SPIRE 2021

On the approximation ratio of LZ-End to LZ77

Takumi Ideue, Takuya Mieno, Mitsuru Funakoshi,
Yuto Nakashima, Shunsuke Inenaga, Masayuki Takeda
Kyushu University, Japan

LZ77 vs LZ-End

LZ77 [Ziv and Lempel, 1977] is the smallest greedy parsing allowing for left-to-right (de)compression.

LZ-End [Kreft and Navarro, 2013] is an LZ77-like parsing allowing for fast substring extraction,
but the number of its phrases is larger than that of LZ77.

Theorem: [This work]

There exist **binary** strings S such that:

$$\frac{z_{\mathrm{End}}(S)}{z_{77}(S)} \to 2 \ \ (|S| \to \infty).$$

$z_{\mathrm{End}}(S)$: # of LZ-End phrases of S
$z_{77}(S)$: # of LZ77 phrases of S

LZ77 [Ziv and Lempel, 1977]

Definition:

The non-overlapping Lempel-Ziv 77 factorization (LZ77) of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{77}(\boldsymbol{T}) = p_1, \ldots, p_z$ of $\boldsymbol{T}$ such that: Each phrase p_i $(1 \leq i \leq z - 1)$ satisfies the following condition.

- $p_i[1, |p_i| - 1]$ is the longest prefix of $p_i \cdots p_z$ which occurs in $p_1 \cdots p_{i-1}$.

The last phrase p_z can be **a suffix** of $\boldsymbol{T}$ which occurs in $p_1 \cdots p_{i-1}$.

z is the number of phrases

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\mathrm{LZ}_{77}(T)$ =	a	b	a	a	b	a	b	b	a	b	b	a	b	a	a	b	b

LZ77 [Ziv and Lempel, 1977]

Definition:

The non-overlapping Lempel-Ziv 77 factorization (LZ77) of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{77}(\boldsymbol{T}) = p_1, \ldots, p_z$ of $\boldsymbol{T}$ such that:
Each phrase p_i $(1 \leq i \leq z - 1)$ satisfies the following condition.

- $p_i[1, |p_i| - 1]$ is the longest prefix of $p_i \cdots p_z$ which occurs in $p_1 \cdots p_{i-1}$.

The last phrase p_z can be **a suffix** of $\boldsymbol{T}$ which occurs in $p_1 \cdots p_{i-1}$.

z is the number of phrases

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\mathrm{LZ}_{77}(T) =$	a |	b	a	a	b	a	b	b	a	b	b	a	b	a	a	b	b

First occurrence

LZ77 [Ziv and Lempel, 1977]

Definition:

The non-overlapping Lempel-Ziv 77 factorization (LZ77) of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{77}(\boldsymbol{T}) = p_1, \ldots, p_z$ of $\boldsymbol{T}$ such that: Each phrase p_i $(1 \leq i \leq z-1)$ satisfies the following condition.

- $p_i[1, |p_i| - 1]$ is the longest prefix of $p_i \cdots p_z$ which occurs in $p_1 \cdots p_{i-1}$.

The last phrase p_z can be **a suffix** of $\boldsymbol{T}$ which occurs in $p_1 \cdots p_{i-1}$.

z is the number of phrases

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\mathrm{LZ}_{77}(T)$ =	a |	b |	a	a	b	a	b	b	a	b	b	a	b	a	a	b	b

First occurrence

LZ77 [Ziv and Lempel, 1977]

Definition:

The non-overlapping Lempel-Ziv 77 factorization (LZ77) of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{77}(\boldsymbol{T}) = p_1, \ldots, p_z$ of $\boldsymbol{T}$ such that:
Each phrase p_i $(1 \leq i \leq z - 1)$ satisfies the following condition.

- $p_i[1, |p_i| - 1]$ is the longest prefix of $p_i \cdots p_z$ which occurs in $p_1 \cdots p_{i-1}$.

The last phrase p_z can be **a suffix** of $\boldsymbol{T}$ which occurs in $p_1 \cdots p_{i-1}$.

z is the number of phrases

E.g.)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

$\mathrm{LZ}_{77}(T)$ = a | b | a a | b a b b a b b a b a a b b

The longest prefix of $p_3 \cdots$

LZ77 [Ziv and Lempel, 1977]

Definition:

The non-overlapping Lempel-Ziv 77 factorization (LZ77) of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{77}(\boldsymbol{T}) = p_1, \ldots, p_z$ of $\boldsymbol{T}$ such that: Each phrase p_i $(1 \le i \le z - 1)$ satisfies the following condition.

- $p_i[1, |p_i| - 1]$ is the longest prefix of $p_i \cdots p_z$ which occurs in $p_1 \cdots p_{i-1}$.

The last phrase p_z can be **a suffix** of $\boldsymbol{T}$ which occurs in $p_1 \cdots p_{i-1}$.

z is the number of phrases

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\mathrm{LZ}_{77}(T) =$	a |	b |	a	a |	b	a	b |	b	a	b	b	a	b	a	a	b	b

The longest prefix of $p_4 \cdots$

LZ77 [Ziv and Lempel, 1977]

Definition:

The non-overlapping Lempel-Ziv 77 factorization (LZ77) of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{77}(\boldsymbol{T}) = p_1, \ldots, p_z$ of $\boldsymbol{T}$ such that: Each phrase p_i $(1 \leq i \leq z - 1)$ satisfies the following condition.

- $p_i[1, |p_i| - 1]$ is the longest prefix of $p_i \cdots p_z$ which occurs in $p_1 \cdots p_{i-1}$.

The last phrase p_z can be **a suffix** of $\boldsymbol{T}$ which occurs in $p_1 \cdots p_{i-1}$.

z is the number of phrases

E.g.)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

$\mathrm{LZ}_{77}(T) =$ a | b | a a | b a b | b a b b a b a | a b b

LZ77 [Ziv and Lempel, 1977]

Definition:

The non-overlapping Lempel-Ziv 77 factorization (LZ77) of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{77}(\boldsymbol{T}) = p_1, \ldots, p_z$ of $\boldsymbol{T}$ such that: Each phrase p_i $(1 \leq i \leq z-1)$ satisfies the following condition.

- $p_i[1, |p_i| - 1]$ is the longest prefix of $p_i \cdots p_z$ which occurs in $p_1 \cdots p_{i-1}$.

The last phrase p_z can be **a suffix** of $\boldsymbol{T}$ which occurs in $p_1 \cdots p_{i-1}$.

z is the number of phrases

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\mathrm{LZ}_{77}(T) =$	a	| b	| a	a	| b	a	b	| b	a	b	b	a	b	a	| a	b	b

Non-overlapping

LZ77 [Ziv and Lempel, 1977]

Definition:

The non-overlapping Lempel-Ziv 77 factorization (LZ77) of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{77}(\boldsymbol{T}) = p_1, \ldots, p_z$ of $\boldsymbol{T}$ such that:
Each phrase p_i $(1 \le i \le z-1)$ satisfies the following condition.

- $p_i[1, |p_i| - 1]$ is the longest prefix of $p_i \cdots p_z$ which occurs in $p_1 \cdots p_{i-1}$.

