Algoritmer og Datastrukturer 2
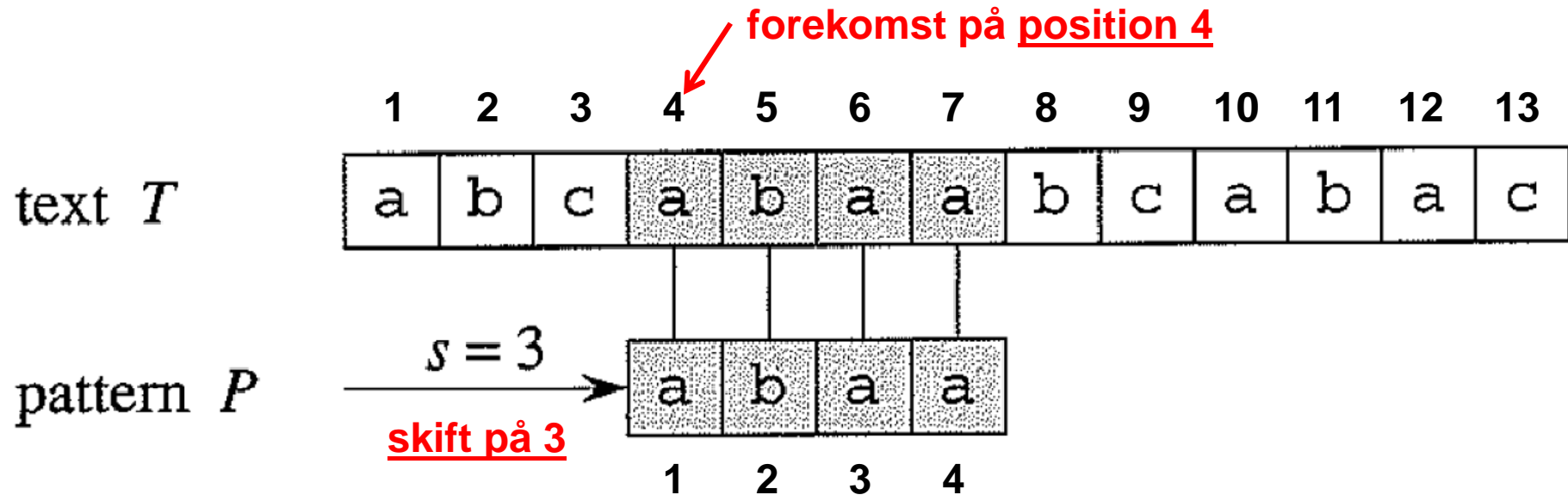
Gerth Stølting Brodal

Mønstergenkendelse [CLRS, kapitel 32.1-32.2, 32.4]

AARHUS UNIVERSITET

Mønster genkendelse

forekomst på position 4

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |

text T

| a | b | c | a | b | a | a | b | c | a | b | a | c |

$s = 3$

pattern P

skift på 3

| a | b | a | a |

| 1 | 2 | 3 | 4 |

Input: Tekst T af længde n og mønster P af længde m

Output: Alle positioner i T hvor P forekommer

Naive Algoritme

NAIVE-STRING-MATCHER(T, P)

