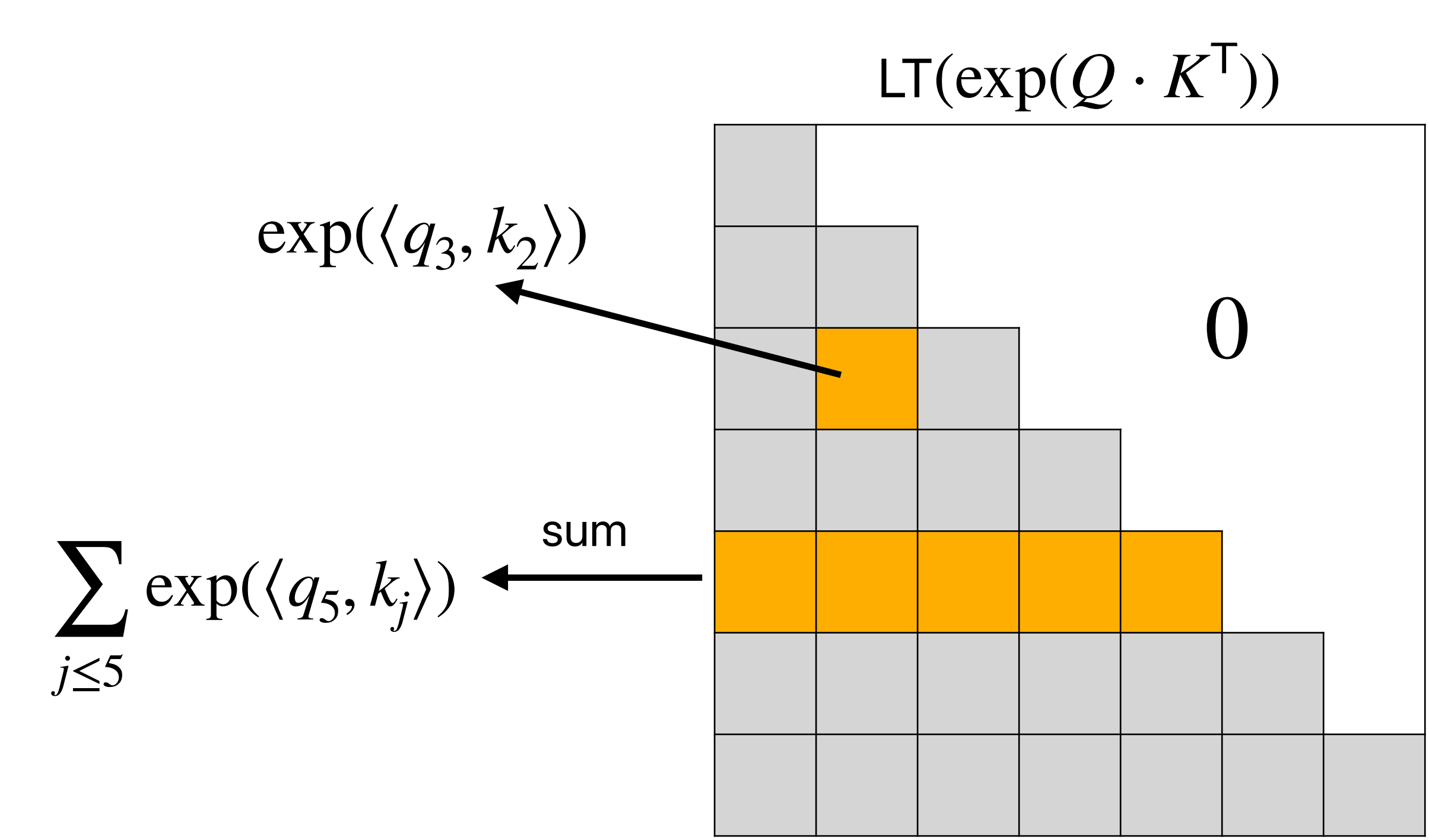
