Move structure, part 2: Querying

Ben Langmead

For original Keynote files email me ben.langmead@gmail.com

Move structure queries

$\langle\ p,\ \ell,\ \pi,\ \xi\ \rangle$

M

0	1	46	12
1	3	34	8
4	9	52	13
13	3	19	5
16	3	1	1
19	3	16	4
22	3	61	14
25	9	37	9
34	3	49	13
37	2	47	12
39	1	0	0
40	6	4	2
⋮	⋮	⋮	⋮

How to compute $\pi(i)$ (LF[i]) for any i?

Including *repeated* π, i.e. $\pi^x(i)$

Takaaki Nishimoto and Yasuo Tabei. Optimal-Time Queries on BWT-Runs Compressed Indexes. In 48th International Colloquium on Automata, Languages, and Programming (ICALP 2021). Leibniz International Proceedings in Informatics (LIPIcs), Volume 198, pp. 101:1-101:15

Move structure queries

If k is an offset & i is the ID of the P-run containing k, then

$$\pi(k) =$$

run	p	ℓ	π	ξ
0	0	1	46	12
1	1	3	34	8
2	4	9	52	13
3	13	3	19	5
⋮	⋮	⋮	⋮	⋮

$k = 10, \; i = 2$

P 110010000000010010010010010000000000000100101100000100100000000000100

Q 110010000010010010010010000000000000000100100000000110100100000000100

