



$seq$  = sequence length  
 $d_{model}$  = size of the embedding vector  
 $h$  = number of heads  
 $d_k = d_v$  =  $d_{model} / h$