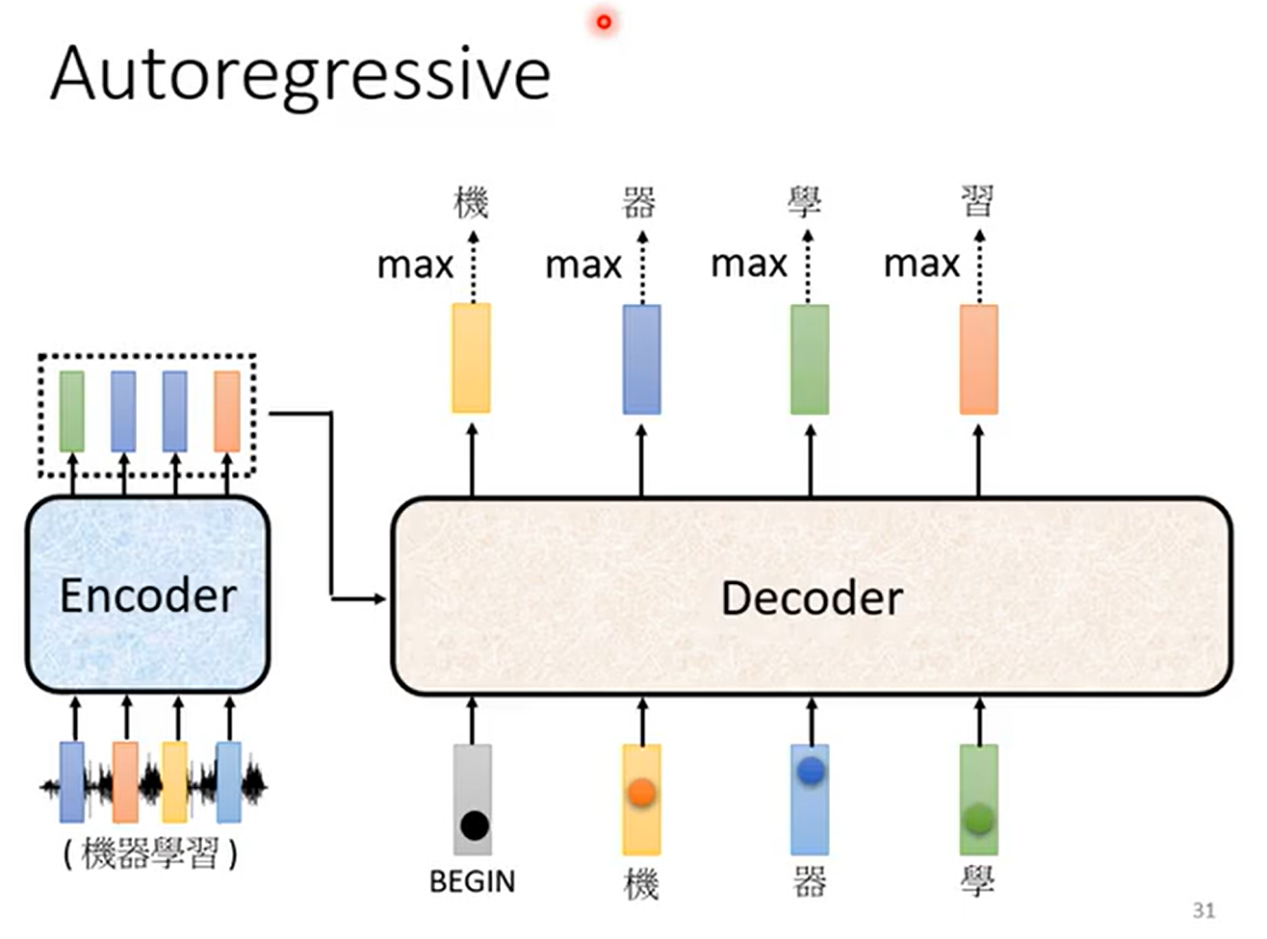
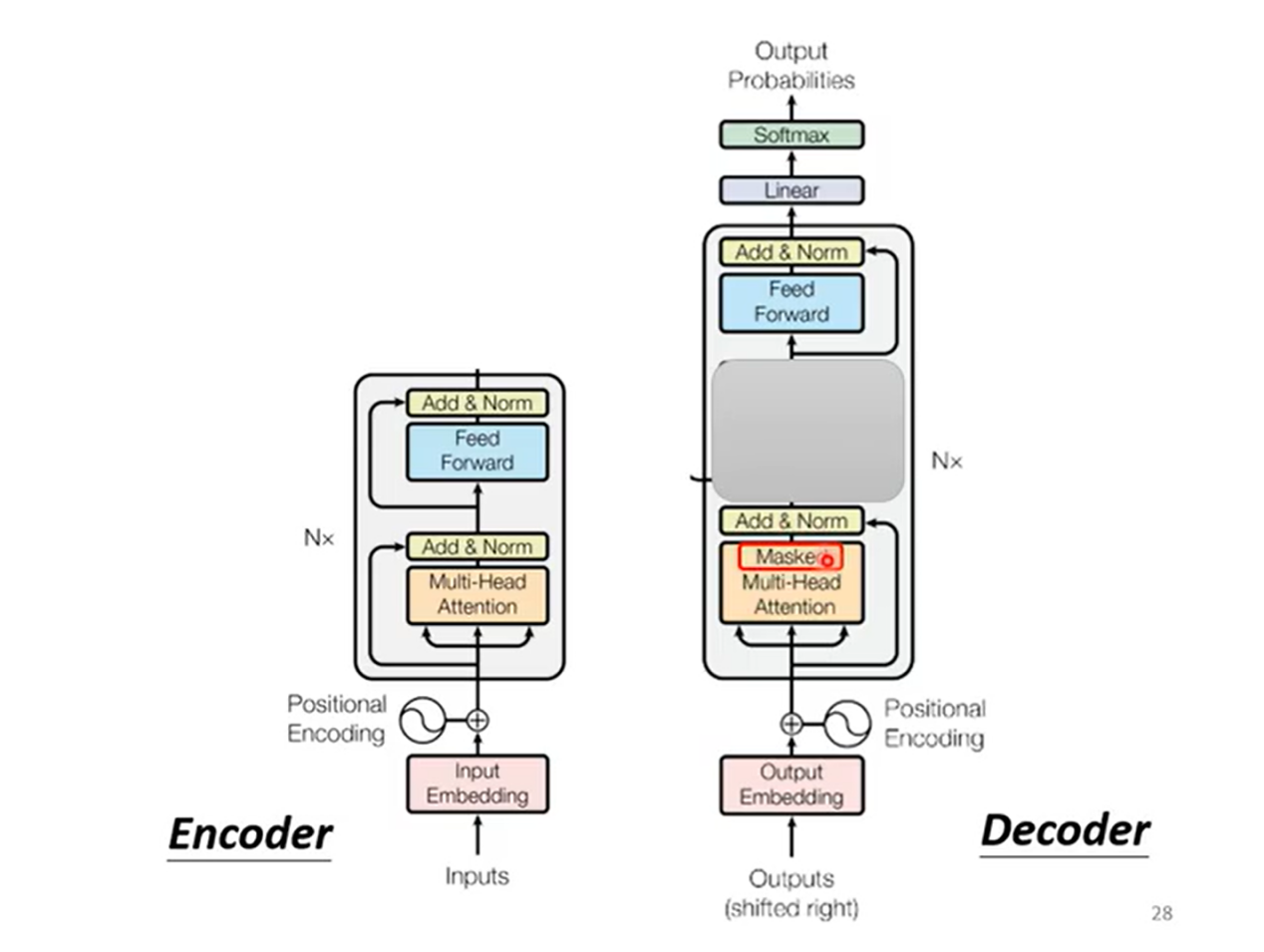