Move structure queries

If k is an offset & i is the ID of the P-run containing k, then

$$\pi(k) = M[i]\,.\,\pi + (k - M[i]\,.\,p)$$

run	p	ℓ	π	ξ
0	0	1	46	12
1	1	3	34	8
2	4	9	52	13
3	13	3	19	5
⋮	⋮	⋮	⋮	⋮

$k = 10, \ i = 2$

P 1100100000000100100100100100000000100101100000100100000000000100

Q 1100100000100100100100100000000000001001000000001101001000000000100

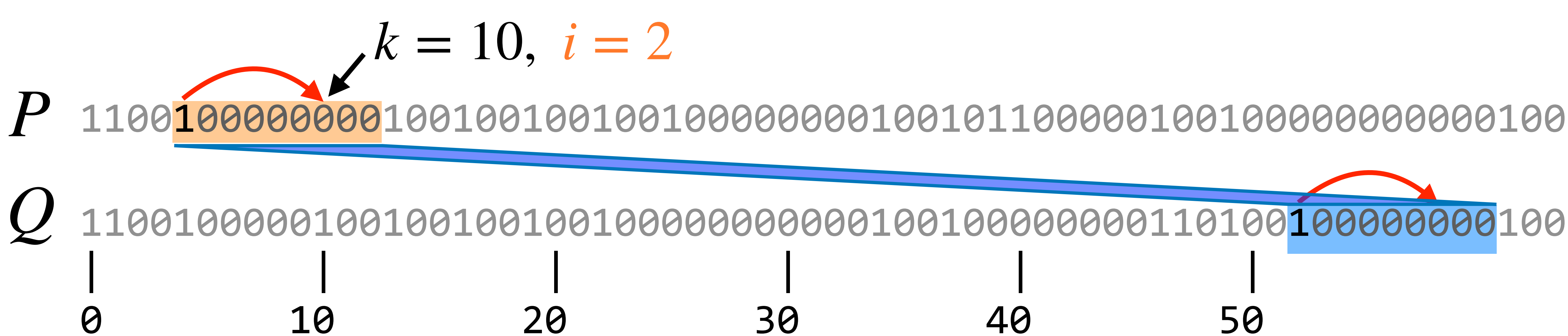
Move structure queries

If k is an offset & i is the ID of the P-run containing k, then

$$\pi(k) = M[i]\,.\,\pi + (k - M[i]\,.\,p)$$

$$52 + 6 = 58$$

run	p	ℓ	π	ξ
0	0	1	46	12
1	1	3	34	8
2	4	9	**52**	13
3	13	3	19	5
⋮	⋮	⋮	⋮	⋮

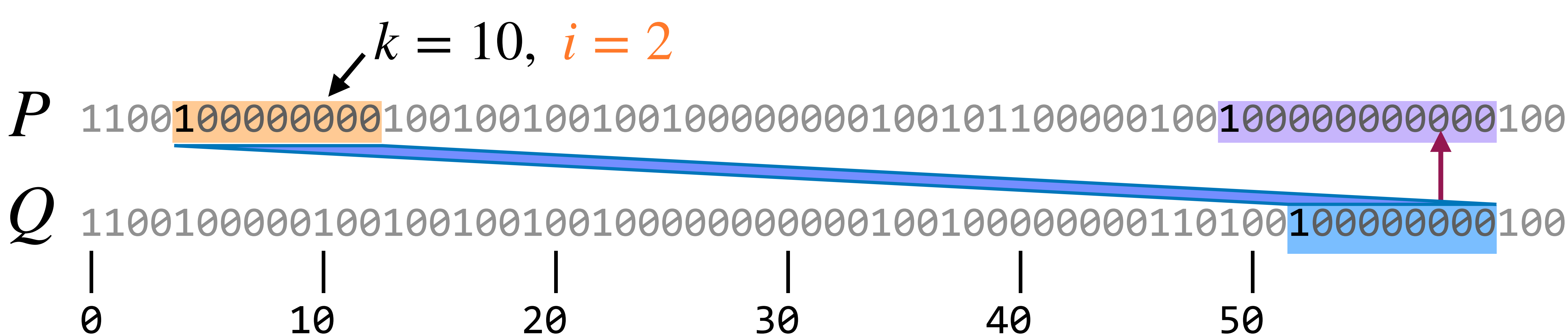


Move structure queries

k' (the "new" k) equals 58.

What is i'?

run	p,	ℓ,	π,	ξ
0	0	1	46	12
1	1	3	34	8
2	4	9	**52**	13
3	13	3	19	5
⋮	⋮	⋮	⋮	⋮



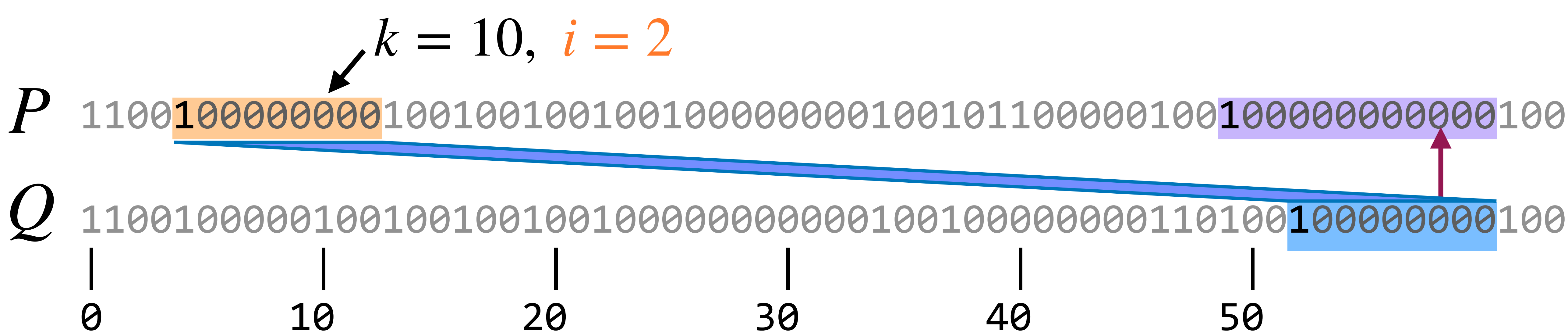
Move structure queries

k' (the "new" k) equals 58.

What is i'?

