension _the occurs twice in T.

These two occurrences of "the" are not net occurrences in T.

Net occurrence

Definition

In a string T, an occurrence of a substring S = T[i ... j] is a net occurrence if S is repeated, while both left extension T[i − 1 ... j] and right extension T[i ... j + 1] are unique in T.

e.g. T = the_theoretical_theme

The only net occurrence of "the", since both left extension and right extension are unique.

Note. In this talk, we assume that occurrence T[0 ... j] and T[i ... |T| + 1] to be unique in T.

Net Frequency

Definition

In a string T, Net Frequency(NF) of a substring S = T[i ... j] is the number of net occurrence of in T.

e.g. T = the_theoretical_theme

There are three occurrence of "the" in T, but NF of "the" is one.

Net Frequency

Definition

In a string T, Net Frequency(NF) of a substring S = T[i ... j] is the number of net occurrence of in T.

e.g. T = the_theoretical_theme

There are three occurrence of “the” in T, but NF of “the” is one.

Only the is used as a definite article, while the other the are prefixes of other words, which are “theoretical” and “theme”.

Net Frequency

Definition
In a string T, Net Frequency(NF) of a substring S = T[i ... j] is the number of net occurrence of in T.

e.g. T = the_theoretical_theme

There are three occurrence of "the" in T, but NF of "the" is one.

Only the is used as a definite article, while the other the are prefixes of other words, which are "theoretical" and "theme".

Originally, NF is motivated by Chinese natural language processing tasks. [Lin and Yu, 2001]

Recent works

- There are some algorithms which compute the NF.

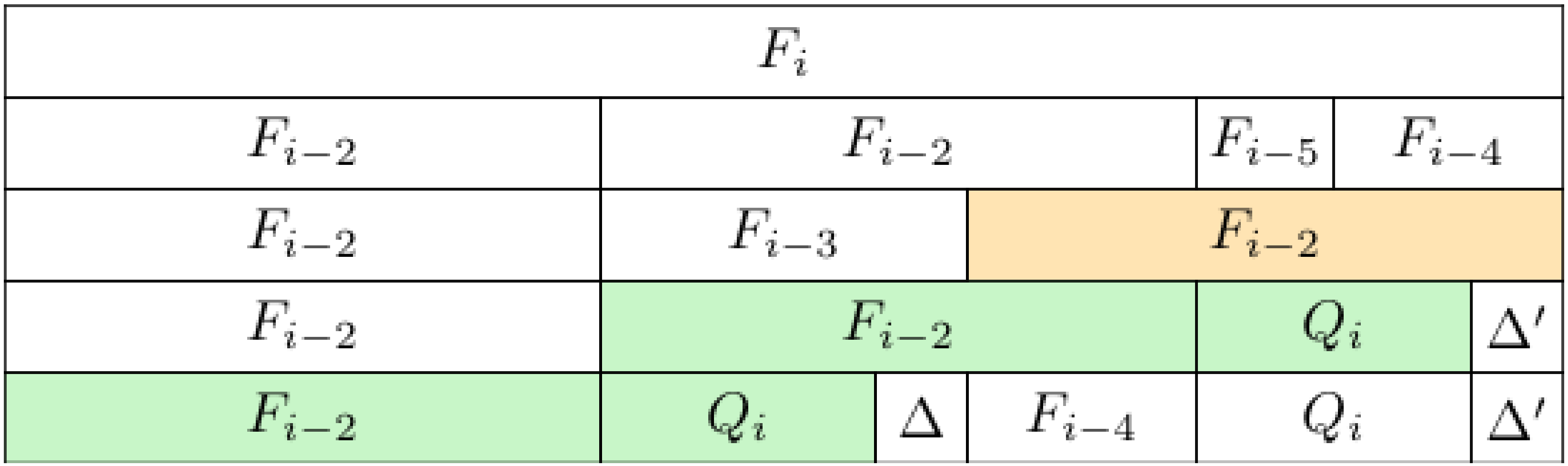
Offline settings: [Guo et al., CPM 2024], [Ohlebusch et al., SPIRE 2024]

Online settings: [Guo et al., SPIRE 2024] [Inenaga, ArXiv 2024]

[Mieno and Inenaga, CPM 2025]

- Also, Mieno and Inenaga (2025) characterized net occurrences in terms of the minimal unique substrings (MUSs).

Net occurrence in Fibonacci words



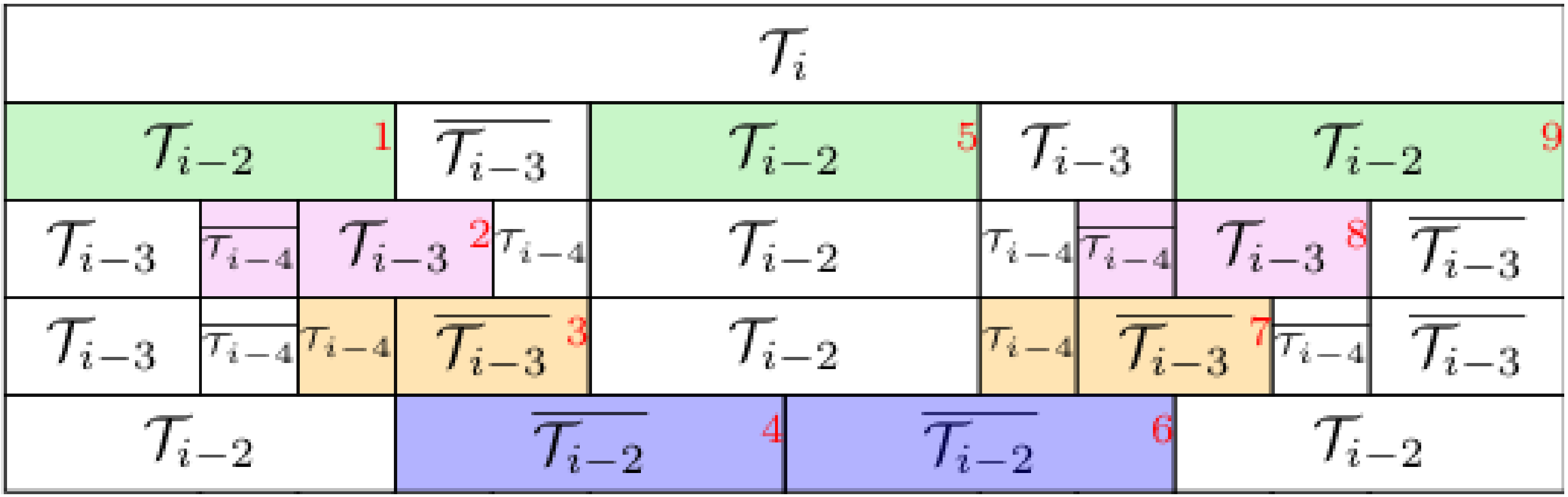
$Q_i := F_{i\text{-}3}[1 \,..\, |F_{i\text{-}3}| - 2]$

We proved:

(1) All three colored occurrences are Net occurrences.

(2) Other occurrences are not Net occurrences.

Net occurrence in Thue-Morse words



$\mathcal{T}_i := \mathrm{TM}_i$

We proved:

(1) All nine colored occurrences are Net occurrences.

(2) Other occurrences are not Net occurrences.