The last phrase p_z can be **a suffix** of $\boldsymbol{T}$ which occurs in $p_1 \cdots p_{i-1}$.

z is the number of phrases

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\mathrm{LZ}_{77}(T)$ =	a |	b |	a	a |	b	a	b |	b	a	b	b |	a	b	a	a	b	b

The longest prefix of $p_5 \cdots$

LZ77 [Ziv and Lempel, 1977]

Definition:

The non-overlapping Lempel-Ziv 77 factorization (LZ77) of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{77}(\boldsymbol{T}) = p_1, \ldots, p_z$ of $\boldsymbol{T}$ such that: Each phrase p_i $(1 \leq i \leq z - 1)$ satisfies the following condition.

- $p_i[1, |p_i| - 1]$ is the longest prefix of $p_i \cdots p_z$ which occurs in $p_1 \cdots p_{i-1}$.

The last phrase p_z can be **a suffix** of $\boldsymbol{T}$ which occurs in $p_1 \cdots p_{i-1}$.

z is the number of phrases

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\mathrm{LZ}_{77}(T)$ =	a |	b |	a	a |	b	a	b |	b	a	b	b |	a	b	a	a	b	b |

The longest prefix of p_6

LZ77 [Ziv and Lempel, 1977]

Definition:

The non-overlapping Lempel-Ziv 77 factorization (LZ77) of a string $\boldsymbol{T}$ is the factorization $LZ_{77}(\boldsymbol{T}) = p_1, \ldots, p_z$ of $\boldsymbol{T}$ such that: Each phrase p_i $(1 \le i \le z - 1)$ satisfies the following condition.

- $p_i[1, |p_i| - 1]$ is the longest prefix of $p_i \cdots p_z$ which occurs in $p_1 \cdots p_{i-1}$.

The last phrase p_z can be **a suffix** of $\boldsymbol{T}$ which occurs in $p_1 \cdots p_{i-1}$.

z is the number of phrases

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$LZ_{77}(T) =$	a |	b |	a	a |	b	a	b |	b	a	b	b |	a	b	a	a	b	b |
	p_1	p_2	p_3		p_4			p_5				p_6					

$z_{77}(T) = 6$

The number of the LZ77 phrases of T

LZ-End [Kreft and Navarro, 2013]

Definition:

The LZ-End factorization of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{\mathrm{End}}(\boldsymbol{T}) = q_1, \ldots, q_{z'}$ of $\boldsymbol{T}$ such that:

z' is the number of phrases

Each phrase q_i $(1 \leq i \leq z' - 1)$ satisfies the following condition.

- $q_i[1, |q_i| - 1]$ is the longest prefix of $q_i \cdots q_{z'}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < i$.

The last phrase $q_{z'}$ can be **a suffix** of $\boldsymbol{T}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < z'$.

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\mathrm{LZ}_{\mathrm{End}}(T) =$	a	b	a	a	b	a	b	b	a	b	b	a	b	a	a	b	b

LZ-End [Kreft and Navarro, 2013]

Definition:

z' is the number of phrases

The LZ-End factorization of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{\mathrm{End}}(\boldsymbol{T}) = q_1, \ldots, q_{z'}$ of $\boldsymbol{T}$ such that:

Each phrase q_i $(1 \le i \le z' - 1)$ satisfies the following condition.

- $q_i[1, |q_i| - 1]$ is the longest prefix of $q_i \cdots q_{z'}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < i$.

The last phrase $q_{z'}$ can be **a suffix** of $\boldsymbol{T}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < z'$.

E.g.)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

$\mathrm{LZ}_{\mathrm{End}}(T)$ = a | b a a b a b b a b b a b a a b b

First occurrence

LZ-End [Kreft and Navarro, 2013]

Definition:

z' is the number of phrases

The LZ-End factorization of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{\mathrm{End}}(\boldsymbol{T}) = q_1, \ldots, q_{z'}$ of $\boldsymbol{T}$ such that:

Each phrase q_i ($1 \le i \le z' - 1$) satisfies the following condition.

- $q_i[1, |q_i| - 1]$ is the longest prefix of $q_i \cdots q_{z'}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < i$.

The last phrase $q_{z'}$ can be **a suffix** of $\boldsymbol{T}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < z'$.

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\mathrm{LZ}_{\mathrm{End}}(T) =$	a |	b |	a	a	b	a	b	b	a	b	b	a	b	a	a	b	b

First occurrence

LZ-End [Kreft and Navarro, 2013]

Definition:

z' is the number of phrases

The LZ-End factorization of a string $\boldsymbol{T}$ is the factorization $LZ_{End}(\boldsymbol{T}) = q_1, \ldots, q_{z'}$ of $\boldsymbol{T}$ such that:

Each phrase q_i $(1 \leq i \leq z' - 1)$ satisfies the following condition.

- $q_i[1, |q_i| - 1]$ is the longest prefix of $q_i \cdots q_{z'}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < i$.

The last phrase $q_{z'}$ can be **a suffix** of $\boldsymbol{T}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < z'$.

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$LZ_{End}(T)$ =	a |	b |	a	a |	b	a	b	b	a	b	b	a	b	a	a	b	b

Suffix of q_1

The longest prefix of $q_3 \cdots$

LZ-End [Kreft and Navarro, 2013]

Definition:

z' is the number of phrases

The LZ-End factorization of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{\mathrm{End}}(\boldsymbol{T}) = q_1, \ldots, q_{z'}$ of $\boldsymbol{T}$ such that:

Each phrase q_i $(1 \leq i \leq z' - 1)$ satisfies the following condition.

- $q_i[1, |q_i| - 1]$ is the longest prefix of $q_i \cdots q_{z'}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < i$.

The last phrase $q_{z'}$ can be **a suffix** of $\boldsymbol{T}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < z'$.

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\mathrm{LZ}_{\mathrm{End}}(T)$ =	a |	b |	a	a |	b	a |	b	b	a	b	b	a	b	a	a	b	b

Suffix of $q_1 q_2$

The longest prefix of $q_4 \cdots$

LZ-End [Kreft and Navarro, 2013]

Definition:

The LZ-End factorization of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{\mathrm{End}}(\boldsymbol{T}) = q_1, \ldots, q_{z'}$ of $\boldsymbol{T}$ such that:

z' is the number of phrases

Each phrase q_i $(1 \le i \le z' - 1)$ satisfies the following condition.

- $q_i[1, |q_i| - 1]$ is the longest prefix of $q_i \cdots q_{z'}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < i$.

The last phrase $q_{z'}$ can be **a suffix** of $\boldsymbol{T}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < z'$.

E.g.)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

$\mathrm{LZ}_{\mathrm{End}}(T)$ = a | b | a a | b a | b b | a b b a b a a b b

Suffix of $q_1 q_2$

The longest prefix of $q_5 \cdots$

LZ-End [Kreft and Navarro, 2013]

Definition:

The LZ-End factorization of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{\mathrm{End}}(\boldsymbol{T}) = q_1, \ldots, q_{z'}$ of $\boldsymbol{T}$ such that:

z' is the number of phrases

Each phrase q_i $(1 \le i \le z' - 1)$ satisfies the following condition.

- $q_i[1, |q_i| - 1]$ is the longest prefix of $q_i \cdots q_{z'}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < i$.

The last phrase $q_{z'}$ can be **a suffix** of $\boldsymbol{T}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < z'$.