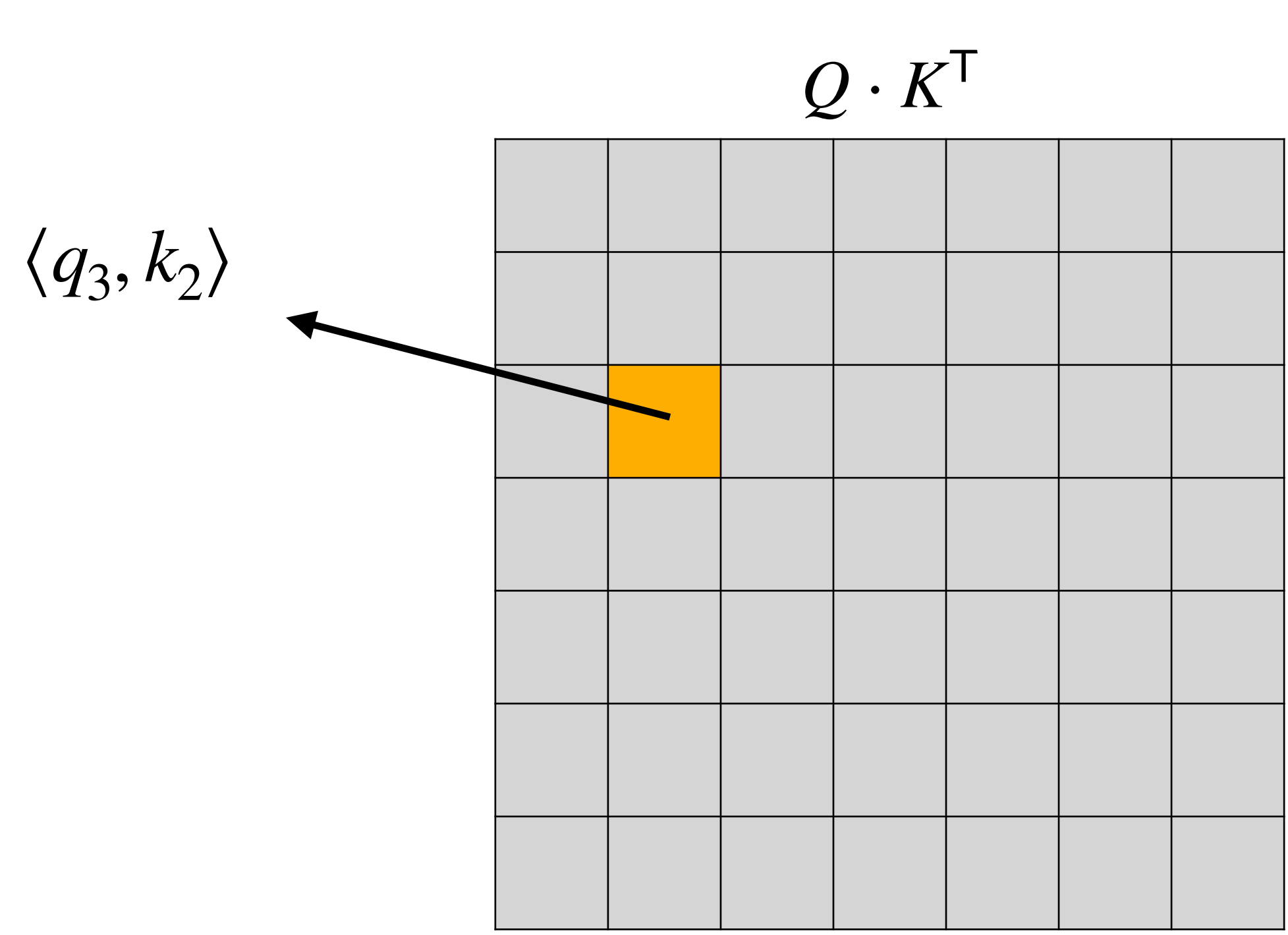