Basic properties of Net occurrences 1/2

Lemma 1
In a string T, if an occurrence of string S' is a proper **super-**occurrence of a net occurrence of string S in T, then occurrence of S' is not a net occurrence.

e.g. T = the_theoretical_theme

↑ net occurrence in T

An occurrence of "the_" is not a net occurrence since it must be unique in T.

Basic properties of Net occurrences 2/2

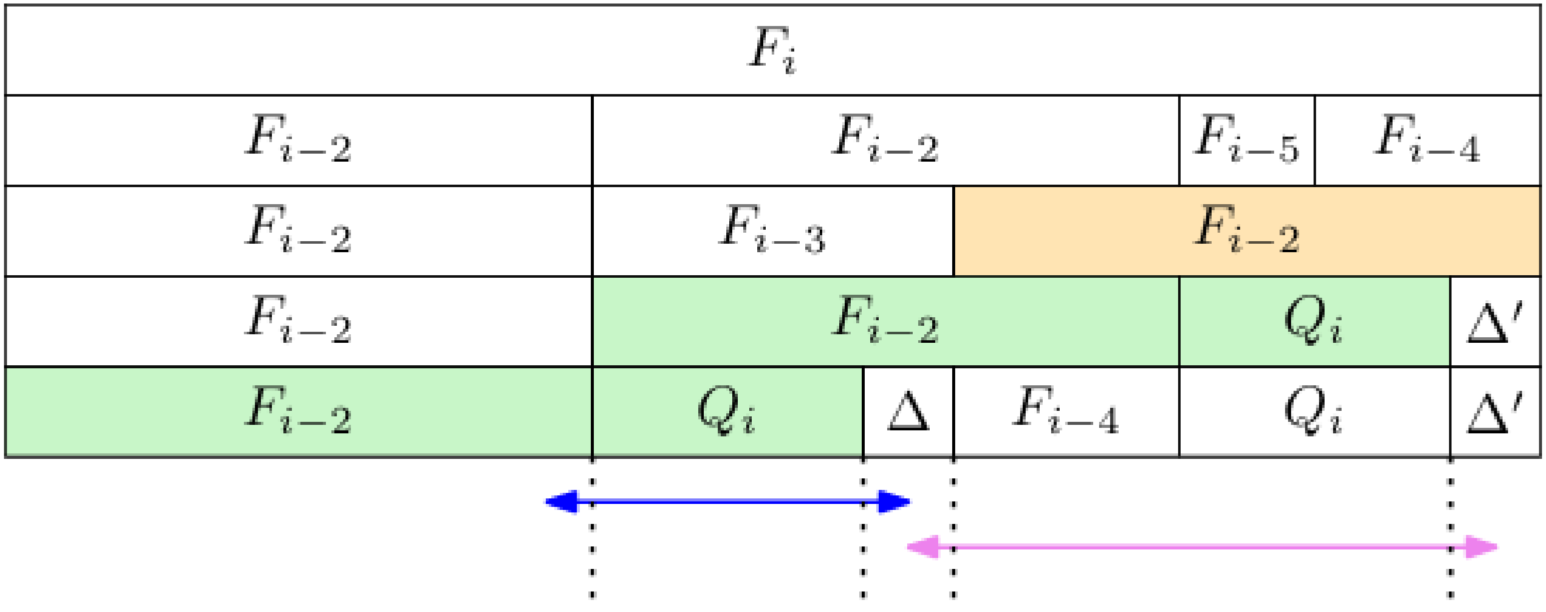
Lemma 2
In a string T, if an occurrence of string S’ is a proper **sub-**occurrence of a net occurrence of string S in T, then occurrence of S’ is not a net occurrence.

e.g. T = the_theoretical_theme

↑ net occurrence in T

An occurrence of “he” is not a net occurrence since it must be repeated in T, but left or right extension is also repeated in T.

Net occurrence in Fibonacci words (restate)



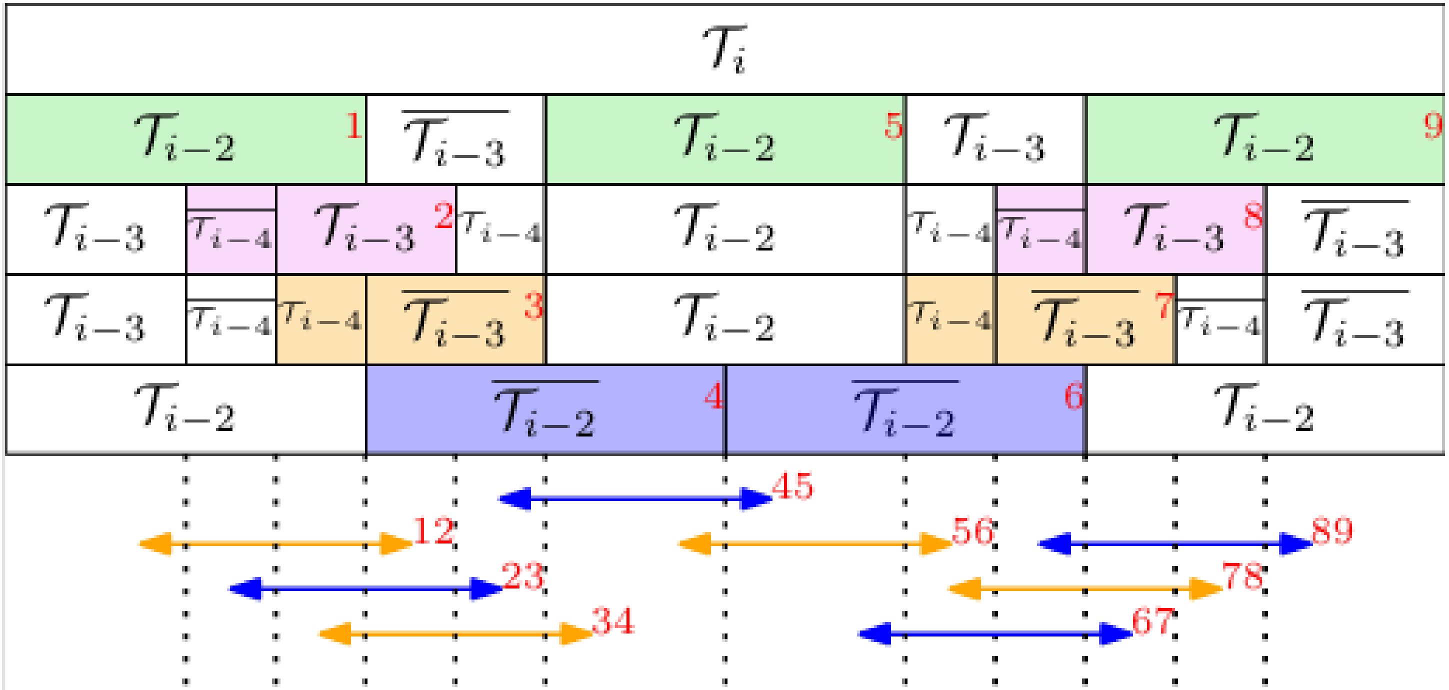
$Q_i := F_{i\text{-}3}[1 .. |F_{i\text{-}3}| - 2]$

We proved:

(1) All three colored occurrences are Net occurrences.

(2)' All of occurrences, which are shown in colored arrow are not Net occurrences.

Net occurrence in Thue-Morse words(restate)



$\mathcal{T}_i := \mathrm{TM}_i$

We proved:

(1) All nine colored occurrences are Net occurrences.

