

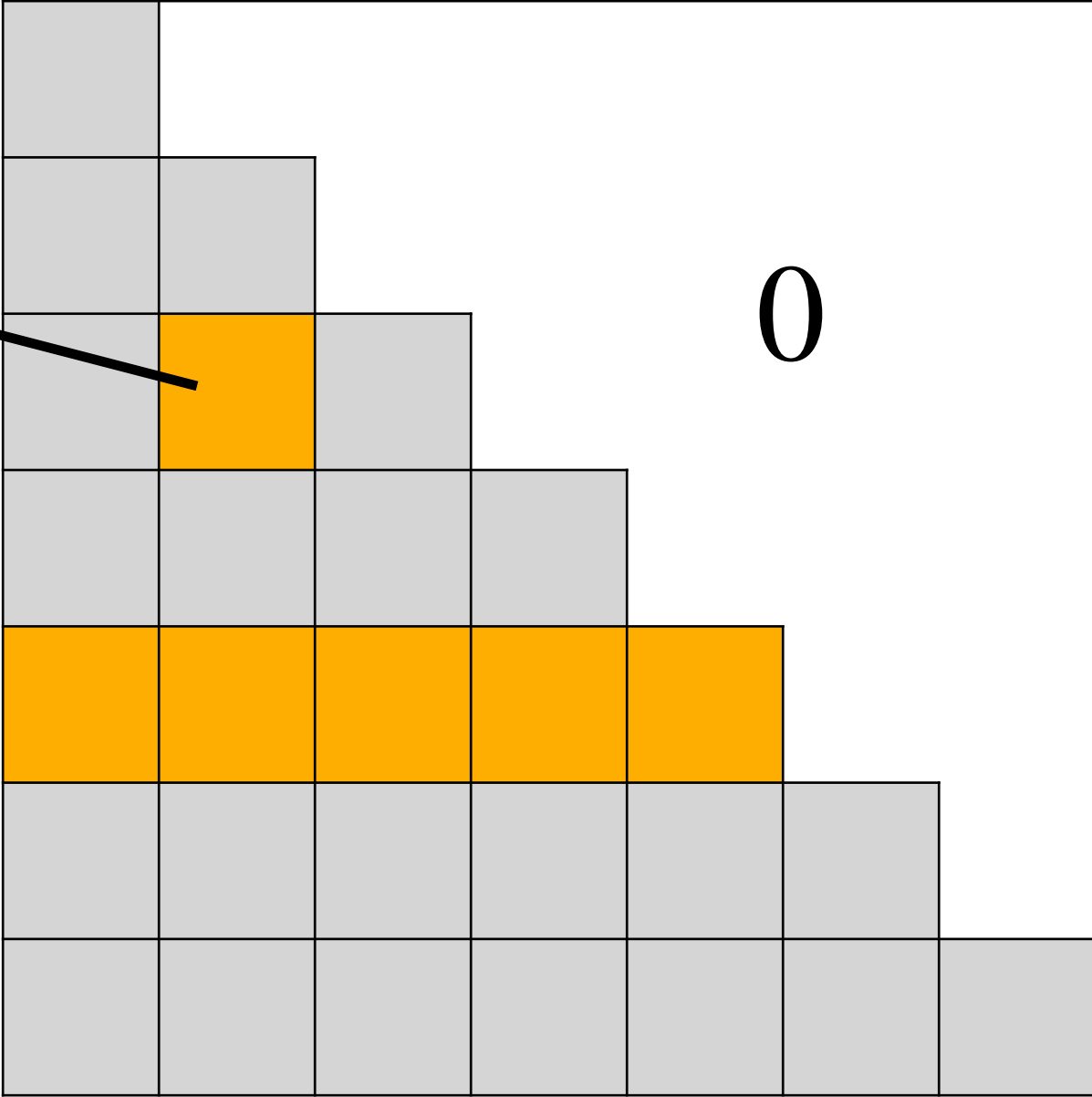
$$\text{LT}(\exp(Q \cdot K^T))$$

$$\exp(\langle q_3, k_2 \rangle)$$

0

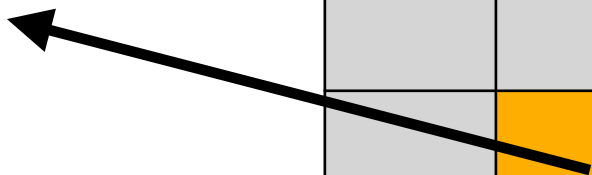
$$\sum_{j \leq 5} \exp(\langle q_5, k_j \rangle)$$