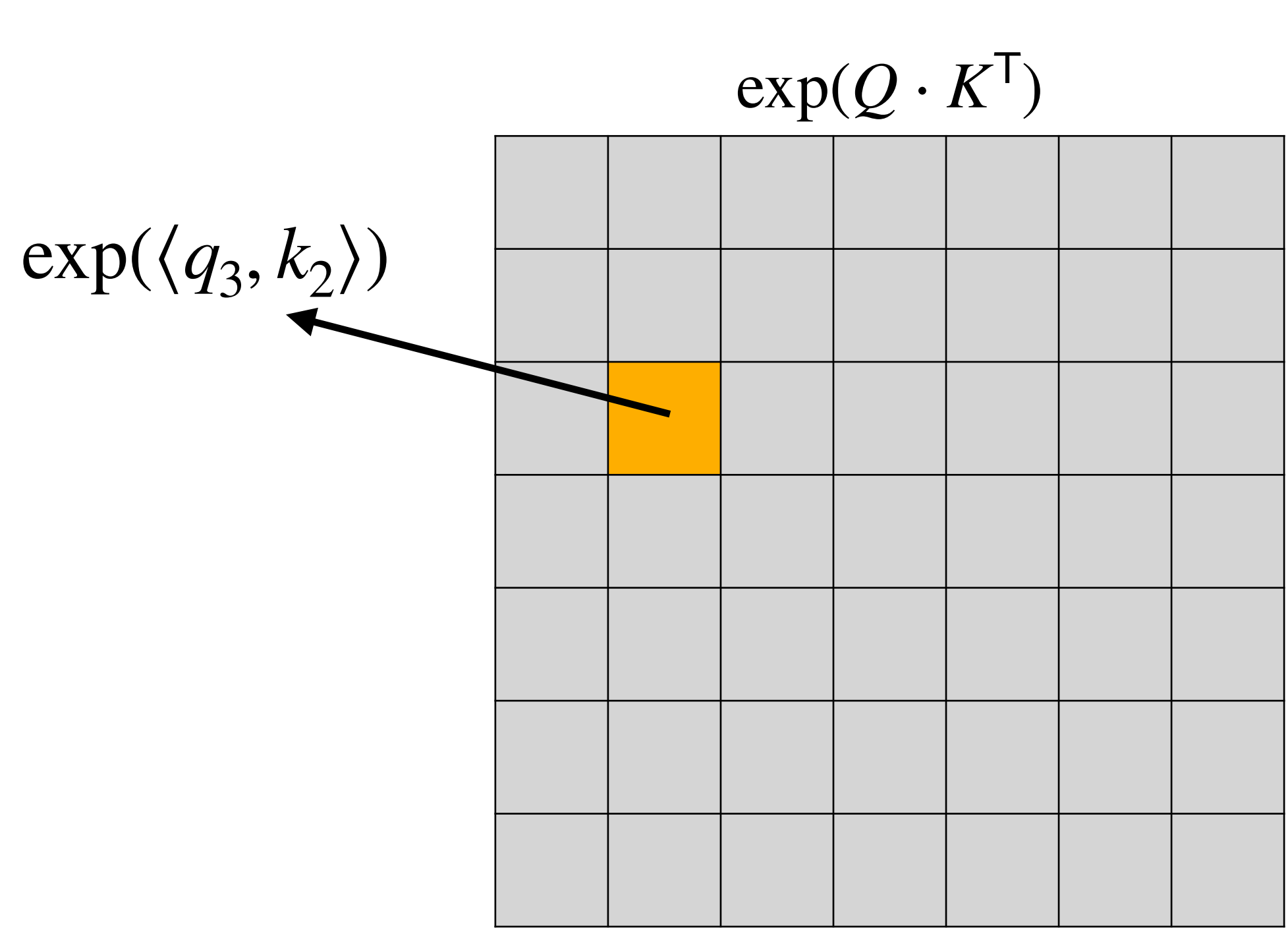