(2)' All of occurrences, which are shown in colored arrow are not Net occurrences.

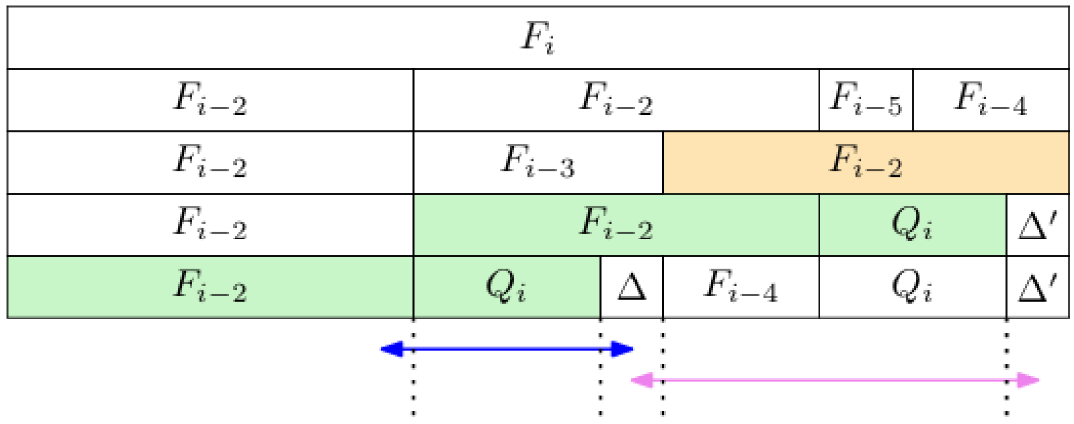
How to find Net occurrences?

To confirm an occurrence of substring T[i ... j] is a net occurrences in T, we have to prove

1. T[i ... j] is repeated in T.
2. T[i − 1 ... j] is unique in T.
3. T[i ... j + 1] is unique in T.

If we can count the number of occurrences of all of the substrings T[i ... j] in T, we can prove the above and find all of the Net occurrences in T.

Sketch of the proof of the Net occurrence in Fibonacci and Thue-Morse words



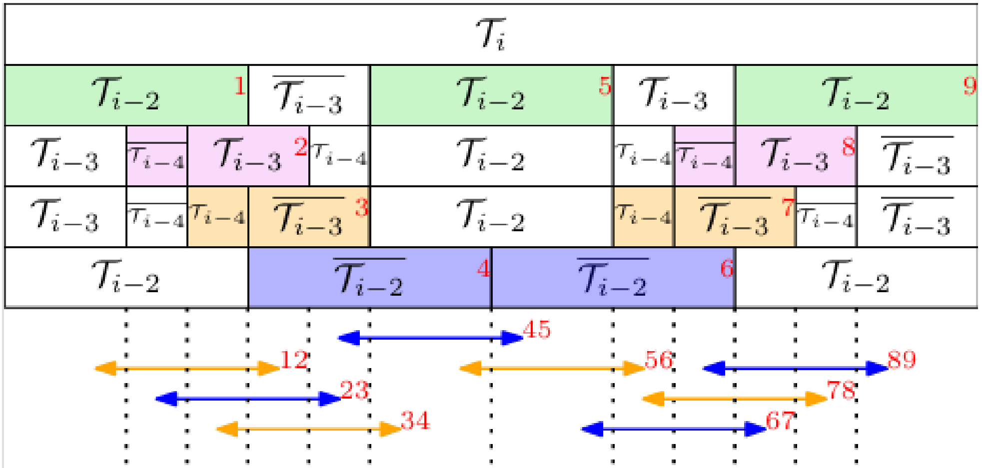
We proved:

(1) All colored occurrences are Net occurrences.

(2)' All of occurrences, which are shown in colored arrow are Net occurrences.

We proved the above by counting the number of occurrences of substrings in F_i and TM_i carefully.

Sketch of the proof of the Net occurrence in Fibonacci and Thue-Morse words



We proved:

(1) All colored occurrences are Net occurrences.

(2)' All of occurrences, which are shown in colored arrow are Net occurrences.

We proved the above by counting the number of occurrences of substrings in F_i and TM_i carefully.

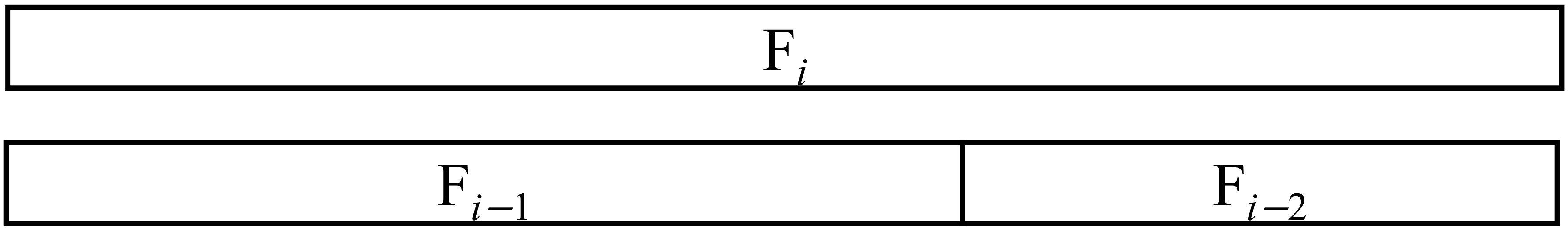
To count these substrings, we characterized the occurrences of in F_j and TM_j in F_i and TM_i where $j \leqq i$.

Factorization of F_i with smaller order

F_i

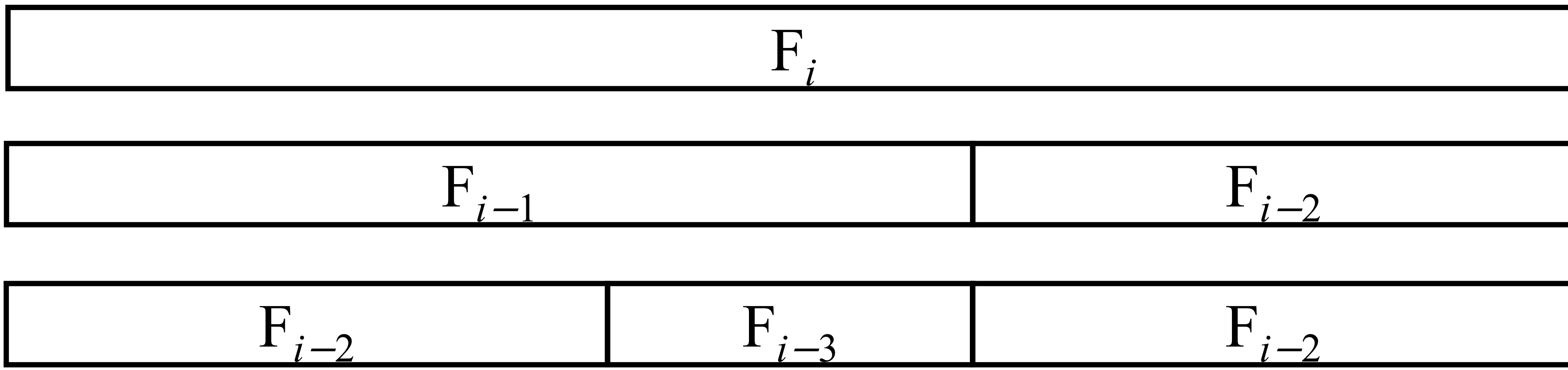
There is a factorization of F_i , where all factors are F_{j+1} or F_j when $j \leqq i$.