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\mathrm{LZ}_{\mathrm{End}}(T)$ =	a |	b |	a	a |	b	a |	b	b |	a	b	b	a |	b	a	a	b	b

Suffix of $q_1q_2q_3q_4q_5$

The longest prefix of $q_6 \cdots$

LZ-End [Kreft and Navarro, 2013]

Definition:

The LZ-End factorization of a string $\boldsymbol{T}$ is the factorization $LZ_{End}(\boldsymbol{T}) = q_1, \ldots, q_{z'}$ of $\boldsymbol{T}$ such that:

z' is the number of phrases

Each phrase q_i $(1 \leq i \leq z' - 1)$ satisfies the following condition.

- $q_i[1, |q_i| - 1]$ is the longest prefix of $q_i \cdots q_{z'}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < i$.

The last phrase $q_{z'}$ can be **a suffix** of $\boldsymbol{T}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < z'$.

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$LZ_{End}(T)$ =	a |	b |	a	a |	b	a |	b	b |	a	b	b	a |	b	a	a	b |	b

Suffix of $q_1q_2q_3$

The longest prefix of $q_7 \cdots$

LZ-End [Kreft and Navarro, 2013]

Definition:

z' is the number of phrases

The LZ-End factorization of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{\mathrm{End}}(\boldsymbol{T}) = q_1, \ldots, q_{z'}$ of $\boldsymbol{T}$ such that:

Each phrase q_i $(1 \leq i \leq z' - 1)$ satisfies the following condition.

- $q_i[1, |q_i| - 1]$ is the longest prefix of $q_i \cdots q_{z'}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < i$.

The last phrase $q_{z'}$ can be **a suffix** of $\boldsymbol{T}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < z'$.

E.g.)

$$\begin{array}{lcccccccccccccccccc} & & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 \\ \mathrm{LZ}_{\mathrm{End}}(T) & = & a & |\,b\,| & a & a\,| & b & a\,| & b & b\,| & a & b & b & a\,| & b & a & a & b\,| & b\,| \end{array}$$

Suffix of $q_1 q_2$

The longest prefix of q_8 and a suffix of T

LZ-End [Kreft and Navarro, 2013]

Definition:

The LZ-End factorization of a string $\boldsymbol{T}$ is the factorization $\mathrm{LZ}_{\mathrm{End}}(\boldsymbol{T}) = q_1, \ldots, q_{z'}$ of $\boldsymbol{T}$ such that:

Each phrase q_i $(1 \leq i \leq z' - 1)$ satisfies the following condition.

- $q_i[1, |q_i| - 1]$ is the longest prefix of $q_i \cdots q_{z'}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < i$.

The last phrase $q_{z'}$ can be **a suffix** of $\boldsymbol{T}$ which occurs as a suffix of $q_1 \cdots q_j$ for some $j < z'$.

z' is the number of phrases

E.g.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\mathrm{LZ}_{\mathrm{End}}(T)$ =	a |	b |	a	a |	b	a |	b	b |	a	b	b	a |	b	a	a	b |	b |

Phrases: q_1 = a, q_2 = b, q_3 = aa, q_4 = ba, q_5 = bb, q_6 = abba, q_7 = baab, q_8 = b

$z_{\mathrm{End}}(T) = 8$ — The number of the LZ-End phrases of T

The ratio $z_{\mathrm{End}} / z_{77}$

It is known that $z_{\mathrm{End}}(T) \geq z_{77}(T)$ for any string T.
Then how much is the gap between them?
To analyze this, we consider the ratio $z_{\mathrm{End}} / z_{77}$.

E.g.)

$z_{77}(T) = 6$

$LZ_{77}(T) =$ a | b | a a | b a b | b a b b | a b a a b b |

$LZ_{\mathrm{End}}(T) =$ a | b | a a | b a | b b | a b b a | b a a b | b |

$z_{\mathrm{End}}(T) = 8$

In this case,

$$\frac{z_{\mathrm{End}}(T)}{z_{77}(T)} = \frac{8}{6} = 1.333\cdots.$$

Previous work

Theorem 1: [Kreft and Navarro, 2013]

There exist strings T of alphabet size $\sigma = \frac{|T|}{3} + 1$ such that:

$$\frac{z_{\mathrm{End}}(T)}{z_{77}(T)} \to 2 \ \ (|T| \to \infty).$$

Previous work

Theorem 1: [Kreft and Navarro, 2013]

There exist strings T of alphabet size $\sigma = \frac{|T|}{3} + 1$ such that:

$$\frac{z_{\mathrm{End}}(T)}{z_{77}(T)} \to 2 \ \ (|T| \to \infty).$$

$\Sigma = \{1, 2, \ldots, \sigma\}$

E.g.) $T = 1\ 1\ 2\ \ 1\ 1\ 3\ \ 2\ 1\ 4\ \ 3\ 2\ 5\ \ 4\ 3\ 6\ \ \ldots\ \ (\sigma-2)(\sigma-3)\sigma$

$\mathrm{LZ}_{77}(T) = 1 \mid 1\ 2 \mid 1\ 1\ 3 \mid 2\ 1\ 4 \mid 3\ 2\ 5 \mid 4\ 3\ 6 \mid \ldots \mid (\sigma-2)(\sigma-3)\sigma \mid$

$\mathrm{LZ}_{\mathrm{End}}(T) = 1 \mid 1\ 2 \mid 1\ 1 \mid 3 \mid 2\ 1 \mid 4 \mid 3\ 2 \mid 5 \mid 4\ 3 \mid 6 \mid \ldots \mid (\sigma-2)(\sigma-3) \mid \sigma \mid$

Previous work

Theorem 1: [Kreft and Navarro, 2013]

There exist strings T of alphabet size $\sigma = \frac{|T|}{3} + 1$ such that:

$$\frac{z_{\mathrm{End}}(T)}{z_{77}(T)} \to 2 \ \ (|T| \to \infty).$$

$\Sigma = \{1, 2, \ldots, \sigma\}$

E.g.) $T = 1\,1\,2 \quad 1\,1\,3 \quad 2\,1\,4 \quad 3\,2\,5 \quad 4\,3\,6 \quad \ldots \quad (\sigma-2)(\sigma-3)\sigma$

1

$\mathrm{LZ}_{77}(T) = 1 \,|\, 1\,2 \,|\, 1\,1\,3 \,|\, 2\,1\,4 \,|\, 3\,2\,5 \,|\, 4\,3\,6 \,|\, \ldots \,|\, (\sigma-2)(\sigma-3)\sigma \,|$

$\mathrm{LZ}_{\mathrm{End}}(T) = 1 \,|\, 1\,2 \,|\, 1\,1 \,|\, 3 \,|\, 2\,1 \,|\, 4 \,|\, 3\,2 \,|\, 5 \,|\, 4\,3 \,|\, 6 \,|\, \ldots \,|\, (\sigma-2)(\sigma-3) \,|\, \sigma \,|$

2

Previous work

Theorem 1: [Kreft and Navarro, 2013]

There exist strings T of alphabet size $\sigma = \frac{|T|}{3} + 1$ such that:

$$\frac{z_{\mathrm{End}}(T)}{z_{77}(T)} \to 2 \ \ (|T| \to \infty).$$

$\Sigma = \{1, 2, \ldots, \sigma\}$

E.g.) $T = 1\ 1\ 2\quad 1\ 1\ 3\quad 2\ 1\ 4\quad 3\ 2\ 5\quad 4\ 3\ 6\quad \ldots\quad (\sigma-2)(\sigma-3)\sigma$