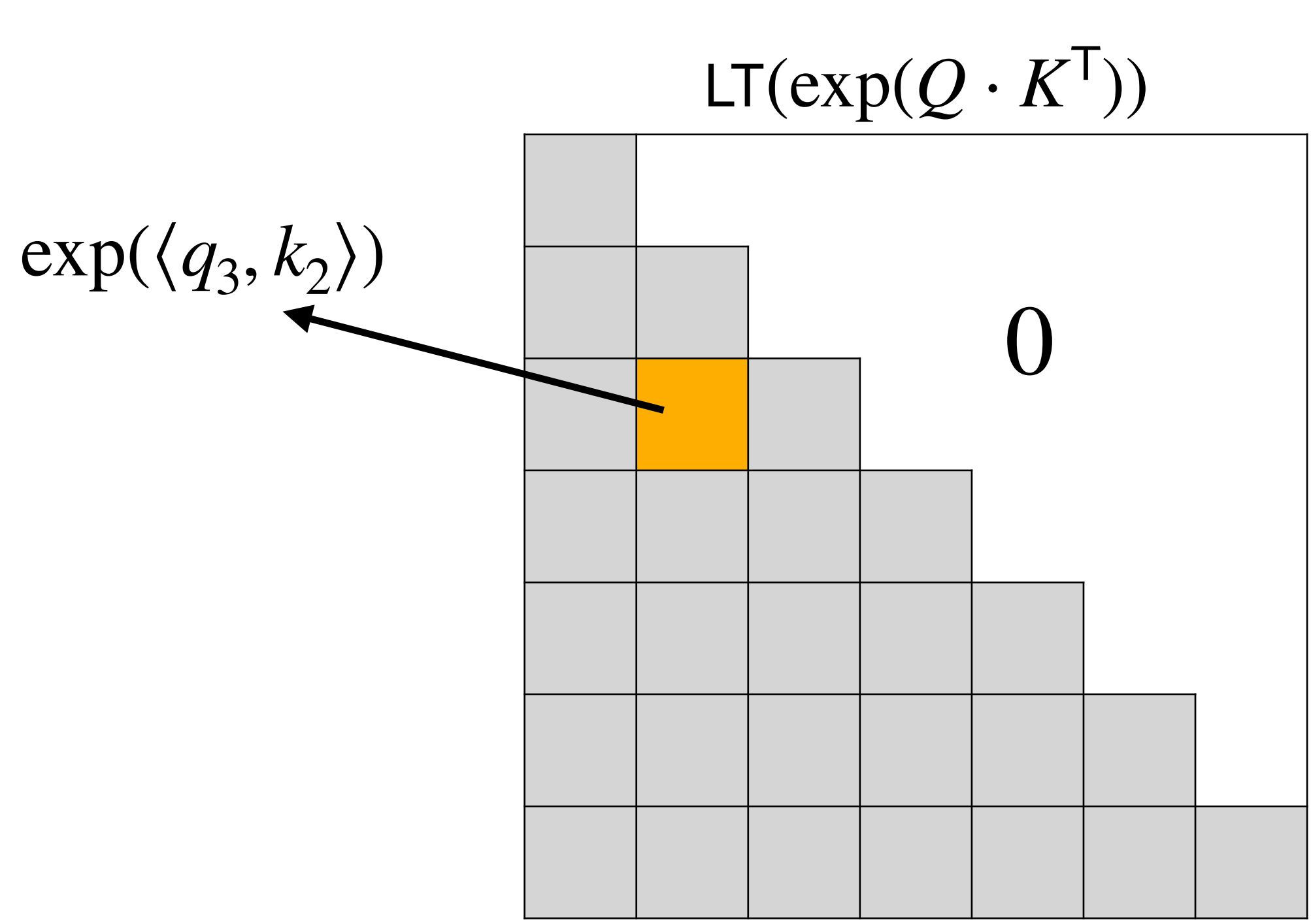