Factorization of F_i with smaller order



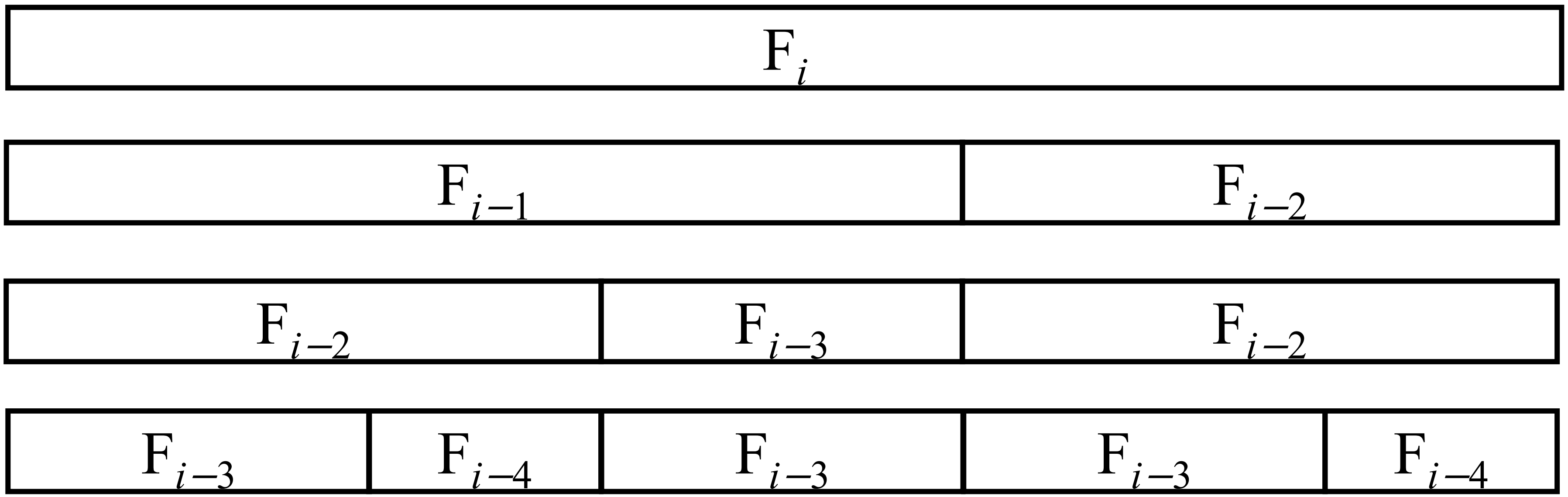
There is a factorization of F_i , where all factors are F_{j+1} or F_j when $j \leqq i$.

Factorization of F_i with smaller order



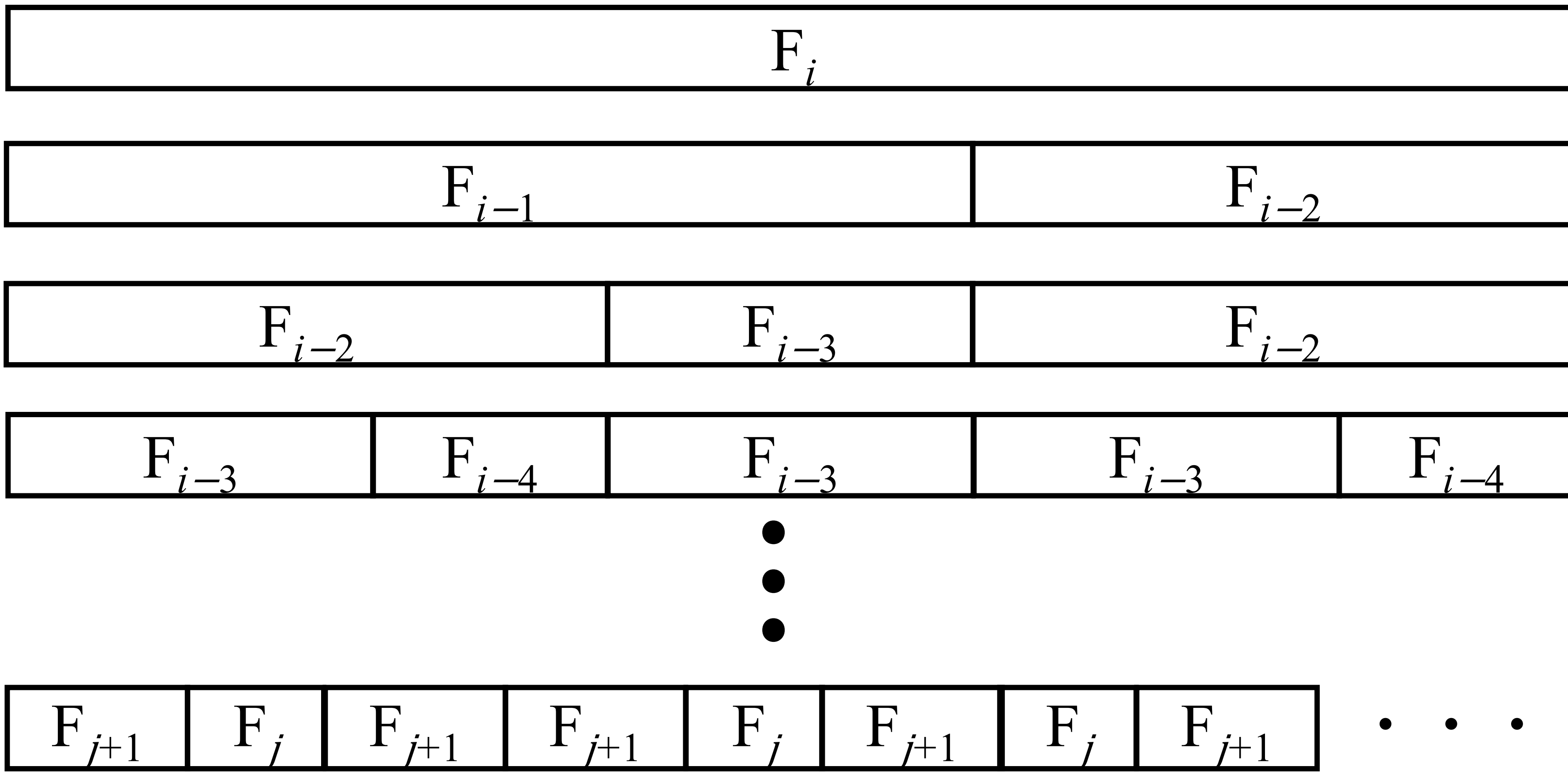
There is a factorization of F_i , where all factors are F_{j+1} or F_j when $j \leqq i$.