Is it $M[i]\,.\,\xi$?

run	p	ℓ	π	ξ
0	0	1	46	12
1	1	3	34	8
2	4	9	**52**	**13**
3	13	3	19	5
⋮	⋮	⋮	⋮	⋮



Move structure queries

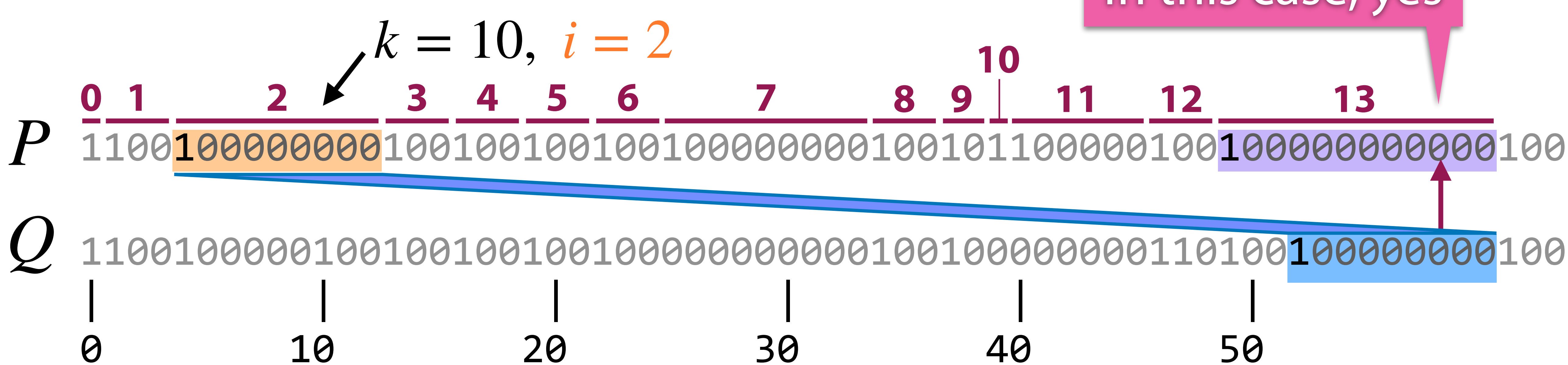
k' (the "new" k) equals 58.

What is i'?

Is it $M[i]\,.\,\xi$?

run	p	ℓ	π	ξ
0	0	1	46	12
1	1	3	34	8
2	**4**	**9**	**52**	**13**
3	13	3	19	5
⋮	⋮	⋮	⋮	⋮

In this case, yes



Move structure queries

$$k' \leftarrow M[i]\,.\,\pi \;+\; (k - M[i]\,.\,p)$$

$\langle\, p,$	$\ell\,,$	$\pi,$	$\xi\,\rangle$
0	1	46	12
1	3	34	8
4	9	52	13
13	3	19	5
16	3	1	1
19	3	16	4
22	3	61	14
25	9	37	9
34	3	49	13
37	2	47	12
39	1	0	0
40	6	4	2
⋮	⋮	⋮	⋮

New example

$k = 31,\; i = 7$

P 1100100000000100100100100100000000100101100000100100000000000100

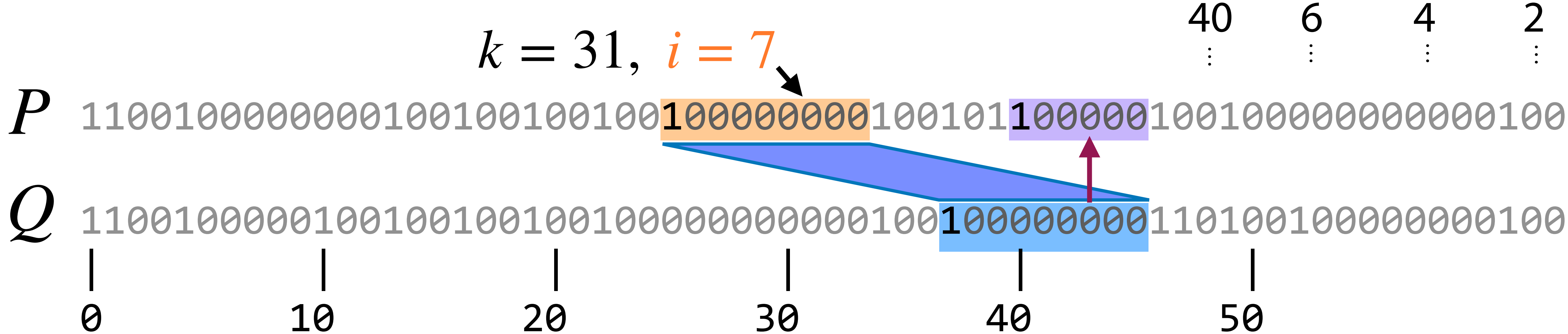
Q 1100100000100100100100100000000000100100000000110100100000000100

