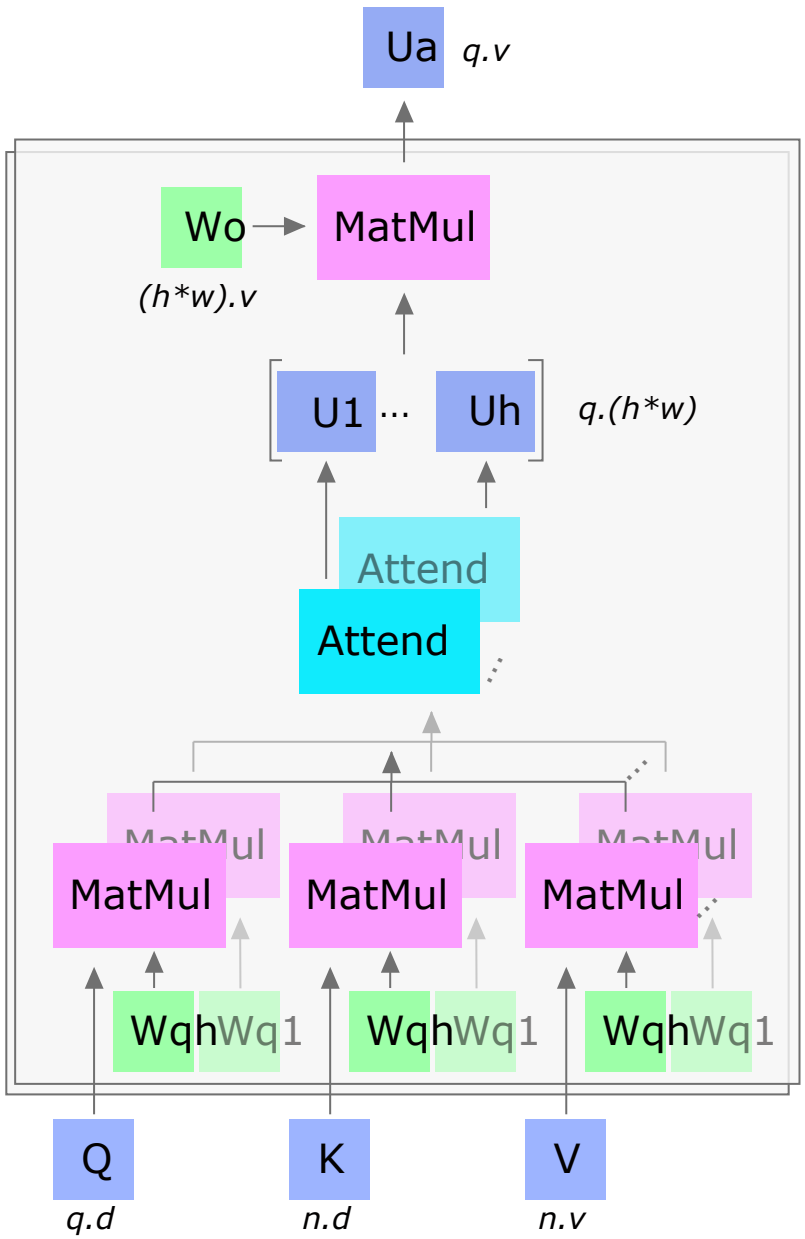
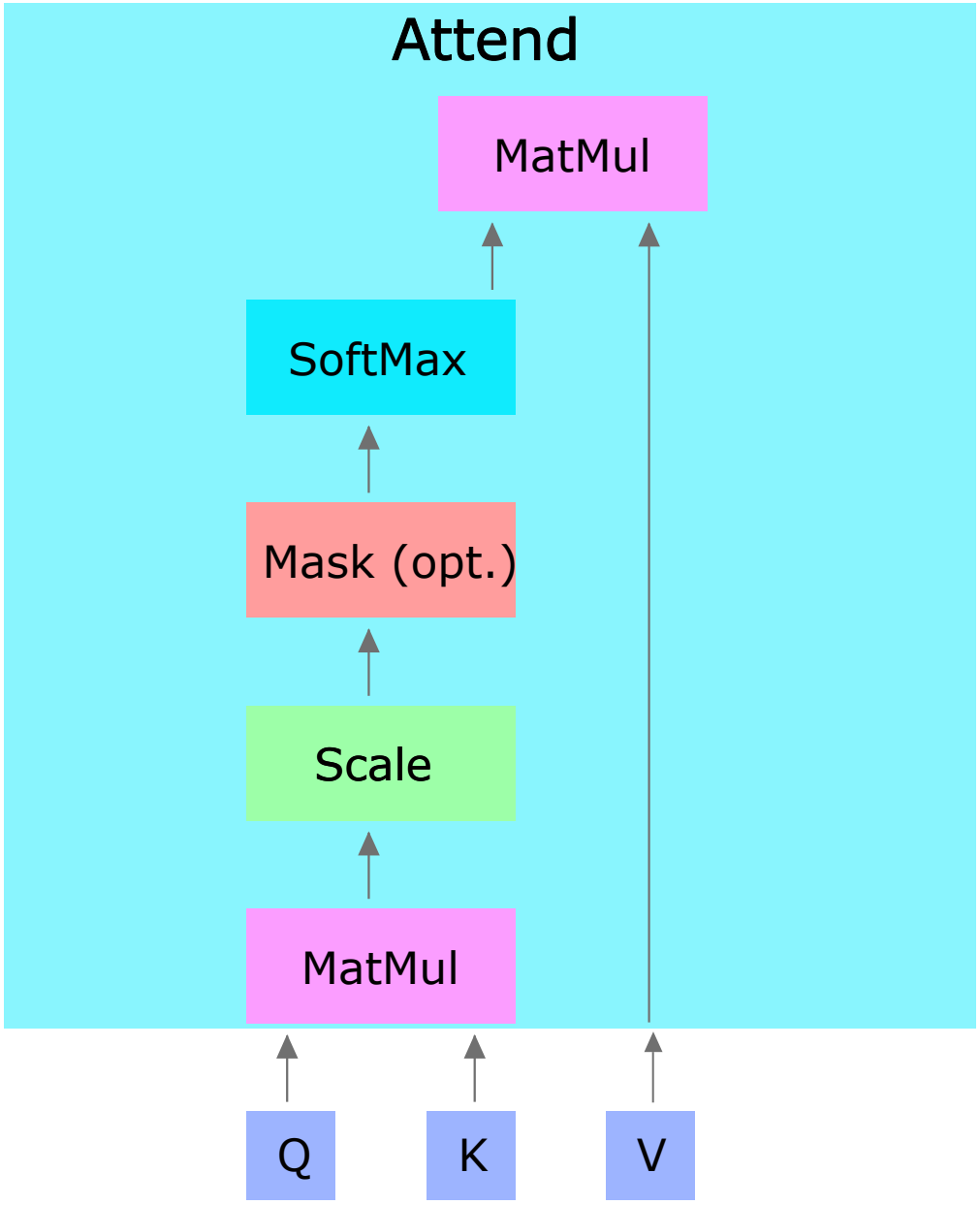

Ua
q.v
Wo
MatMul
(h*w).v
U1
...
Uh
q.(h*w)
Attend
Attend
MatMul
MatMul
MatMul
WqhWq1
WqhWq1
WqhWq1
Q
K
V
q.d
n.d
n.v


Attend
MatMul
SoftMax
Mask (opt.)
Scale
MatMul
Q
K
V