sum



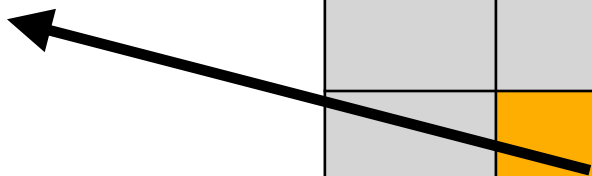
$$Q \cdot K^{\top}$$

$$\langle q_3, k_2 \rangle$$



$$\exp(Q \cdot K^T)$$

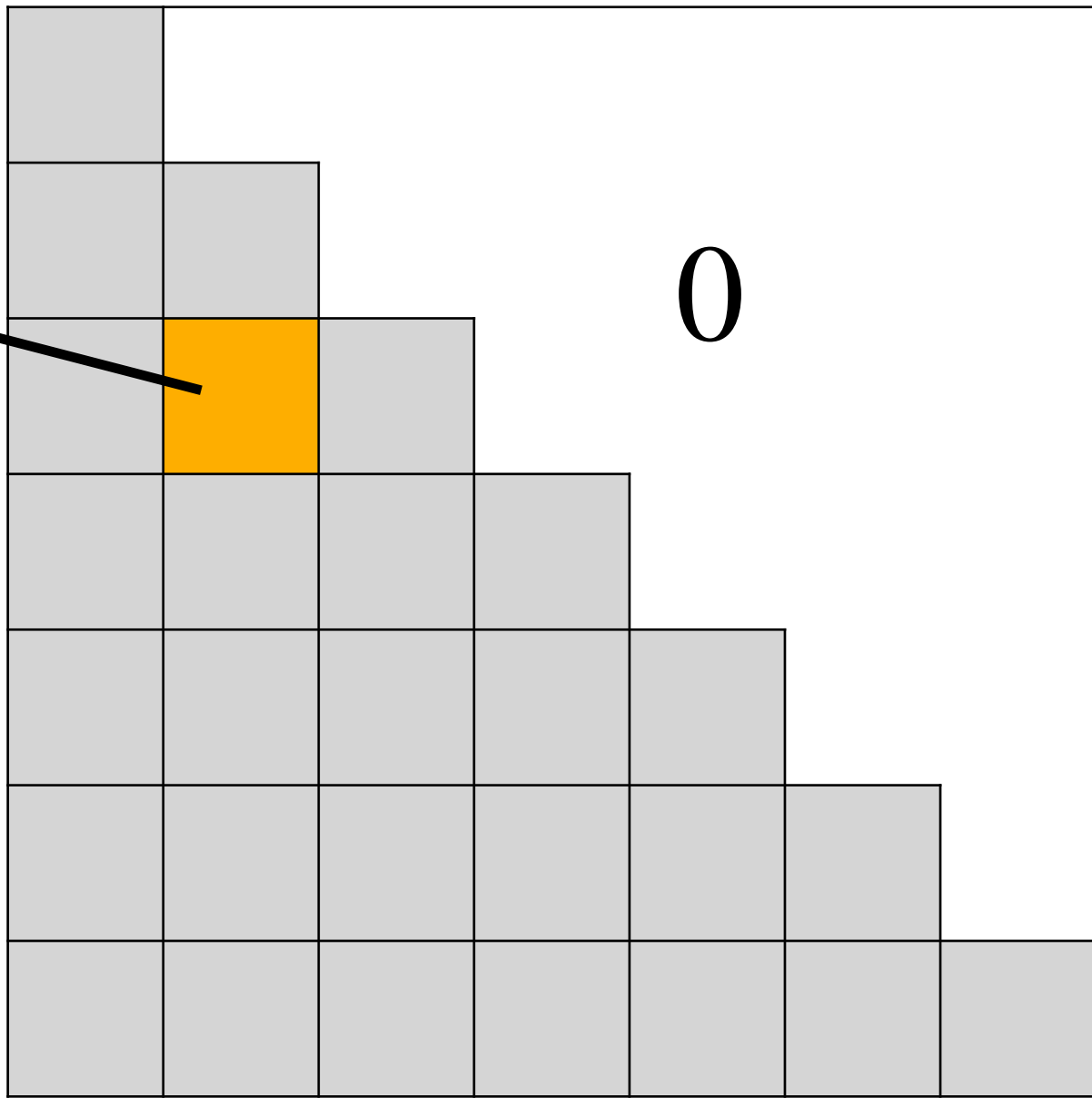