1 $n \leftarrow length[T]$
2 $m \leftarrow length[P]$
3 **for** $s \leftarrow 0$ **to** $n - m$
4   **do if** $P[1 .. m] = T[s + 1 .. s + m]$
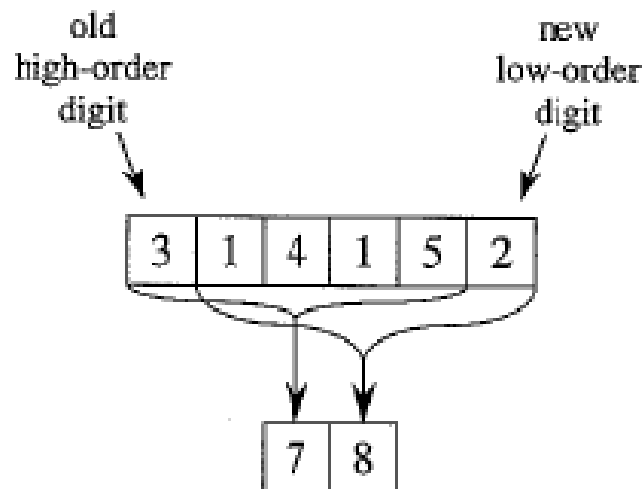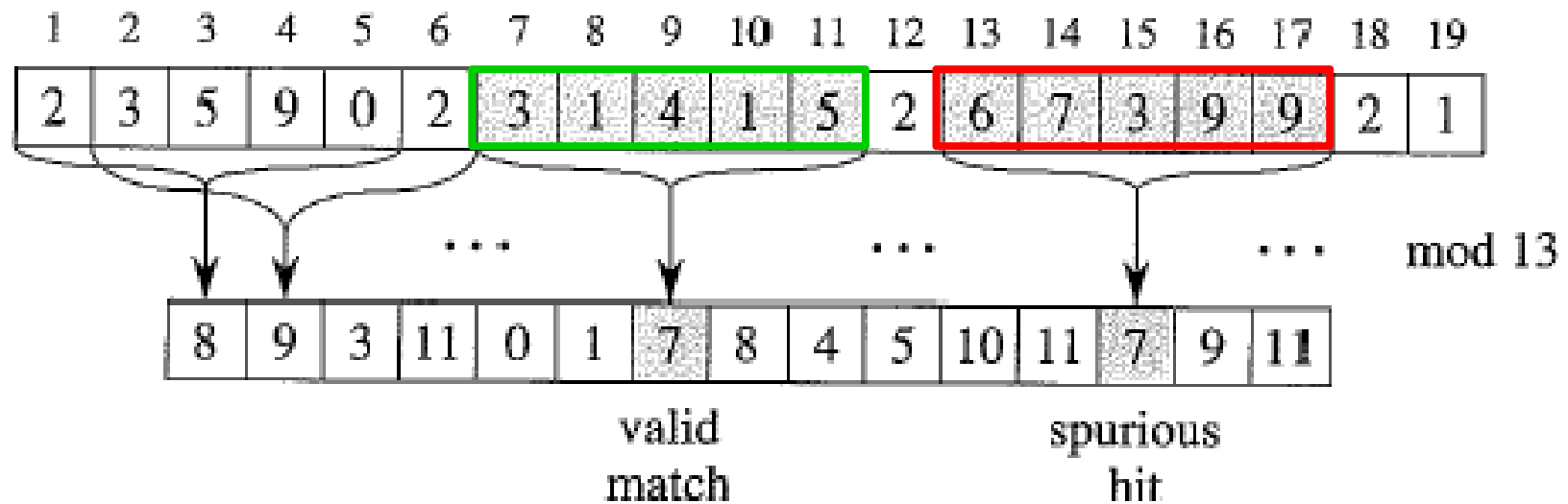5   **then** print "Pattern occurs with shift"

Rabin-Karp

RABIN-KARP-MATCHER(T, P, d, q)

1    $n \leftarrow length[T]$
2    $m \leftarrow length[P]$
3    $h \leftarrow d^{m-1} \bmod q$
4    $p \leftarrow 0$
5    $t_0 \leftarrow 0$
6   **for** $i \leftarrow 1$ **to** m $\triangleright$ Preprocessing.
7   **do** $p \leftarrow (dp + P[i]) \bmod q$
8          $t_0 \leftarrow (dt_0 + T[i]) \bmod q$
9   **for** $s \leftarrow 0$ **to** $n - m$ $\triangleright$ Matching.
10   **do if** $p = t_s$
11   **then if** $P[1 .. m] = T[s+1 .. s+m]$
12   **then** print "Pattern occurs with shift" s
13   **if** $s < n - m$
14   **then** $t_{s+1} \leftarrow (d(t_s - T[s+1]h) + T[s+m+1]) \bmod q$

Rabin-Karp: Eksempel

old high-order digit ... new low-order digit

old high-order digit ... shift ... new low-order digit

$$14152 \equiv (31415 - 3 \cdot 10000) \cdot 10 + 2 \pmod{13}$$
$$\equiv (7 - 3 \cdot 3) \cdot 10 + 2 \pmod{13}$$
$$\equiv 8 \pmod{13}$$

Knuth-Morris-Pratt

KMP-MATCHER(T, P)

1 $n \leftarrow length[T]$

2 $m \leftarrow length[P]$

3 $\pi \leftarrow$ COMPUTE-PREFIX-FUNCTION(P)

4 $q \leftarrow 0$ ▷ Number of characters matched.