Factorization of F_i with smaller order



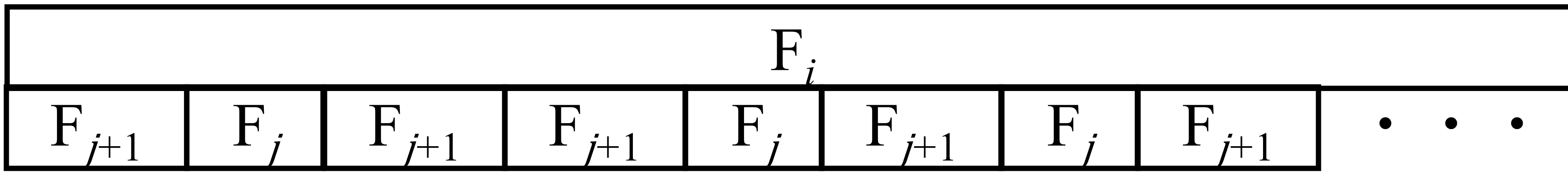
There is a factorization of F_i, where all factors are F_{j+1} or F_j when $j \leqq i$.

Factorization of F_i with smaller order

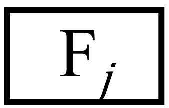


There is a factorization of F_i, where all factors are F_{j+1} or F_j when $j \leqq i$.

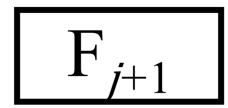
Occurrences of Fibonacci Words of smaller order



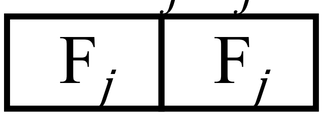
Case (1) F_j occurs in F_j



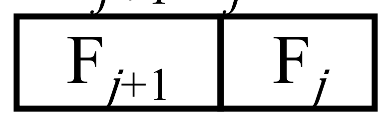
Case (2) F_j occurs in F_{j+1}



Case (3) F_j occurs across the border of F_j F_j

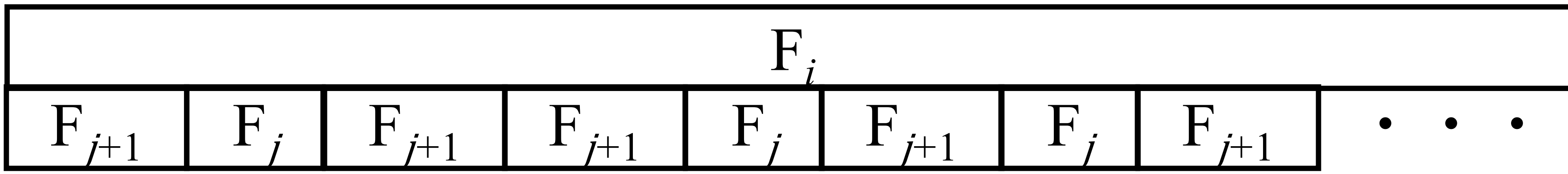


Case (4) F_j occurs across the border of F_{j+1} F_j

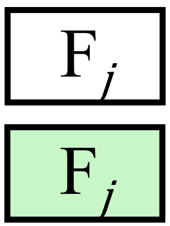


Since $f_j < f_{j+1}$, there are only four cases where F_j , occurs in F_i.

Occurrences of Fibonacci Words of smaller order

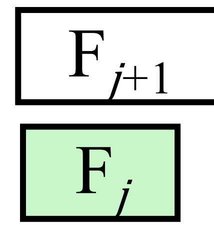


Case (1) F_j occurs in F_j



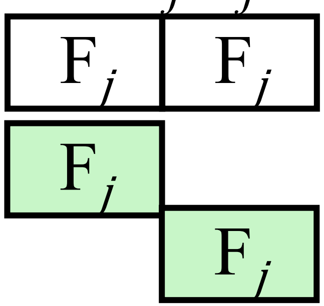
only once

Case (2) F_j occurs in F_{j+1}



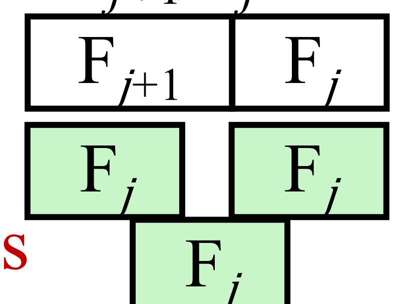
only once [Navarro et al., 2020]

Case (3) F_j occurs across the border of F_j F_j



only twice [Droubay, 1995]

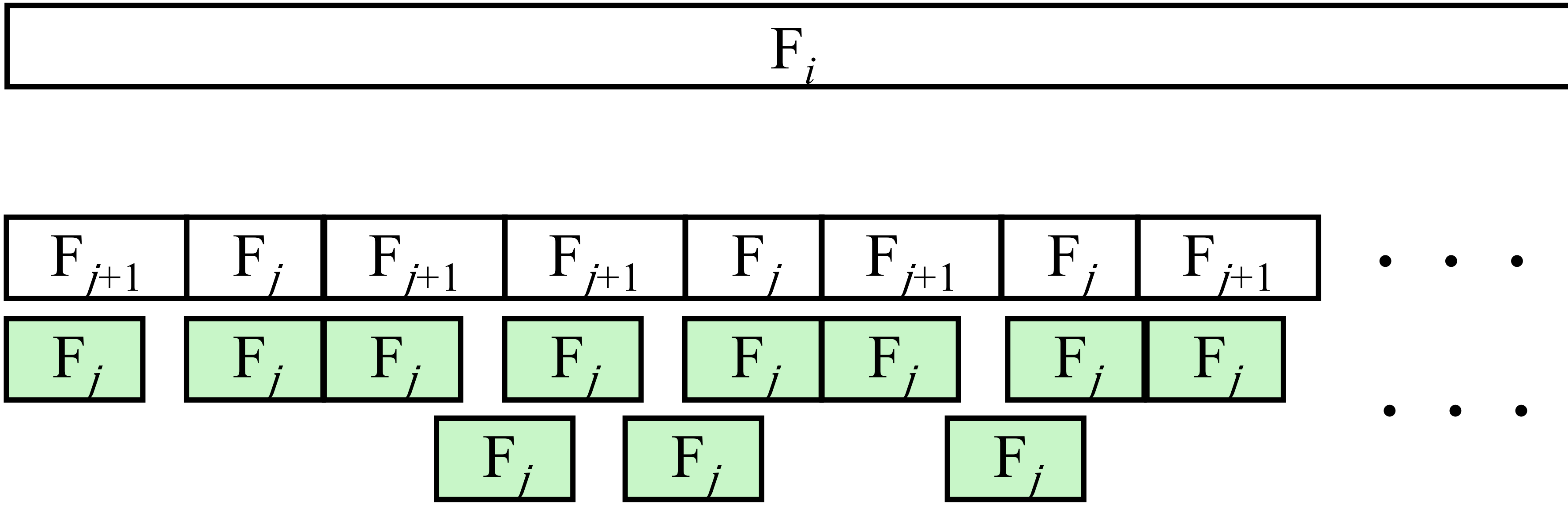
Case (4) F_j occurs across the border of F_{j+1} F_j



only three times [Navarro et al., 2020]

We can characterize the occurrence of F_j in all cases using some theorem about Fibonacci words.

Occurrences of Fibonacci Words of smaller order



Therefore, we can characterize all the occurrence of F_j in F_i .