0 10 20 30 40 50

Move structure queries

$$k' \leftarrow M[i]\,.\,\pi \;+\; (k - M[i]\,.\,p) = 37 + 6 = 43$$

$\langle\, p,$	$\ell,$	$\pi,$	$\xi\,\rangle$
0	1	46	12
1	3	34	8
4	9	52	13
13	3	19	5
16	3	1	1
19	3	16	4
22	3	61	14
25	9	37	9
34	3	49	13
37	2	47	12
39	1	0	0
40	6	4	2
⋮	⋮	⋮	⋮



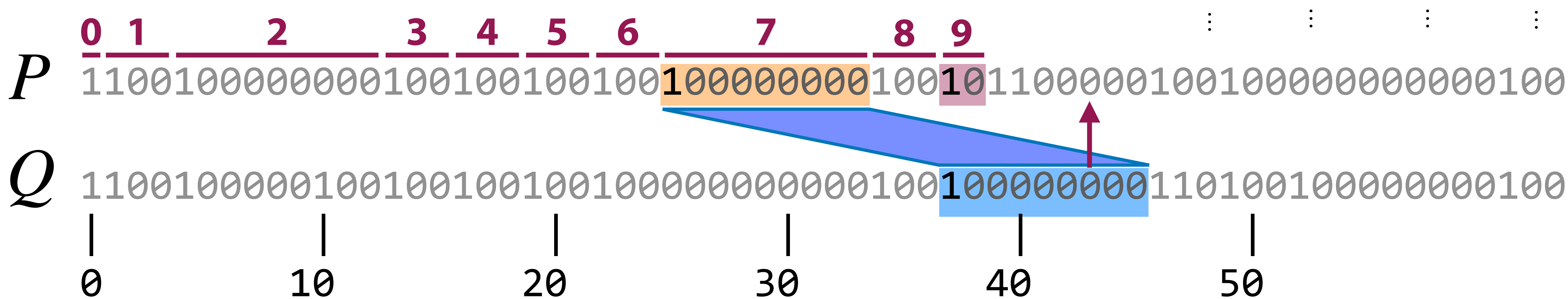
Move structure queries

$$k' \leftarrow M[i]\,.\,\pi \;+\; (k - M[i]\,.\,p) = 37 + 6 = 43$$

$$M[i]\,.\,\xi = 9$$

Run 9 doesn't cover
$k' = 43$; Just $\{37,38\}$!

$\langle\, p,$	$\ell\,,$	$\pi,$	$\xi\,\rangle$
0	1	46	12
1	3	34	8
4	9	52	13
13	3	19	5
16	3	1	1
19	3	16	4
22	3	61	14
25	9	37	9
34	3	49	13
37	2	47	12
39	1	0	0
40	6	4	2
⋮	⋮	⋮	⋮