5 **for** $i \leftarrow 1$ **to** n ▷ Scan the text from left to right.

6   **do while** $q > 0$ and $P[q + 1] \neq T[i]$

7   **do** $q \leftarrow \pi[q]$ ▷ Next character does not match.

8   **if** $P[q + 1] = T[i]$

9   **then** $q \leftarrow q + 1$ ▷ Next character matches.

10   **if** $q = m$ ▷ Is all of P matched?

11   **then** print "Pattern occurs with shift" $i - m$

12             $q \leftarrow \pi[q]$ ▷ Look for the next match.

Knuth-Morris-Pratt: Eksempel

(a)

(b)

(c)
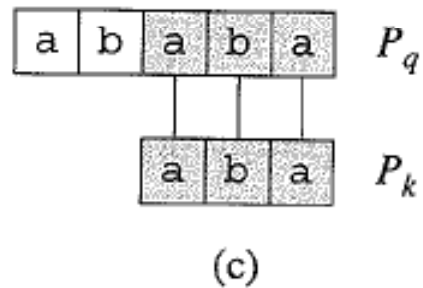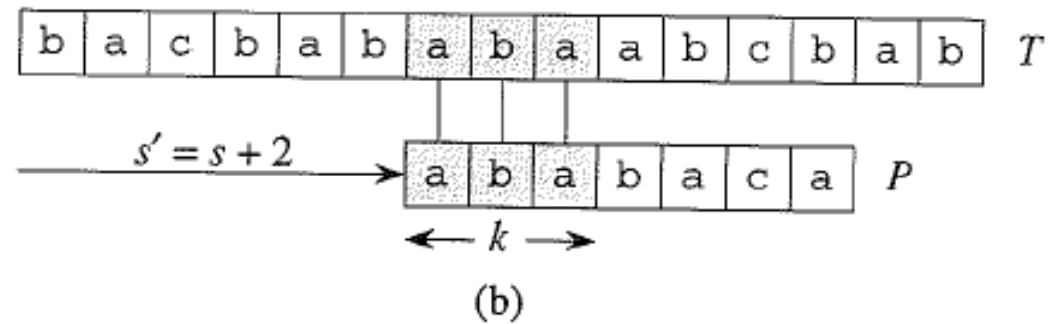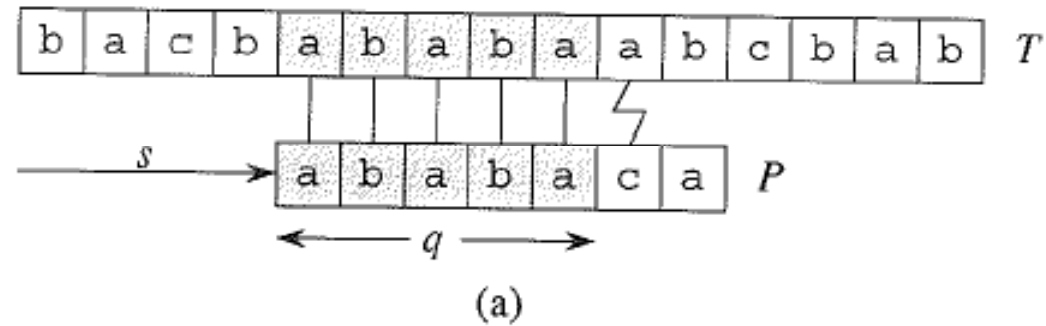
Knuth-Morris-Pratt:
Beregning af prefix funktionen
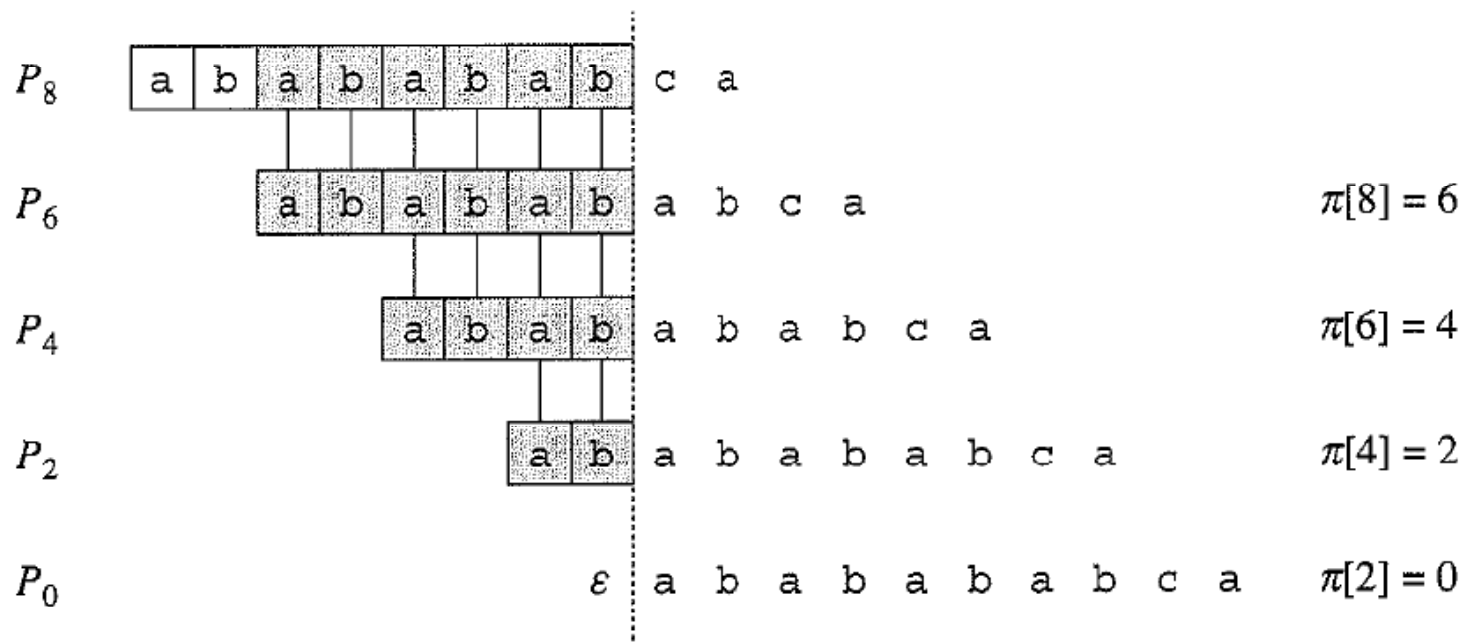
COMPUTE-PREFIX-FUNCTION(P)