$z_{77}(T) = \sigma$

$LZ_{77}(T) = 1 | 1\ 2 | 1\ 1\ 3 | 2\ 1\ 4 | 3\ 2\ 5 | 4\ 3\ 6 | \ldots | (\sigma-2)(\sigma-3)\sigma |$

$LZ_{\mathrm{End}}(T) = 1 | 1\ 2 | 1\ 1 | 3 | 2\ 1 | 4 | 3\ 2 | 5 | 4\ 3 | 6 | \ldots | (\sigma-2)(\sigma-3) | \sigma |$

$z_{\mathrm{End}}(T) = 2(\sigma - 1)$

Previous work

Theorem 1: [Kreft and Navarro, 2013]

There exist strings T of alphabet size $\sigma = \frac{|T|}{3} + 1$ such that:

$$\frac{z_{\mathrm{End}}(T)}{z_{77}(T)} \to 2 \ \ (|T| \to \infty).$$

$\Sigma = \{1, 2, \ldots, \sigma\}$

E.g.) $T = 1\,1\,2 \quad 1\,1\,3 \quad 2\,1\,4 \quad 3\,2\,5 \quad 4\,3\,6 \quad \ldots \quad (\sigma-2)(\sigma-3)\sigma$

$z_{77}(T) = \sigma$

$LZ_{77}(T) = 1 | 1\,2 \,|\, 1\,1\,3 \,|\, 2\,1\,4 \,|\, 3\,2\,5 \,|\, 4\,3\,6 \,|\, \ldots \,|\, (\sigma-2)(\sigma-3)\sigma |$

$LZ_{\mathrm{End}}(T) = 1 | 1\,2 \,|\, 1\,1 | 3 \,|\, 2\,1 | 4 \,|\, 3\,2 | 5 \,|\, 4\,3 | 6 \,|\, \ldots \,|\, (\sigma-2)(\sigma-3) | \sigma |$

$z_{\mathrm{End}}(T) = 2(\sigma - 1)$

$$\frac{z_{\mathrm{End}}(T)}{z_{77}(T)} = \frac{2(\sigma-1)}{\sigma} = 2 \ \ (|T| \to \infty, \sigma \to \infty)$$

Main result

Theorem 1: [Kreft and Navarro, 2013]

There exist strings T of alphabet size $\sigma = \frac{|T|}{3} + 1$ such that:

$$\frac{z_{\mathrm{End}}(T)}{z_{77}(T)} \to 2 \ \ (|T| \to \infty).$$

Theorem 2: [This work]

There exist strings S of alphabet size $\sigma = \mathbf{2}$ such that:

$$\frac{z_{\mathrm{End}}(S)}{z_{77}(S)} \to 2 \ \ (|S| \to \infty).$$

The string S in Theorem 2 is the period-doubling sequence.

Period-doubling sequence [Boston, 1980]

Definition:

The k-th period-doubling sequence S_k over $\Sigma = \{\mathrm{a}, \mathrm{b}\}$ is defined as follows:

- $S_0 = \mathrm{a}$
- $S_k = S_{k-1} \cdot S_{k-1}[1, n_{k-1}-1] \cdot \bar{\mathrm{c}} \quad (k \geq 1)$

n_{k-1} is the length of S_{k-1}, that is $n_{k-1} = |S_{k-1}|$.
c is the last character of S_{k-1}, that is c = $S_{k-1}[n_{k-1}]$.
$\bar{\mathrm{c}}$ is bit-flipped character of c.

Intuition:
Copy the first half and flip the last character.

Period-doubling sequence [Boston, 1980]

$S_0 = \mathrm{a}$

Intuition:
Copy the first half and flip the last character.

Period-doubling sequence [Boston, 1980]

$S_0 = a$

$S_1 = ab$

Intuition:
Copy the first half and flip the last character.

Period-doubling sequence [Boston, 1980]

$S_0 = \text{a}$

$S_1 = \text{a b}$

$S_2 = \text{a b a a}$

Intuition:
Copy the first half and flip the last character.

Period-doubling sequence [Boston, 1980]

$S_0 = \text{a}$

$S_1 = \text{a b}$

$S_2 = \text{a b a a}$

$S_3 = \text{a b a a a b a b}$

Intuition:
Copy the first half and flip the last character.

Period-doubling sequence [Boston, 1980]

Intuition:
Copy the first half and flip the last character.

$S_0 = \text{a}$

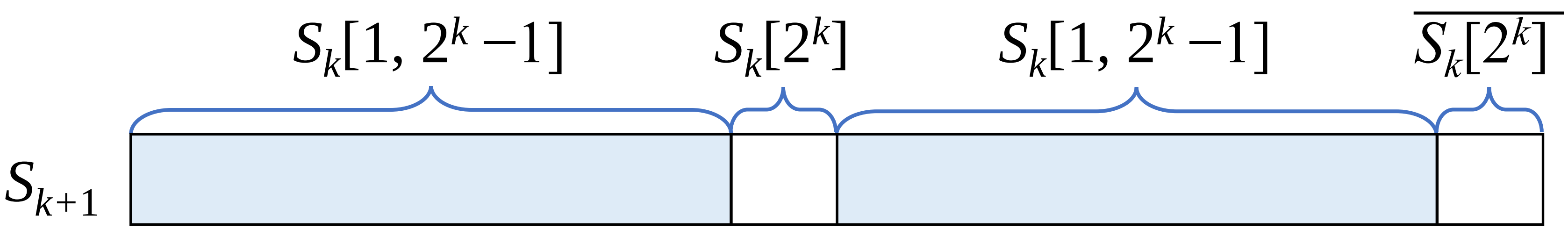
$S_1 = \text{a b}$

$S_2 = \text{a b a a}$

$S_3 = \text{a b a a a b a b}$

$S_4 = \underline{\text{a b a a a b a}}\ \underline{\text{b}}\ \underline{\text{a b a a a b a}}\ \underline{\text{a}}$