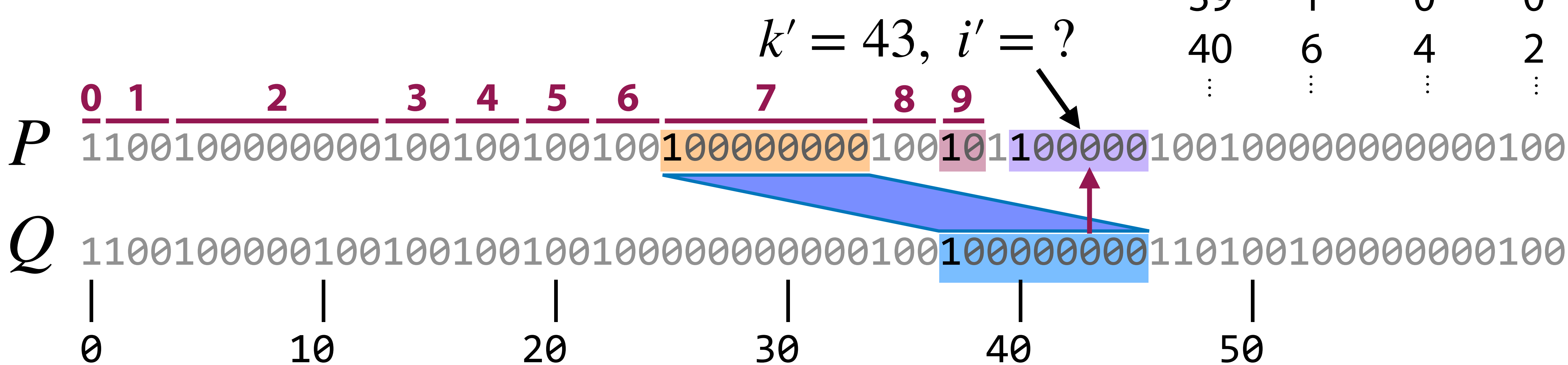
Move structure queries

$$k' \leftarrow M[i].\pi + (k - M[i].p) = 37 + 6 = 43$$

$M[i].\xi = 9$

Run 9 doesn't cover $k' = 43$; Just $\{37,38\}$!

$\langle p,$	$\ell,$	$\pi,$	$\xi \rangle$
0	1	46	12
1	3	34	8
4	9	52	13
13	3	19	5
16	3	1	1
19	3	16	4
22	3	61	14
25	9	37	9
34	3	49	13
37	2	47	12
39	1	0	0
40	6	4	2
⋮	⋮	⋮	⋮



Move structure queries

$$k' \leftarrow M[i].\pi + (k - M[i].p) = 37 + 6 = 43$$

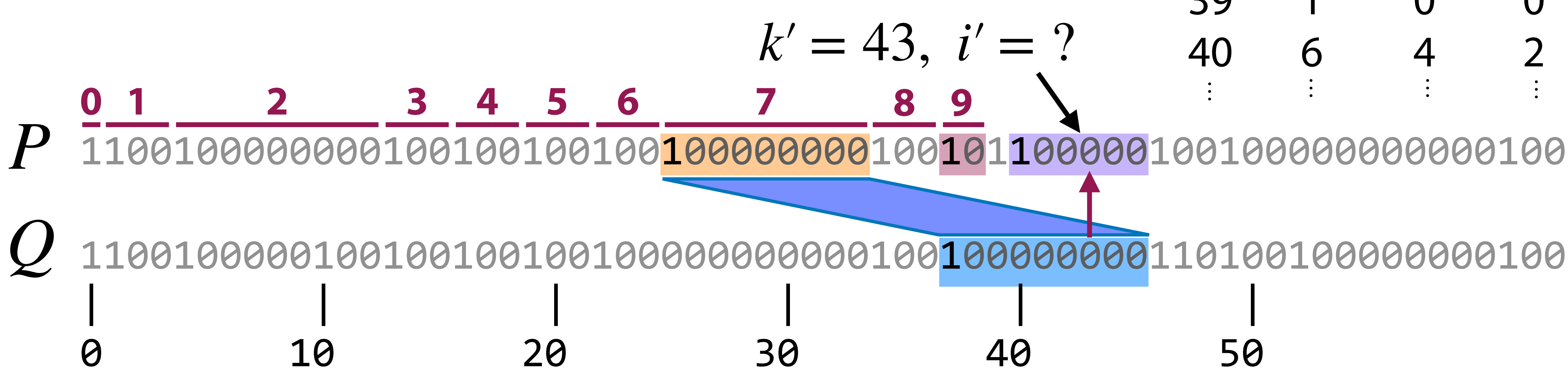
$$M[i].\xi = 9$$

Run 9 doesn't cover $k' = 43$; Just $\{37, 38\}$!