$\mathrm{LT}(\exp(Q \cdot K^{\mathsf{T}}))$
$\exp(\langle q_3, k_2 \rangle)$
0
sum
$\sum_{j \leq 5} \exp(\langle q_5, k_j \rangle)$

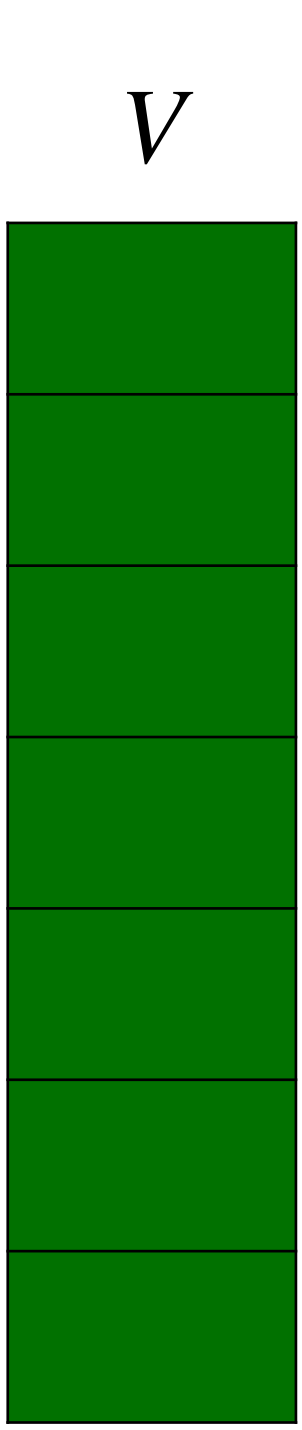

$Q \cdot K^{\mathsf{T}}$
$\langle q_3, k_2 \rangle$

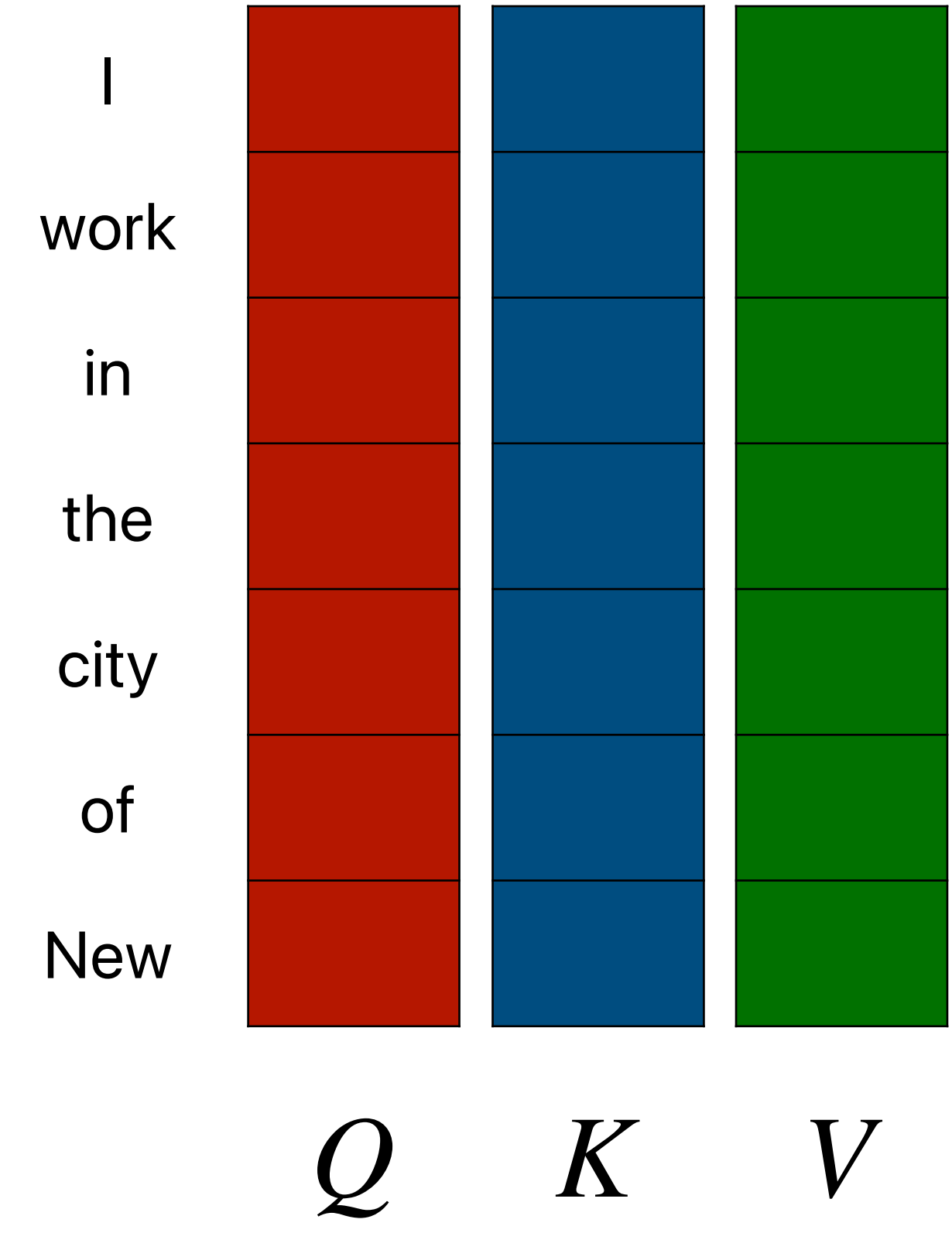

$\exp(Q \cdot K^{\mathsf{T}})$