$\vdots$



LZ77 of period-doubling sequences S_k

LZ77 Phrase = Non-overlapping longest previous occurrence + Single character

$LZ_{77}(S_1)$ = a|b|

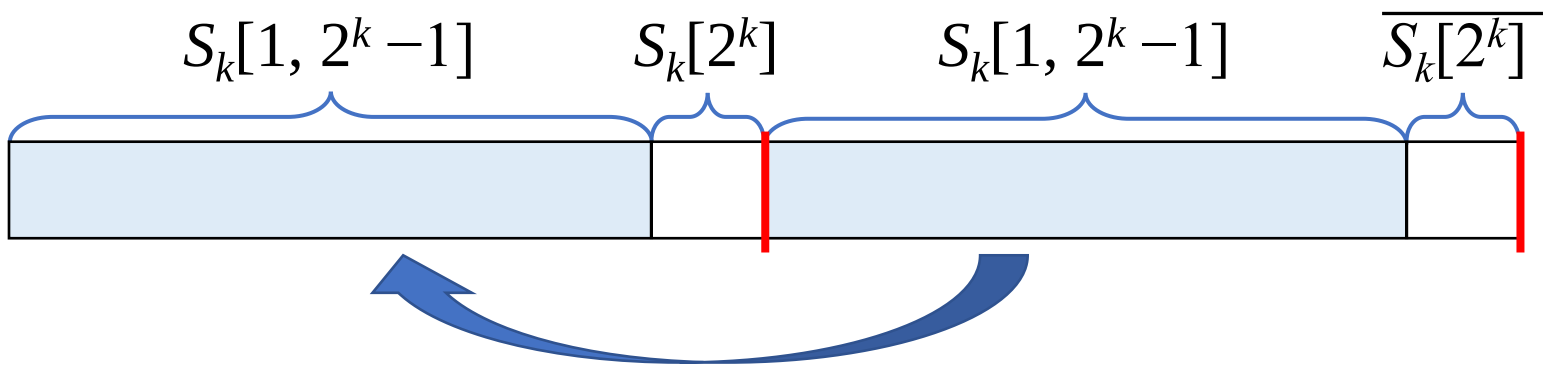
$LZ_{77}(S_2)$ = a|b|a a|

$LZ_{77}(S_3)$ = a|b|a a|a b a b|

$LZ_{77}(S_4)$ = a|b|a a|a b a b|a b a a a b a a|

$LZ_{77}(S_5)$ = a|b|a a|a b a b|a b a a a b a a|a b a a a b a b a b a a a b a b|

$LZ_{77}(S_{k+1})$



From definition of period-doubling sequence and LZ77,
$z_{77}(S_k) = k + 1$.

LZ-End of period-doubling sequences S_k

$\mathrm{LZ}_{\mathrm{End}}(S_1) =$ a|b|

$\mathrm{LZ}_{\mathrm{End}}(S_2) =$ a|b|a a|

$\mathrm{LZ}_{\mathrm{End}}(S_3) =$ a|b|a a|a b a|b|

$\mathrm{LZ}_{\mathrm{End}}(S_4) =$ a|b|a a|a b a|b a b|a a a b a a|

$\mathrm{LZ}_{\mathrm{End}}(S_5) =$ a b a a a b a b a b a a a b a a a b a a a b a b a b a a a b a b

LZ-End of period-doubling sequences S_k

$\mathrm{LZ}_{\mathrm{End}}(S_1) = \mathrm{a}|\mathrm{b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_2) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_3) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_4) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b\,a\,b}|\mathrm{a\,a\,a\,b\,a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_5) = \mathrm{a}|\mathrm{b\,a\,a\,a\,b\,a\,b\,a\,b\,a\,a\,a\,b\,a\,a\,a\,b\,a\,a\,a\,b\,a\,b\,a\,b\,a\,a\,a\,b\,a\,b}$

First occurrence

LZ-End of period-doubling sequences S_k

$\mathrm{LZ}_{\mathrm{End}}(S_1) = \mathrm{a}|\mathrm{b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_2) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_3) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_4) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b\,a\,b}|\mathrm{a\,a\,a\,b\,a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_5) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a\,a\,b\,a\,b\,a\,b\,a\,a\,a\,b\,a\,a\,a\,b\,a\,a\,a\,b\,a\,b\,a\,b\,a\,a\,a\,b\,a\,b}$

First occurrence

LZ-End of period-doubling sequences S_k

$\mathrm{LZ}_{\mathrm{End}}(S_1)$ = a|b|

$\mathrm{LZ}_{\mathrm{End}}(S_2)$ = a|b|a a|

$\mathrm{LZ}_{\mathrm{End}}(S_3)$ = a|b|a a|a b a|b|

$\mathrm{LZ}_{\mathrm{End}}(S_4)$ = a|b|a a|a b a|b a b|a a a b a a|

$\mathrm{LZ}_{\mathrm{End}}(S_5)$ = a|b|a a|a b a b a b a a a b a a a b a a a b a b a b a a a b a b

Ends with the 1st phrase

The longest prefix of the suffix at position 3

LZ-End of period-doubling sequences S_k

$\mathrm{LZ}_{\mathrm{End}}(S_1)$ = a|b|

$\mathrm{LZ}_{\mathrm{End}}(S_2)$ = a|b|a a|

$\mathrm{LZ}_{\mathrm{End}}(S_3)$ = a|b|a a|a b a|b|

$\mathrm{LZ}_{\mathrm{End}}(S_4)$ = a|b|a a|a b a|b a b|a a a b a a|

$\mathrm{LZ}_{\mathrm{End}}(S_5)$ = a|b|a a|a b a|b a b a a a b a a a b a a a b a b a b a a a b a b

Ends with
the 2nd phrase

The longest prefix of
the suffix at position 5

LZ-End of period-doubling sequences S_k

$\mathrm{LZ}_{\mathrm{End}}(S_1) = \mathrm{a}|\mathrm{b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_2) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_3) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_4) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b\,a\,b}|\mathrm{a\,a\,a\,b\,a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_5) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,\underline{b\,a}}|\mathrm{\underline{b\,a}\,b}|\mathrm{a\,a\,a\,b\,a\,a\,a\,b\,a\,a\,a\,b\,a\,b\,a\,b\,a\,a\,a\,b\,a\,b}$

Ends with
the 4th phrase

The longest prefix of
the suffix at position 8

LZ-End of period-doubling sequences S_k

$\mathrm{LZ}_{\mathrm{End}}(S_1) = \mathrm{a}|\mathrm{b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_2) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_3) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_4) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b\,a\,b}|\mathrm{a\,a\,a\,b\,a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_5) = \mathrm{a}|\mathrm{b}|\underline{\mathrm{a\,a}|\mathrm{a\,b\,a}}|\mathrm{b\,a\,b}|\underline{\mathrm{a\,a\,a\,b\,a}}\,\mathrm{a}|\mathrm{a\,b\,a\,a\,a\,b\,a\,b\,a\,b\,a\,a\,a\,b\,a\,b}$

Ends with
the 4th phrase

The longest prefix of
the suffix at position 11

LZ-End of period-doubling sequences S_k

$\mathrm{LZ}_{\mathrm{End}}(S_1) = \mathrm{a|b|}$

$\mathrm{LZ}_{\mathrm{End}}(S_2) = \mathrm{a|b|a\,a|}$

$\mathrm{LZ}_{\mathrm{End}}(S_3) = \mathrm{a|b|a\,a|a\,b\,a|b|}$

$\mathrm{LZ}_{\mathrm{End}}(S_4) = \mathrm{a|b|a\,a|a\,b\,a|b\,a\,b|a\,a\,a\,b\,a\,a|}$

$\mathrm{LZ}_{\mathrm{End}}(S_5) = \underline{\mathrm{a|b|a\,a|a\,b\,a|b\,a\,b}}\mathrm{|a\,a\,a\,b\,a\,a|}\underline{\mathrm{a\,b\,a\,a\,a\,b\,a\,b\,a\,b}}\mathrm{\,a|a\,a\,b\,a\,b}$