$$\exp(\langle q_3, k_2 \rangle)$$



$$\text{LT}(\exp(Q \cdot K^\top))$$

$$\exp(\langle q_3, k_2 \rangle)$$

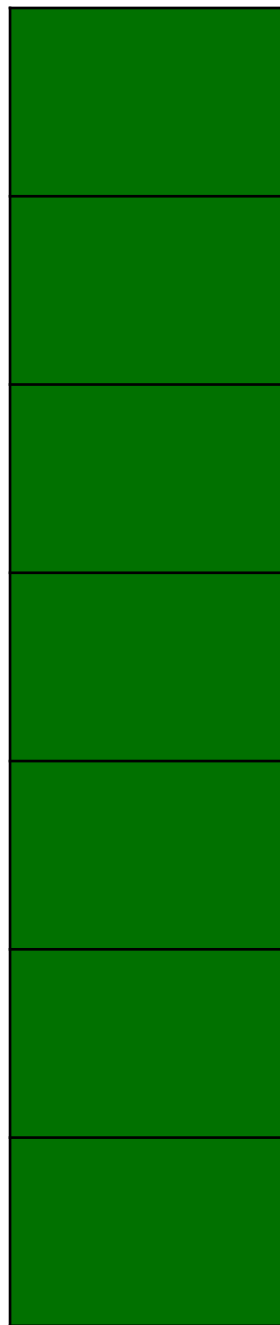
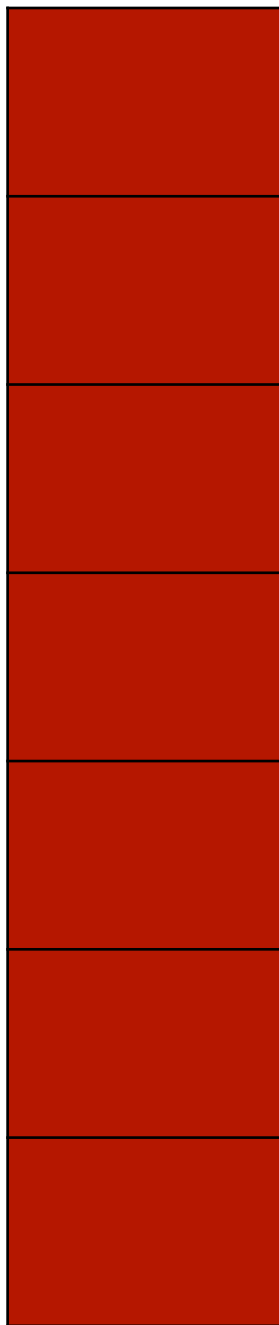
0



