

source code:
`cur = self_attn(j, tokens, d_k, Q, K, V)`
`multi_head_out = concat(multi_head_out, j, tokens, d_k, cur)`