Ends with
the 5th phrase

The longest prefix of
the suffix at position 17

LZ-End of period-doubling sequences S_k

$\mathrm{LZ}_{\mathrm{End}}(S_1) = \mathrm{a}|\mathrm{b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_2) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_3) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_4) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b\,a\,b}|\mathrm{a\,a\,a\,b\,a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_5) = \mathrm{a}|\mathrm{b}|\mathrm{a}\,\underline{\mathrm{a}|\mathrm{a\,b\,a}}|\mathrm{b\,a\,b}|\mathrm{a\,a\,a\,b\,a\,a}|\mathrm{a\,b\,a\,a\,a\,b\,a\,b\,a\,b\,a}|\underline{\mathrm{a\,a\,b\,a}}\,\mathrm{b}|$

Ends with
the 4th phrase

The longest prefix of
the suffix at position 28

LZ-End of period-doubling sequences S_k

$\mathrm{LZ}_{\mathrm{End}}(S_1) = \mathrm{a}|\mathrm{b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_2) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_3) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_4) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b\,a\,b}|\mathrm{a\,a\,a\,b\,a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_5) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b\,a\,b}|\mathrm{a\,a\,a\,b\,a\,a}|\mathrm{a\,b\,a\,a\,a\,b\,a\,b\,a\,b\,a}|\mathrm{a\,a\,b\,a\,b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_6) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b\,a\,b}|\mathrm{a\,a\,a\,b\,a\,a}|\mathrm{a\,b\,a\,a\,a\,b\,a\,b\,a\,b\,a}|\mathrm{a\,a\,b\,a\,b} \vdots \mathrm{a\,b\,a\,a}|\mathrm{a} \ldots$

The last phrase of $\mathrm{LZ}_{\mathrm{End}}(S_k)$ is
not always $\mathrm{LZ}_{\mathrm{End}}(S_{k+1})$ phrase.

LZ-End of period-doubling sequences S_k

$\mathrm{LZ}_{\mathrm{End}}(S_1) = \mathrm{a}|\mathrm{b}|$

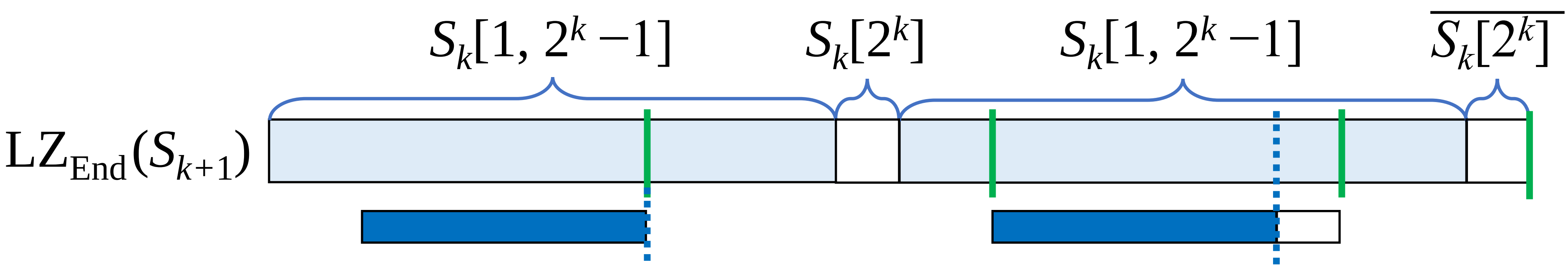
$\mathrm{LZ}_{\mathrm{End}}(S_2) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_3) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_4) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b\,a\,b}|\mathrm{a\,a\,a\,b\,a\,a}|$

$\mathrm{LZ}_{\mathrm{End}}(S_5) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b\,a\,b}|\mathrm{a\,a\,a\,b\,a\,a}|\mathrm{a\,b\,a\,a\,a\,b\,a\,b\,a\,b\,a}|\mathrm{a\,a\,b\,a\,b}|$

$\mathrm{LZ}_{\mathrm{End}}(S_6) = \mathrm{a}|\mathrm{b}|\mathrm{a\,a}|\mathrm{a\,b\,a}|\mathrm{b\,a\,b}|\mathrm{a\,a\,a\,b\,a\,a}|\mathrm{a\,b\,a\,a\,a\,b\,a\,b\,a\,b\,a}|\mathrm{a\,a\,b\,a\,b\,a\,b\,a\,a}|\mathrm{a} \ldots$



From definition of period-doubling sequence and LZ-End,
$z_{\mathrm{End}}(S_k) = 2k - O(\log^* k)$.

LZ-End of period-doubling sequences S_k

Observation 1:

$\mathrm{LZ}_{\mathrm{End}}(S_4)$

a|b|aa|aba|bab|aaabaa|

$\mathrm{LZ}_{\mathrm{End}}(S_5)$

a|b|aa|aba|bab|aaabaa|abaaabababa|aabab|

$\mathrm{LZ}_{\mathrm{End}}(S_6)$

a|b|aa|aba|bab|aaabaa|abaaabababa|aabababaa|abababaaabaaabaaabababaa|abaa|

LZ-End of period-doubling sequences S_k

Observation 1:

$\mathrm{LZ}_{\mathrm{End}}(S_4)$

a|b|aa|aba|bab|**aaabaa**|

$\mathrm{LZ}_{\mathrm{End}}(S_5)$

a|b|aa|aba|bab|aaabaa|abaaabababa|**aabab**|

$\mathrm{LZ}_{\mathrm{End}}(S_6)$

a|b|aa|aba|bab|aaabaa|abaaabababa|aabababaa|abababaaabaaabaaabababaa|**abaa**|

LZ-End of period-doubling sequences S_k

Observation 1:

$\mathrm{LZ}_{\mathrm{End}}(S_4)$

a|b|aa|aba|bab|aaabaa|

6

$\mathrm{LZ}_{\mathrm{End}}(S_5)$

a|b|aa|aba|bab|aaabaa|abaaababababa|aabab|

5

$\mathrm{LZ}_{\mathrm{End}}(S_6)$

a|b|aa|aba|bab|aaabaa|abaaababababa|aabababaa|abababaaabaaabaaababababaa|abaa|

4

The length of the last LZ-End phrase decreases by **1** until the length of the last phrase becomes **1**.

Increase of LZ-End phrases

Observation 2:

S_4

LZ77: a|b|aa|abab|abaaabaa|

LZ-End: a|b|aa|aba|bab|aaabaa|

S_5

LZ77: a|b|aa|abab|abaaabaa|abaaabababaaabab|

LZ-End: a|b|aa|aba|bab|aaabaa|abaaabababa|aabab|

S_6

LZ77: a|b|aa|abab|abaaabaa|abaaabababaaabab|abaaabababaaabaaabaaabababaaabaa|

LZ-End: a|b|aa|aba|bab|aaabaa|abaaabababa|aabababaa|abababaaabaaabaaabababaa|abaa|

Increase of LZ-End phrases

Observation 2:

S_4 — $z_{77}(S_4) = 5$, $z_{End}(S_4) = 6$

LZ77: a|b|aa|abab|abaaabaa|

LZ-End: a|b|aa|aba|bab|aaabaa|

S_5 — $z_{77}(S_5) = 6$, $z_{End}(S_5) = 8$

LZ77: a|b|aa|abab|abaaabaa|abaaabababaaabab|

LZ-End: a|b|aa|aba|bab|aaabaa|abaaabababa|aabab|

S_6 — $z_{77}(S_6) = 7$, $z_{End}(S_6) = 10$

LZ77: a|b|aa|abab|abaaabaa|abaaabababaaabab|abaaabababaaabaaabaaabababaaabaa|

LZ-End: a|b|aa|aba|bab|aaabaa|abaaabababa|aabababaa|abababaaabaaabaaabababaa|abaa|

