Attention

attention
score

$$\alpha_{1,2} = q^1 \cdot k^2 \qquad \alpha_{1,3} = q^1 \cdot k^3 \qquad \alpha_{1,4} = q^1 \cdot k^4$$

q^1 $q^1 = W^q a^1$ k^2 $k^2 = W^k a^2$ k^3 $k^3 = W^k a^3$ k^4 $k^4 = W^k a^4$

a^1 a^2 a^3 a^4

Q   **K**

Attention
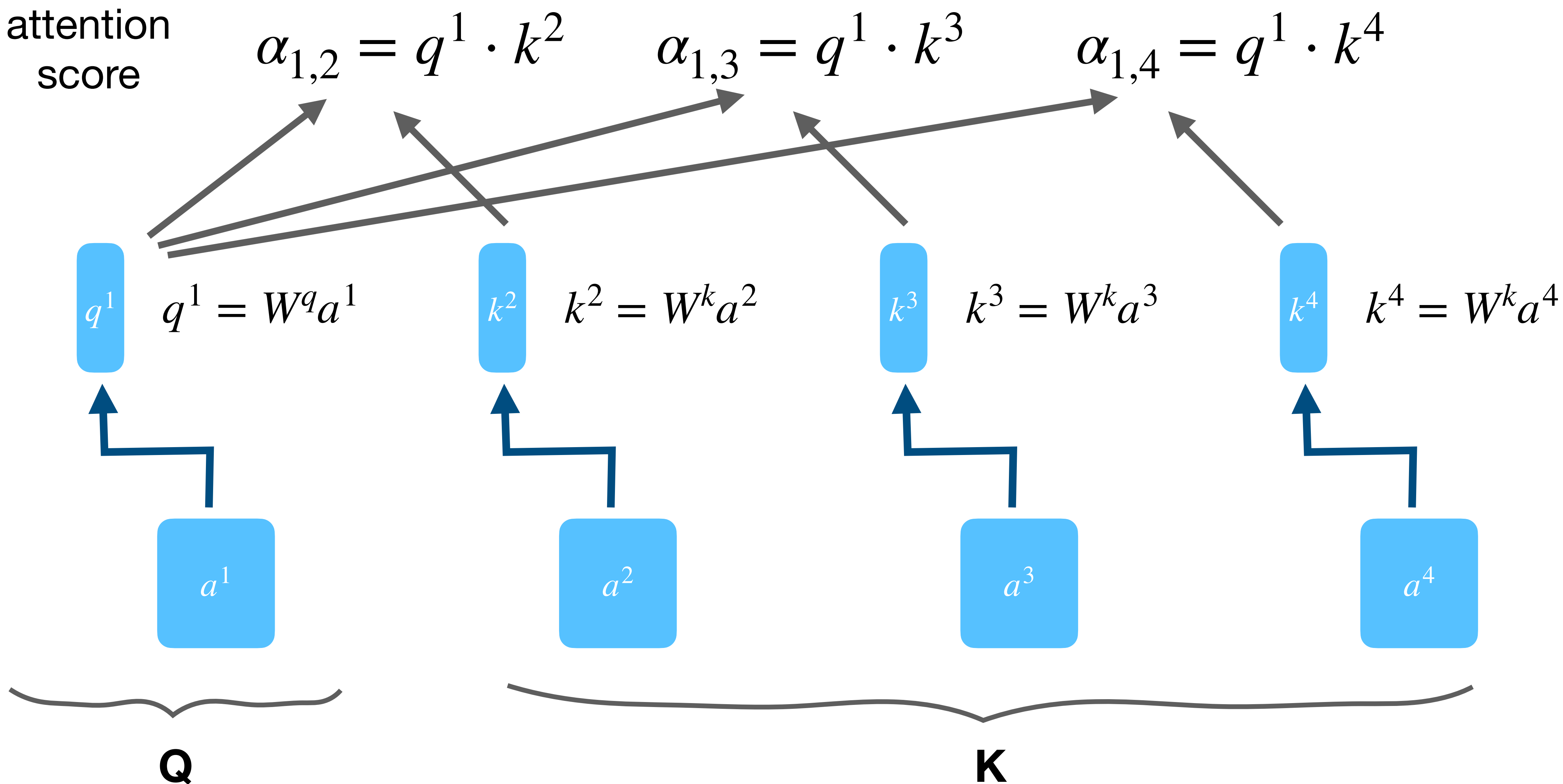
$$\alpha'_{1,i} = \exp(a_{1,i}) / \sum_j \exp(a_{1,j})$$