1    $m \leftarrow length[P]$
2    $\pi[1] \leftarrow 0$
3    $k \leftarrow 0$
4   **for** $q \leftarrow 2$ **to** m
5   **do while** $k > 0$ and $P[k+1] \neq P[q]$
6   **do** $k \leftarrow \pi[k]$
7   **if** $P[k+1] = P[q]$
8   **then** $k \leftarrow k+1$
9         $\pi[q] \leftarrow k$
10   **return** π

Knuth-Morris-Pratt:
Beregning af prefix funktionen

i	1	2	3	4	5	6	7	8	9	10
$P[i]$	a	b	a	b	a	b	a	b	c	a
$\pi[i]$	0	0	1	2	3	4	5	6	0	1

(a)

P_8 a b a b a b a b c a

P_6 a b a b a b a b c a $\pi[8] = 6$

P_4 a b a b a b a b c a $\pi[6] = 4$

P_2 a b a b a b a b c a $\pi[4] = 2$

P_0 ε a b a b a b a b c a $\pi[2] = 0$

(b)

Worst-case tider

Algorithm	Preprocessing time	Matching time	[CLRS]
Naive	0	$O((n - m + 1)m)$	**32.1**
Rabin-Karp	$\Theta(m)$	$O((n - m + 1)m)$	**32.2**
Finite automaton	$O(m\,|\Sigma|)$	$\Theta(n)$	**(32.3)**
Knuth-Morris-Pratt	$\Theta(m)$	$\Theta(n)$	**32.4**