Increase of LZ-End phrases

Observation 2:

Increasing number of phrases:
- LZ77: **1** (for any k)
- LZ-End: **2** (for almost all k)

S_4 $z_{77}(S_4) = \mathbf{5}$, $z_{End}(S_4) = \mathbf{6}$

LZ77: a|b|aa|abab|abaaabaa|

LZ-End: a|b|aa|aba|bab|aaabaa|

LZ77: + **1**

LZ-End: + **2**

S_5 $z_{77}(S_5) = \mathbf{6}$, $z_{End}(S_5) = \mathbf{8}$

LZ77: a|b|aa|abab|abaaabaa|abaaabababaaabab|

LZ-End: a|b|aa|aba|bab|aaabaa|abaaabababa|aabab|

LZ77: + **1**

LZ-End: + **2**

S_6 $z_{77}(S_6) = \mathbf{7}$, $z_{End}(S_6) = \mathbf{10}$

LZ77: a|b|aa|abab|abaaabaa|abaaabababaaabab|abaaabababaaabaaabaaabababaaabaa|

LZ-End: a|b|aa|aba|bab|aaabaa|abaaabababa|aabababaa|abababaaabaaabaaabababaa|abaa|

Table of LZ77 and LZ-End phrases

$z_{\mathrm{End}}(S_k) - z_{\mathrm{End}}(S_{k-1})$

k	z_{End}	z_{77}	$z_{\mathrm{End}} / z_{77}$	Length of the last LZ-End phrase	z_{End} diff
…	…	…	…	…	…
6	10	7	1.428…	4	2
7	12	8	1.5	3	2
8	14	9	1.555…	2	2
9	16	10	1.6	1	2
10	17	11	1.545…	384	1
11	19	12	1.583…	383	2
…	…	…	…	…	…
393	783	394	1.987…	1	2
394	784	395	1.984…	$3*2^{391}$	1
395	786	396	1.984…	$3*2^{391} - 1$	2
…	…	…	…	…	…
$3*2^{391} + 394$	…	…	1.999…	$3*2^{(3*2^{391} + 391)}$	1

Table of LZ77 and LZ-End phrases

$z_{\text{End}}(S_k) - z_{\text{End}}(S_{k-1})$

k	z_{End}	z_{77}	z_{End} / z_{77}	Length of the last LZ-End phrase	z_{End} diff
…	…	…	…	…	…
6	10	7	1.428…	4	2
7	12	8	1.5	3	2
8	14	9	1.555…	2	2
9	16	10	1.6	1	2
10	17	11	1.545…	384	1
11	19	12	1.583…	383	2
…	…	…	…	…	…
393	783	394	1.987…	1	2
394	784	395	1.984…	$3*2^{391}$	1
395	786	396	1.984…	$3*2^{391} - 1$	2
…	…	…	…	…	…
$3*2^{391} + 394$	…	…	1.999…	$3*2^{(3*2^{391} + 391)}$	1

Table of LZ77 and LZ-End phrases

$z_{End}(S_k) - z_{End}(S_{k-1})$

k	z_{End}	z_{77}	z_{End} / z_{77}	Length of the last LZ-End phrase	z_{End} diff
…	…	…	…	…	…
6	10	7	1.428…	4	2
7	12	8	1.5	3	2
8	14	9	1.555…	2	2
9	16	10	1.6	1	2
10	17	11	1.545…	384	1
11	19	12	1.583…	383	2
…	…	…	…	…	…
393	783	394	1.987…	1	2
394	784	395	1.984…	$3*2^{391}$	1
395	786	396	1.984…	$3*2^{391} - 1$	2
…	…	…	…	…	…
$3*2^{391} + 394$	…	…	1.999…	$3*2^{(3*2^{391} + 391)}$	1

Table of LZ77 and LZ-End phrases

$z_{End}(S_k) - z_{End}(S_{k-1})$

k	z_{End}	z_{77}	z_{End} / z_{77}	Length of the last LZ-End phrase	z_{End} diff
…	…	…	…	…	…
6	10	7	1.428…	4	2
7	1[illegible]				2
8	1[illegible]				2
9	1[illegible]				2
10	1[illegible]				1
11	19	12	1.583…	383	2
…	…	…	…	…	…
393	783	394	1.987…	1	2
394	784	395	1.984…	$3*2^{391}$	1
395	786	396	1.984…	$3*2^{391} - 1$	2
…	…	…	…	…	…
$3*2^{391} + 394$	…	…	1.999…	$3*2^{(3*2^{391} + 391)}$	1

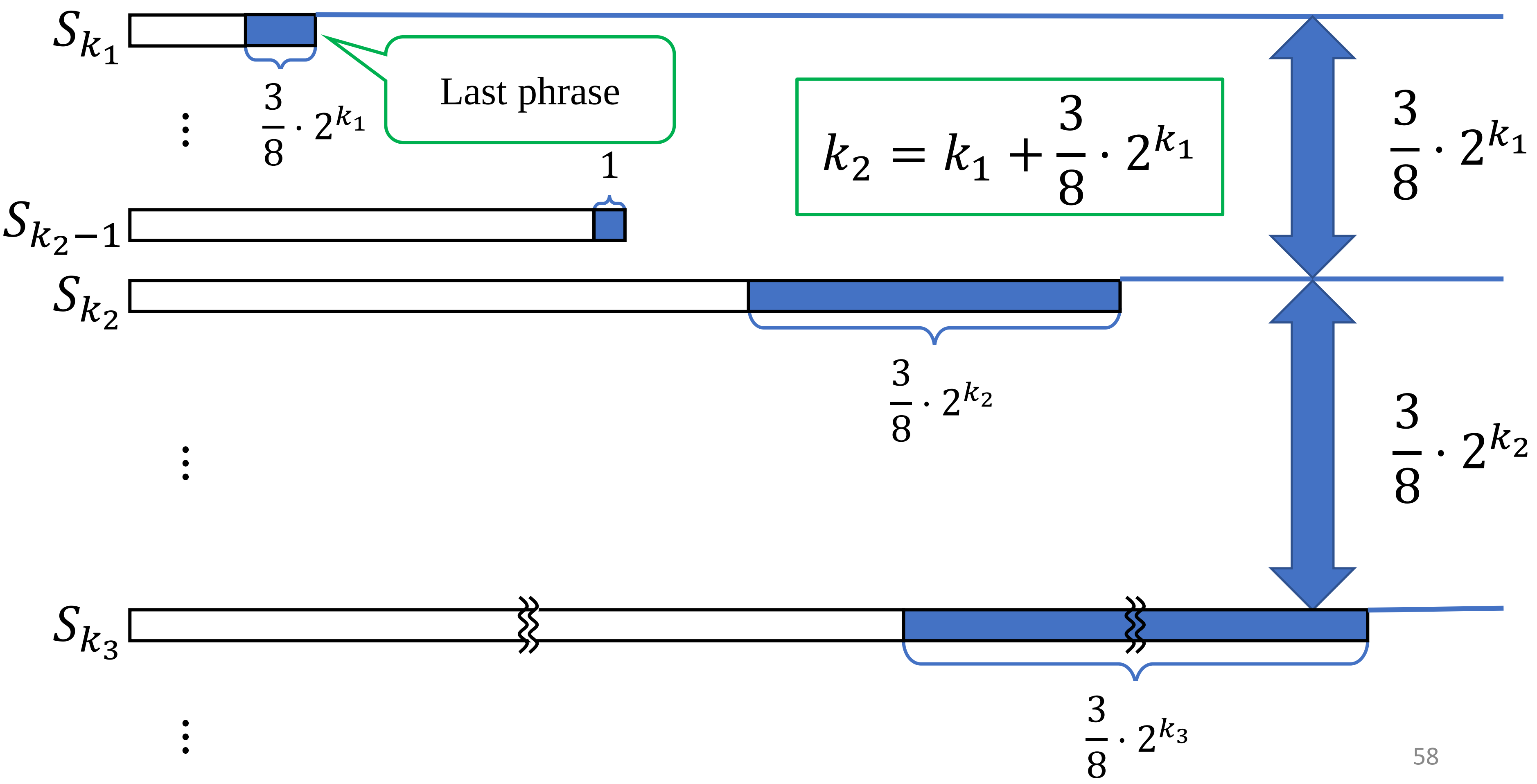
Lemma 1:
The number of **1**'s is $O(\log^* k)$.
Thus, $z_{End}(S_k) = 2k - O(\log^* k)$.