Causal Attention



I
work
in
the
city
of
New

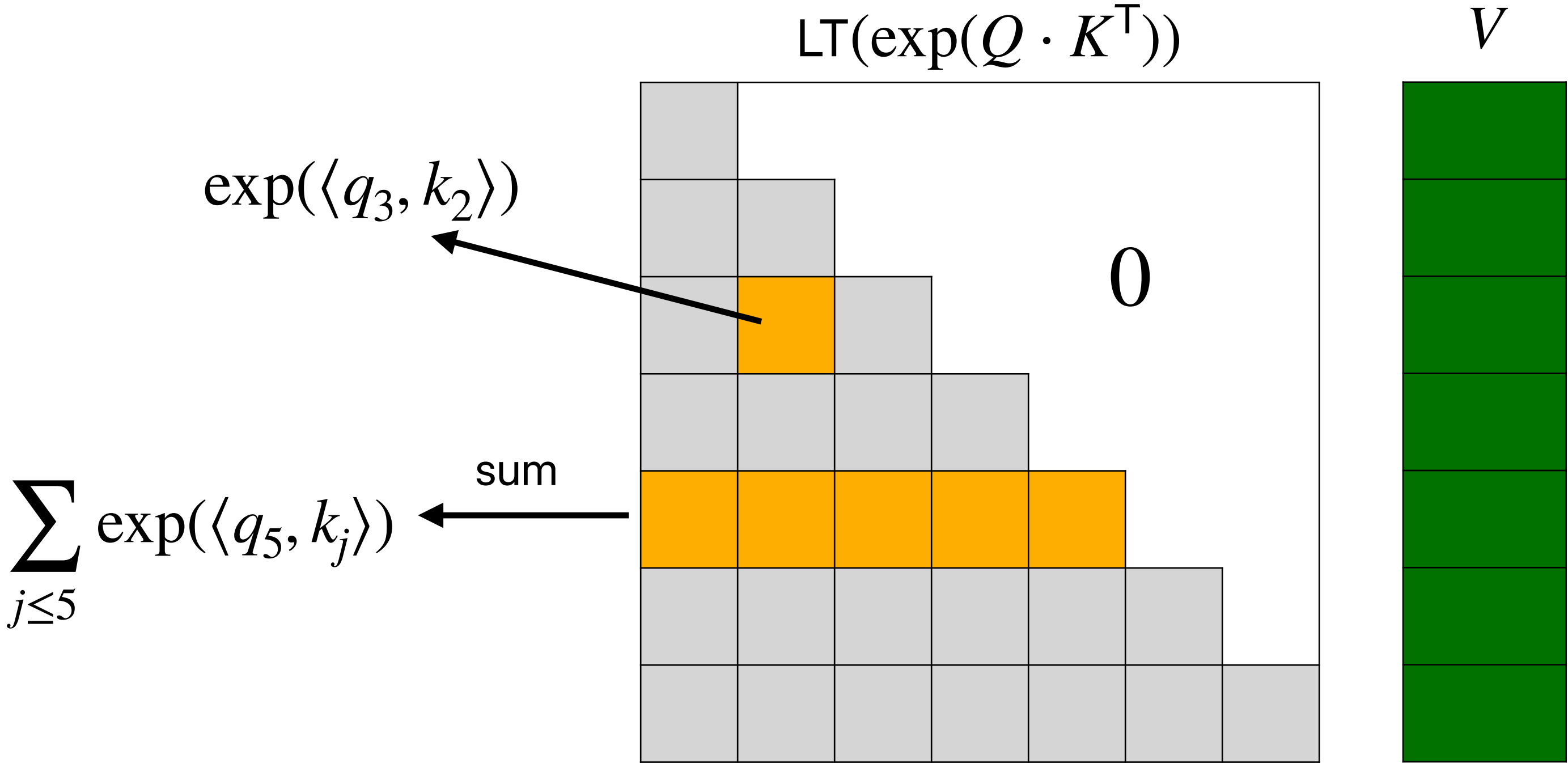


Q

K

V

Causal Attention



Attention Outputs