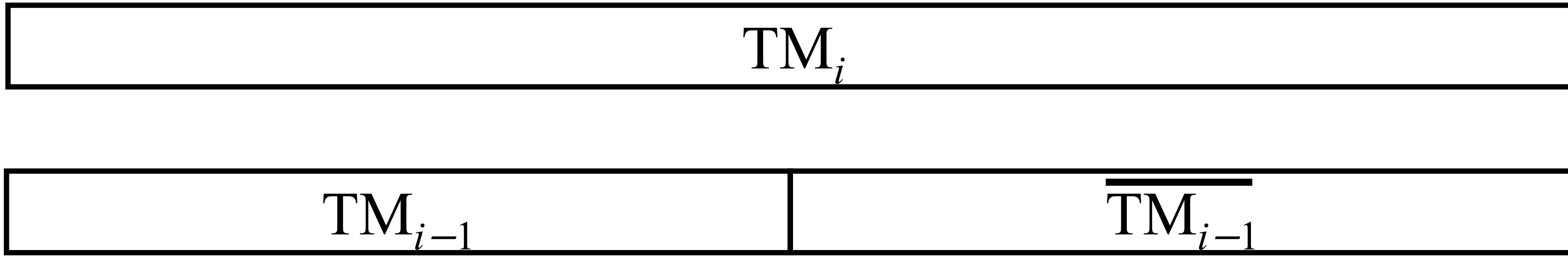
Note : This is also proved in [Iliopoulos et al., 1997] and [Rytter, 2006] in a different way.

Factorization of TM_i with smaller order

TM_i

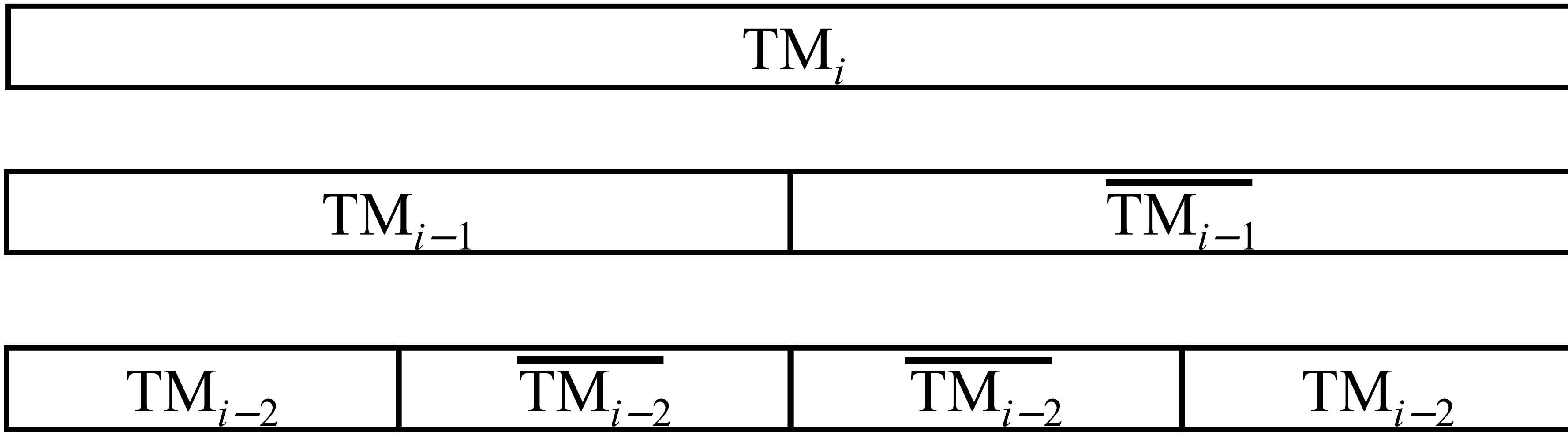
There is a factorization of TM_i, where all factors are TM_j or $\overline{\mathrm{TM}_j}$ when $j \leqq i$.

Factorization of TM_i with smaller order



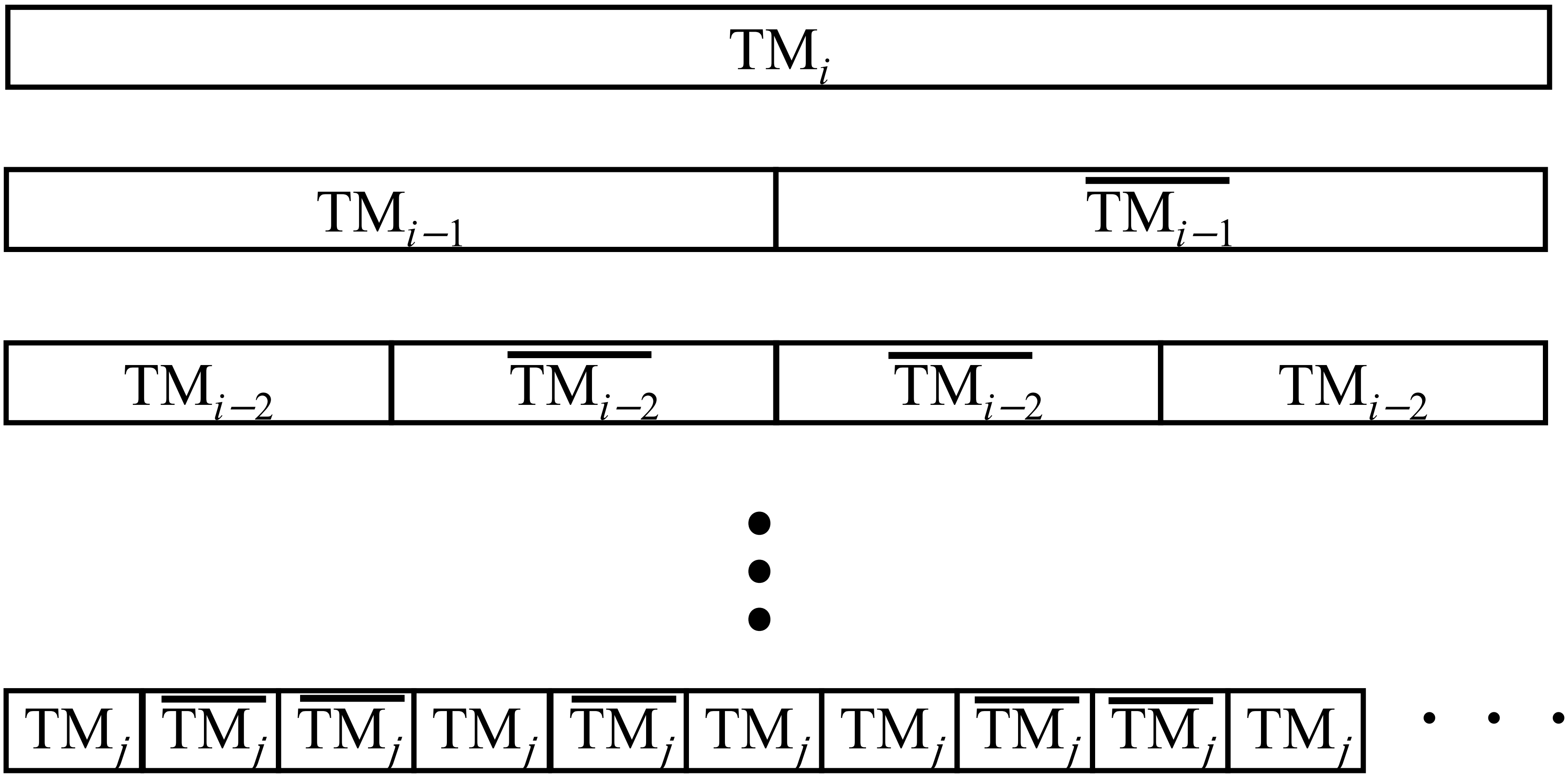
There is a factorization of TM_i , where all factors are TM_j or $\overline{TM_j}$ when $j \leqq i$.