Why $O(\log^* k)$?

$$k_m = O\left(2^{2^{\cdot^{\cdot^{2^k}}}}\right) \Leftrightarrow m = O(\log^* k)$$

Lemma 2:

The maximal length of the last LZ-End phrase is $\frac{3}{8} \cdot 2^k$.



The ratio z_{End} / z_{77}

We obtain the following result.

$$z_{77}(S_k) = k + 1$$

$$z_{End}(S_k) = 2k - O(\log^* k)$$

$$\frac{z_{End}(S_k)}{z_{77}(S_k)} = \frac{2k - O(\log^* k)}{k+1} \to 2 \quad (k \to \infty).$$

Period-doubling sequence

Theorem 2:

There exist strings S of alphabet size $\sigma = \mathbf{2}$ such that:

$$\frac{z_{End}(S)}{z_{77}(S)} \to 2 \quad (|S| \to \infty).$$

Summary and future work

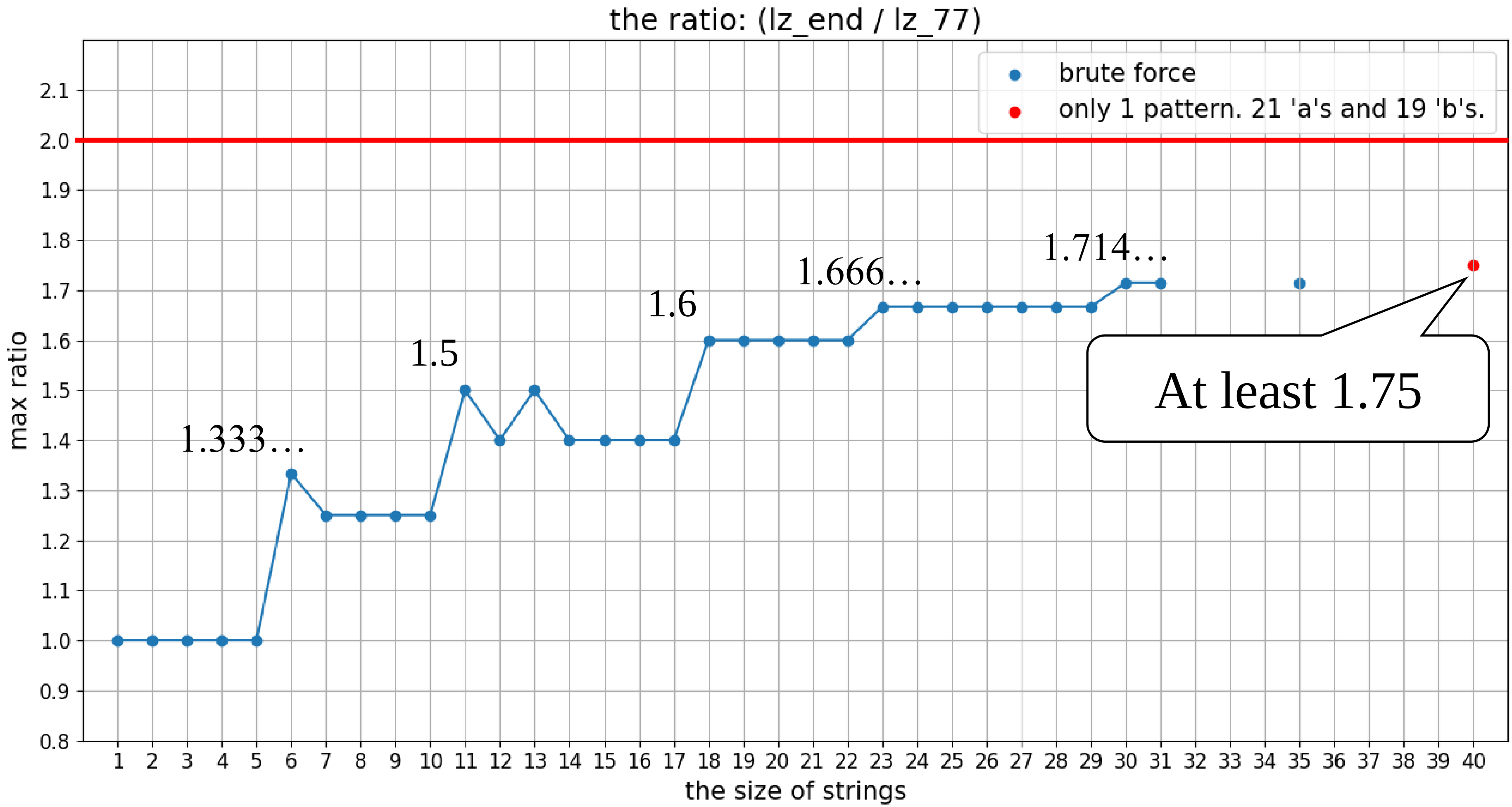
Summary:

- We proved that period-doubling sequence S satisfies that $z_{End}(S) / z_{77}(S)$ asymptotically approaches 2 when the limit as the length of S tends to infinity.
- There also exist other binary sequences S' such that $z_{End}(S') / z_{77}(S')$ asymptotically approaches 2.

Conjecture: [Kreft and Navarro, 2013]

$z_{End}(T) / z_{77}(T) \leq 2$ holds for any string T.

Preliminary experimental result



The ratio seems to asymptotically approach 2.

Summary and future work

Summary:

- We proved that period-doubling sequence S satisfies that $z_{\text{End}}(S)$ / $z_{77}(S)$ asymptotically approaches 2 when the limit as the length of S tends to infinity.
- There also exist other binary sequences S' such that $z_{\text{End}}(S')$ / $z_{77}(S')$ asymptotically approaches 2.

Future work:

- Prove or disprove the conjecture for upper bound.

Conjecture: [Kreft and Navarro, 2013]

$z_{\text{End}}(T)$ / $z_{77}(T) \leq 2$ holds for any string T.