$\langle p,$	$\ell,$	$\pi,$	$\xi \rangle$
0	1	46	12
1	3	34	8
4	9	52	13
13	3	19	5
16	3	1	1
19	3	16	4
22	3	61	14
25	9	37	9
34	3	49	13
37	2	47	12
39	1	0	0
40	6	4	2
⋮	⋮	⋮	⋮

Linear scanning is required

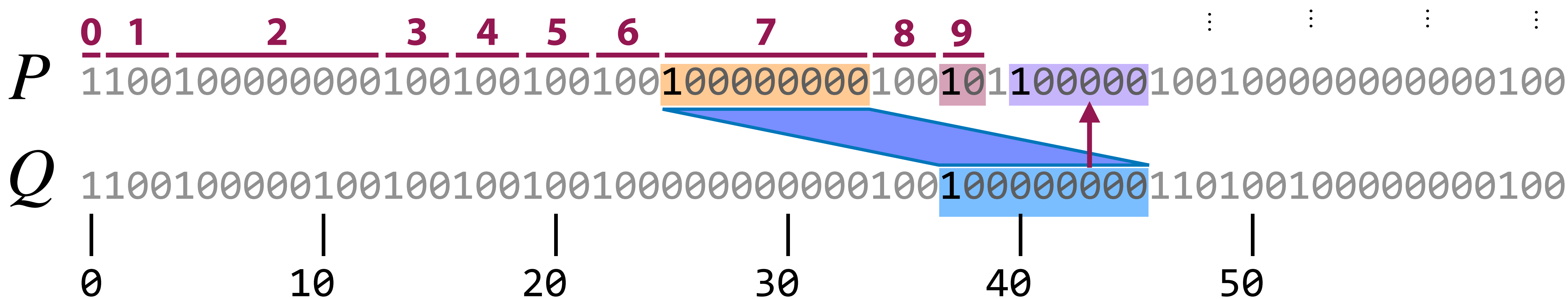
Scan from $M[i].\xi = 9$ until we reach a row containing 43



Move structure queries

Scan from $M[i]\,.\,\xi = 9$ until we reach a row containing 43

$\langle\, p,$	$\ell\,,$	$\pi,$	$\xi\,\rangle$
0	1	46	12
1	3	34	8
4	9	52	13
13	3	19	5
16	3	1	1
19	3	16	4
22	3	61	14
25	9	37	9
34	3	49	13
37	2	47	12
39	1	0	0
40	6	4	2
⋮	⋮	⋮	⋮