Factorization of TM_i with smaller order



There is a factorization of TM_i , where all factors are TM_j or $\overline{\mathrm{TM}_j}$ when $j \leqq i$.

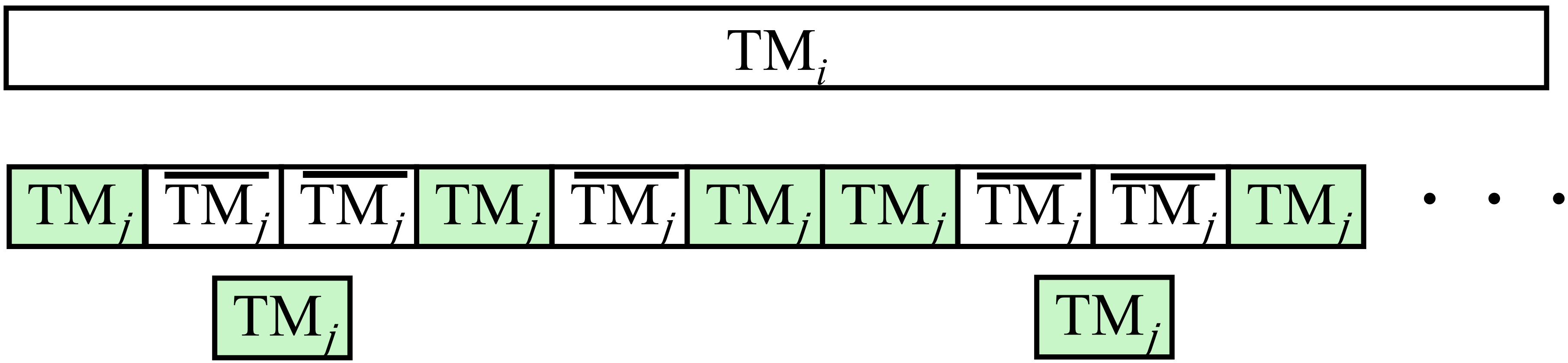
Factorization of TM_i with smaller order



There is a factorization of TM_i , where all factors are TM_j or $\overline{\mathrm{TM}_j}$ when $j \leqq i$.

Occurrences of Thue-Morse Words of smaller order

Theorem [Lothaire, 1997]
TM_i has no overlapping occurrences of the same string.

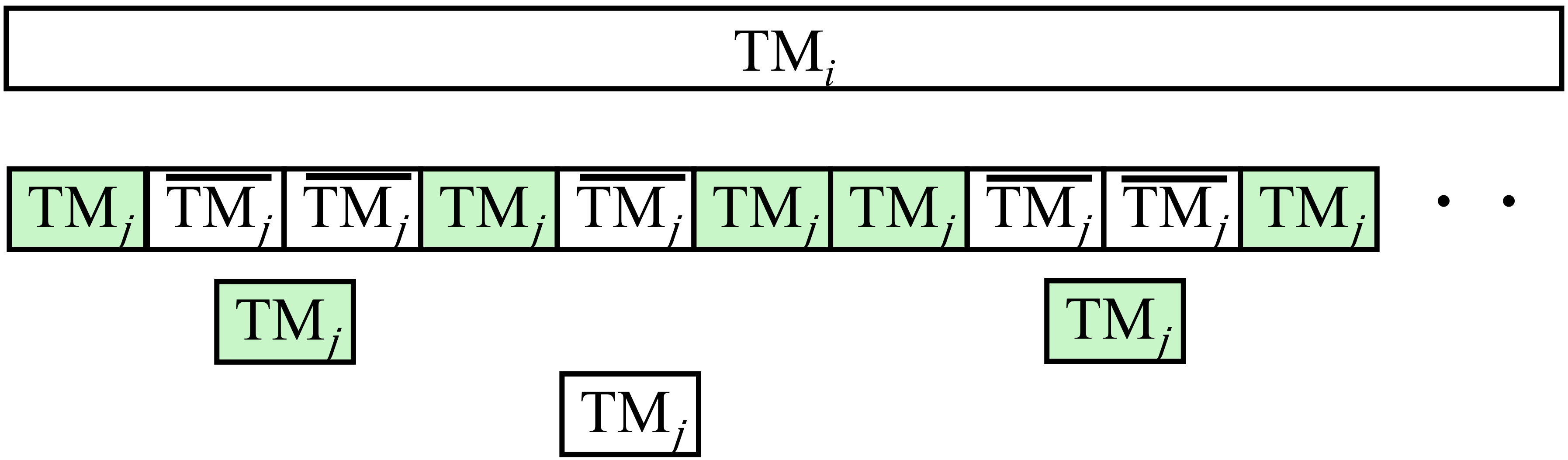


Since Thue-Morse words is overlapping free and TM_j occurs only once in $\overline{TM_j}\,\overline{TM_j}$, we can characterize all of the occurrences of TM_j in TM_i .

Note : This is also proved in [Radoszewski and Rytter, 2012] in a different way.

Occurrences of Thue-Morse Words of smaller order

Theorem [Lothaire, 1997]
TM_i has no overlapping occurrences of the same string.

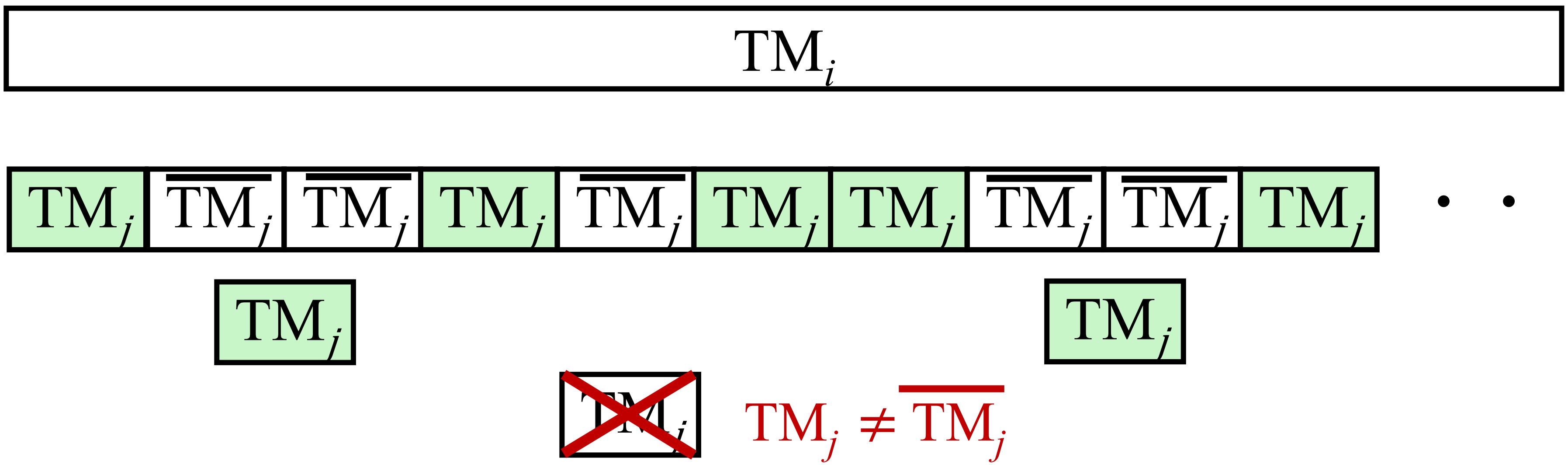


Since Thue-Morse words is overlapping free and TM_j occurs only once in $\overline{TM_j}\,\overline{TM_j}$, we can characterize all of the occurrences of TM_j in TM_i .