$\exp(\langle q_3, k_2 \rangle)$


$\mathrm{LT}(\exp(Q \cdot K^{\mathsf{T}}))$
$\exp(\langle q_3, k_2 \rangle)$
0

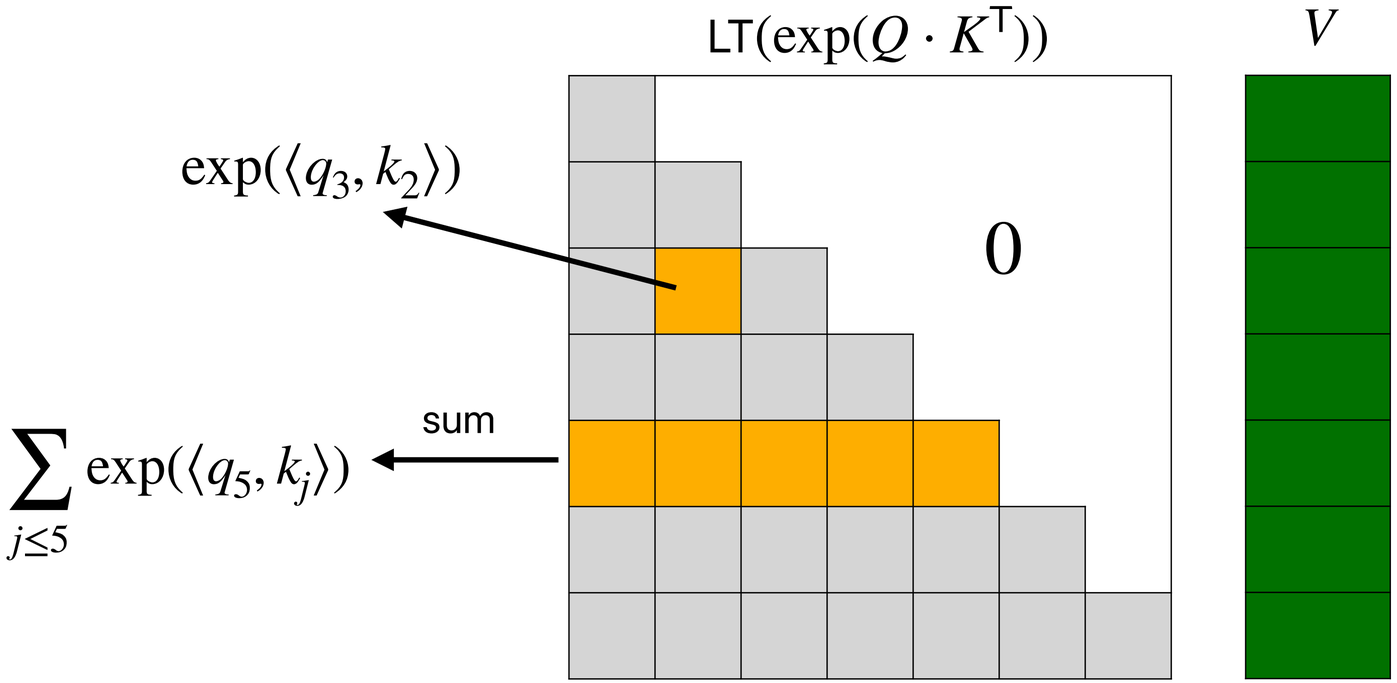
Causal Attention

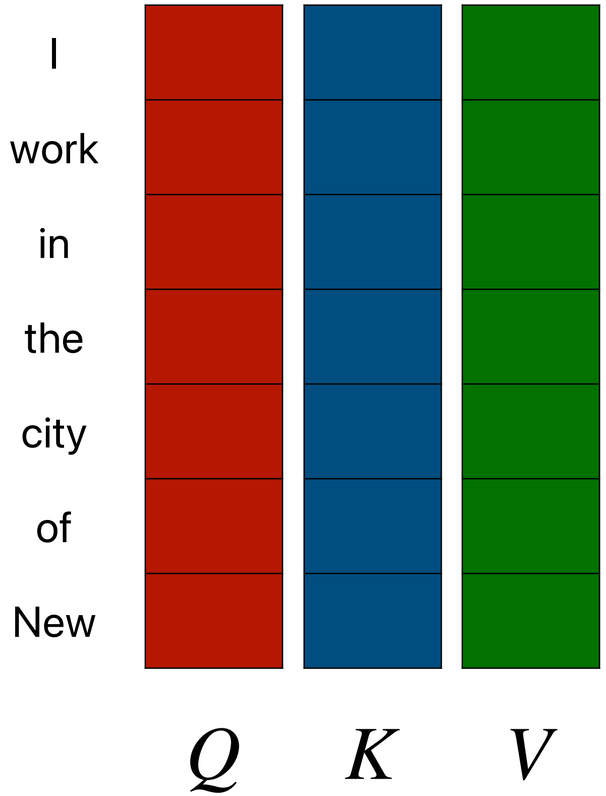
12


V


I
work
in
the
city
of
New
Q
K
V

Causal Attention





Attention Outputs