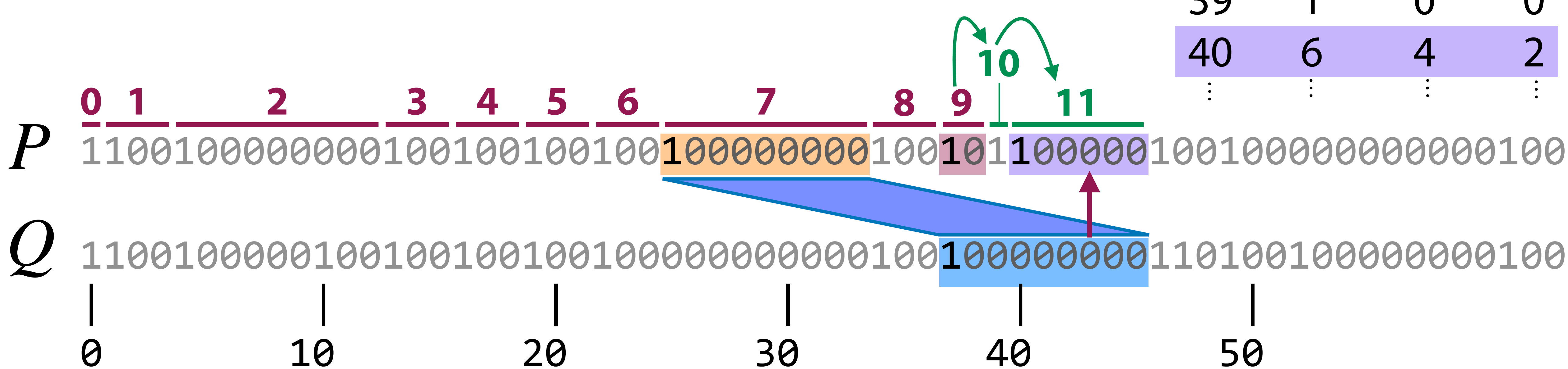
Move structure queries

Scan from $M[i].\xi = 9$ until we reach a row containing 43

Requires 2 scans; correct row is 11

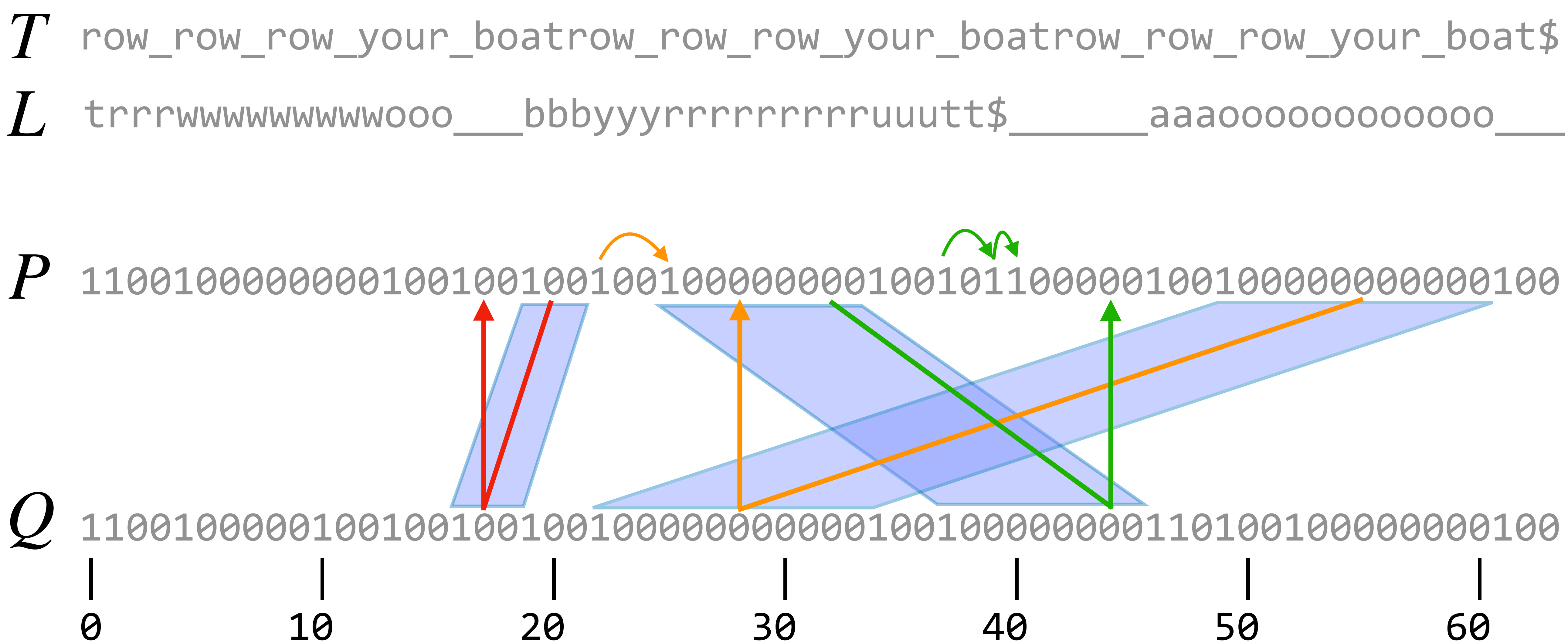
Worst-case LF query time depends on # scans