Note : This is also proved in [Radoszewski and Rytter, 2012] in a different way.

Occurrences of Thue-Morse Words of smaller order

Theorem [Lothaire, 1997]
TM_i has no overlapping occurrences of the same string.

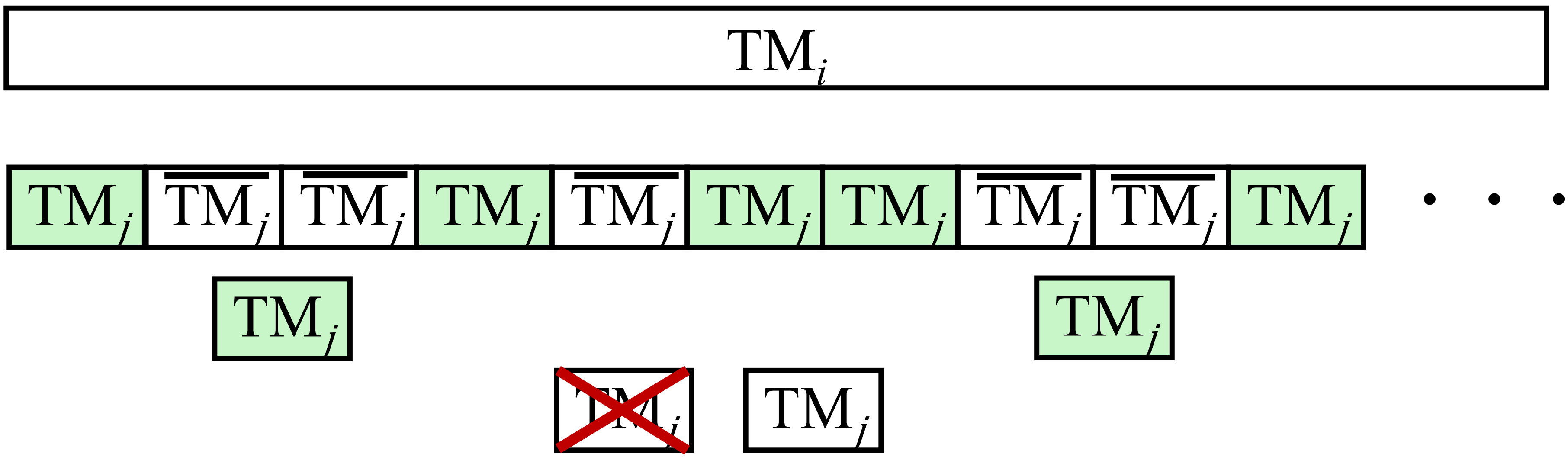


Since Thue-Morse words is overlapping free and TM_j occurs only once in $\overline{TM_j}\,\overline{TM_j}$, we can characterize all of the occurrences of TM_j in TM_i .

Note : This is also proved in [Radoszewski and Rytter, 2012] in a different way.

Occurrences of Thue-Morse Words of smaller order

Theorem [Lothaire, 1997]
TM_i has no overlapping occurrences of the same string.

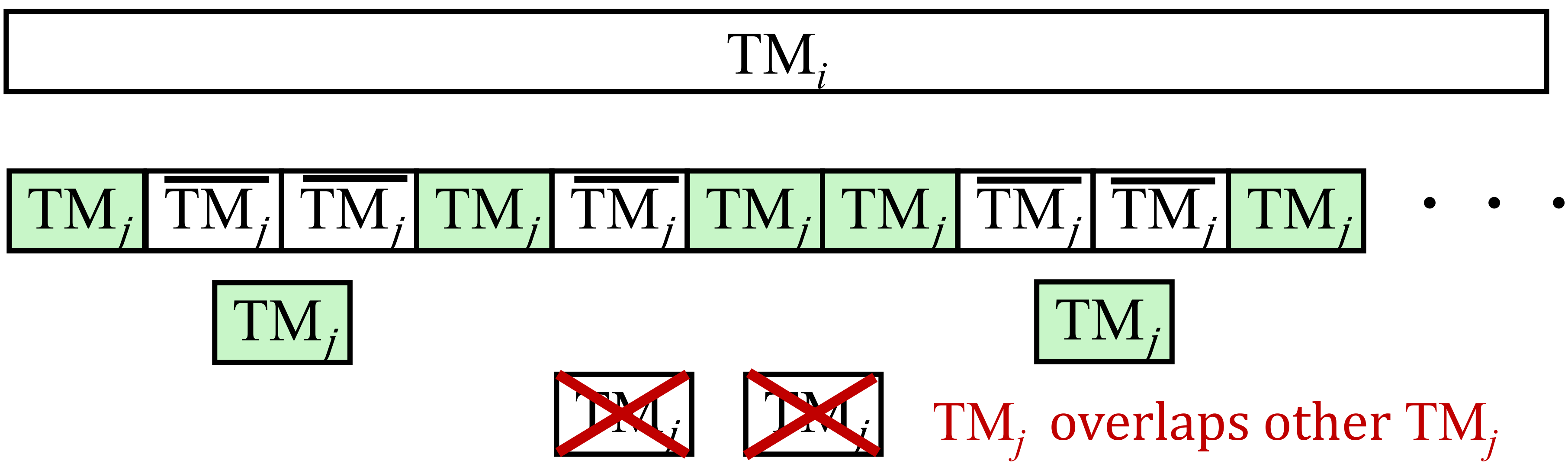


Since Thue-Morse words is overlapping free and TM_j occurs only once in $\overline{TM_j}\,\overline{TM_j}$, we can characterize all of the occurrences of TM_j in TM_i .

Note : This is also proved in [Radoszewski and Rytter, 2012] in a different way.

Occurrences of Thue-Morse Words of smaller order

Theorem [Lothaire, 1997]
TM_i has no overlapping occurrences of the same string.



Since Thue-Morse words is overlapping free and TM_j occurs only once in $\overline{TM_j}\,\overline{TM_j}$, we can characterize all of the occurrences of TM_j in TM_i .

Note : This is also proved in [Radoszewski and Rytter, 2012] in a different way.

Conclusion

- We introduced the net occurrences and its some properties.
- We showed characterization of occurrences of Fibonacci words and Thue-Morse Words of smaller order.
- We showed there are only three net occurrences in Fibonacci words and only nine net occurrences in Thue-Morse words.

Walnut and Net Occurrences

- Walnut is a free software that has been used to prove and disprove many theorems in combinatorics on words.
- Jeffrey Shallit provided alternative proofs using Walnut for our results on net occurrences in Fibonacci and Thue-Morse words.
- Jeffrey and we have been studying the net occurrences in other automatic sequences.

Future works

- Characterization of smaller-order occurrences of other words.

- Net occurrences of other morphic words,
  - Such as *k*-bonacci and Thue-Morse-like words.