p,	ℓ,	π,	ξ
0	1	46	12
1	3	34	8
4	9	52	13
13	3	19	5
16	3	1	1
19	3	16	4
22	3	61	14
25	9	37	9
34	3	49	13
37	2	47	12
39	1	0	0
40	6	4	2
⋮	⋮	⋮	⋮



Move structure queries

In our example, computing $\pi(k)$ can take 0, 1 or 2 steps, depending on k

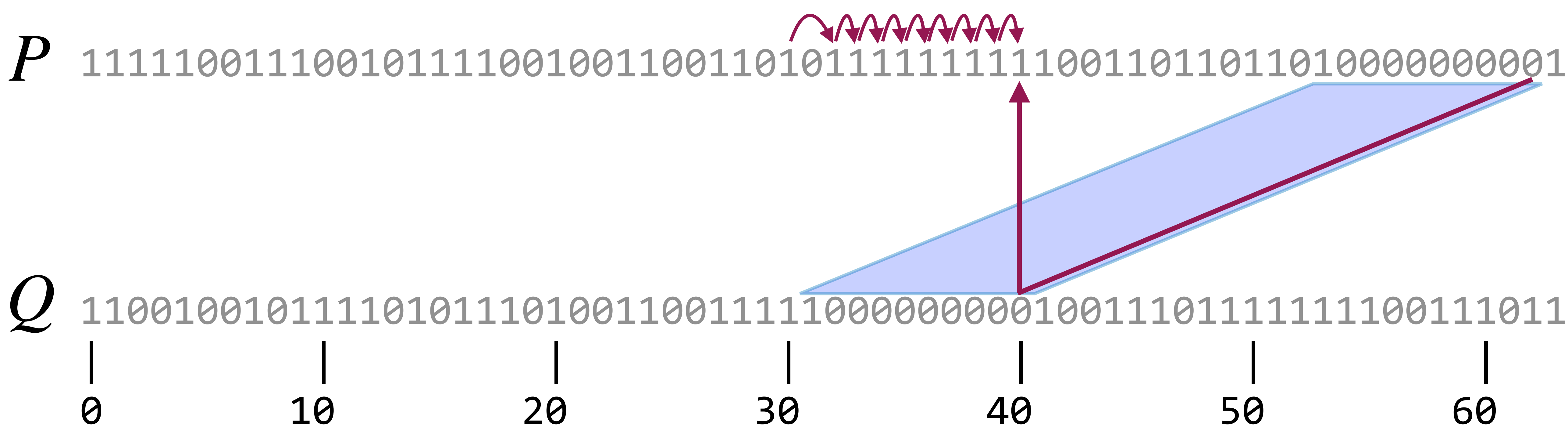


Move structure queries

Here it can take up to 9 steps! Depends on the example…

T row_row_roo__our_boatrou_oou_row__our_boa__ow_oow_ror_your_boat$

L twoarrruw___wwuwrooo___obbbr__rro_y_r_roruuuo__t__$aaoooooooooo_



Move structure queries

p	ℓ	π	ξ
0	1	46	12
1	3	34	8
4	9	52	13
13	3	19	5
16	3	1	1
19	3	16	4
22	3	61	14
25	9	37	9
34	3	49	13
37	2	47	12
39	1	0	0
40	6	4	2
46	3	13	3
49	12	22	2
61	3	10	6

TODO

~~How to query?~~

How to prove $O(1)$ time queries together with $O(r)$?