Seqtoseq:

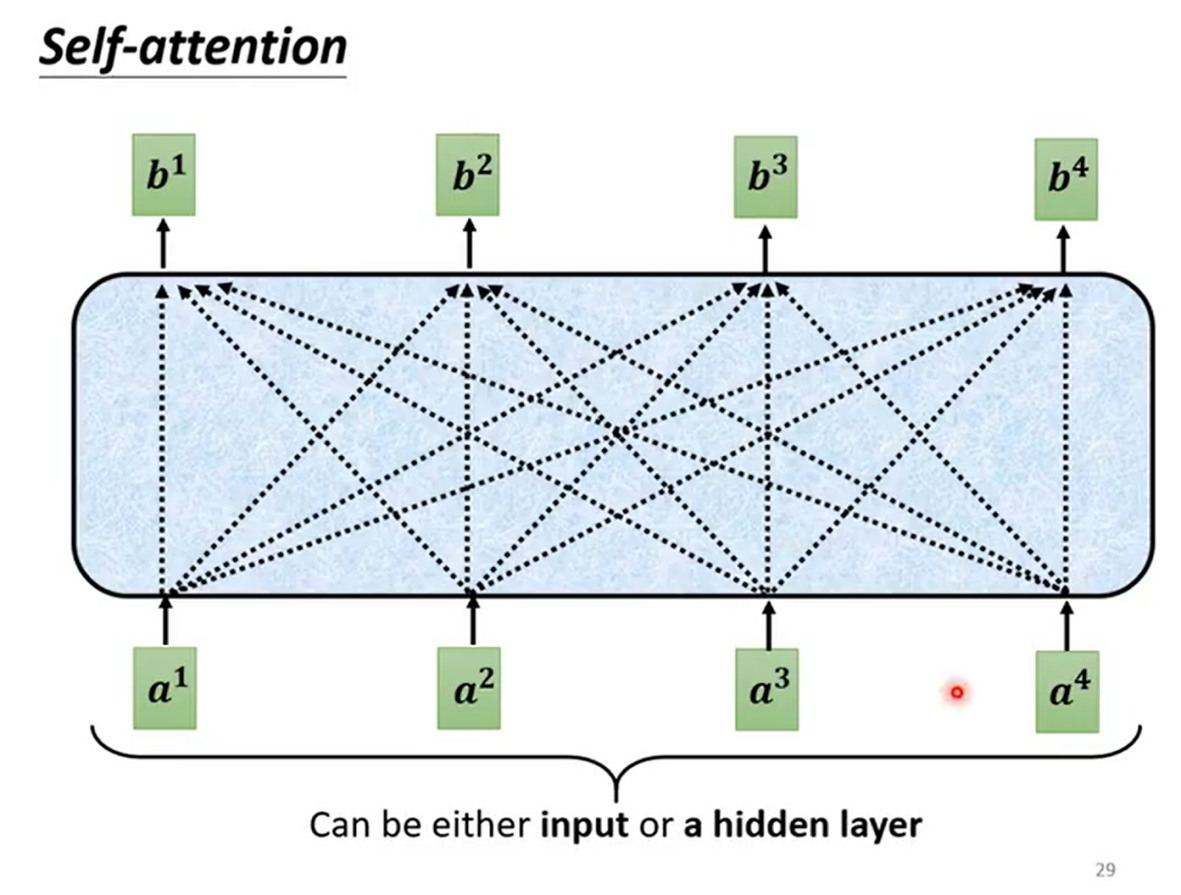
Input--->encoder--->decoder--->output

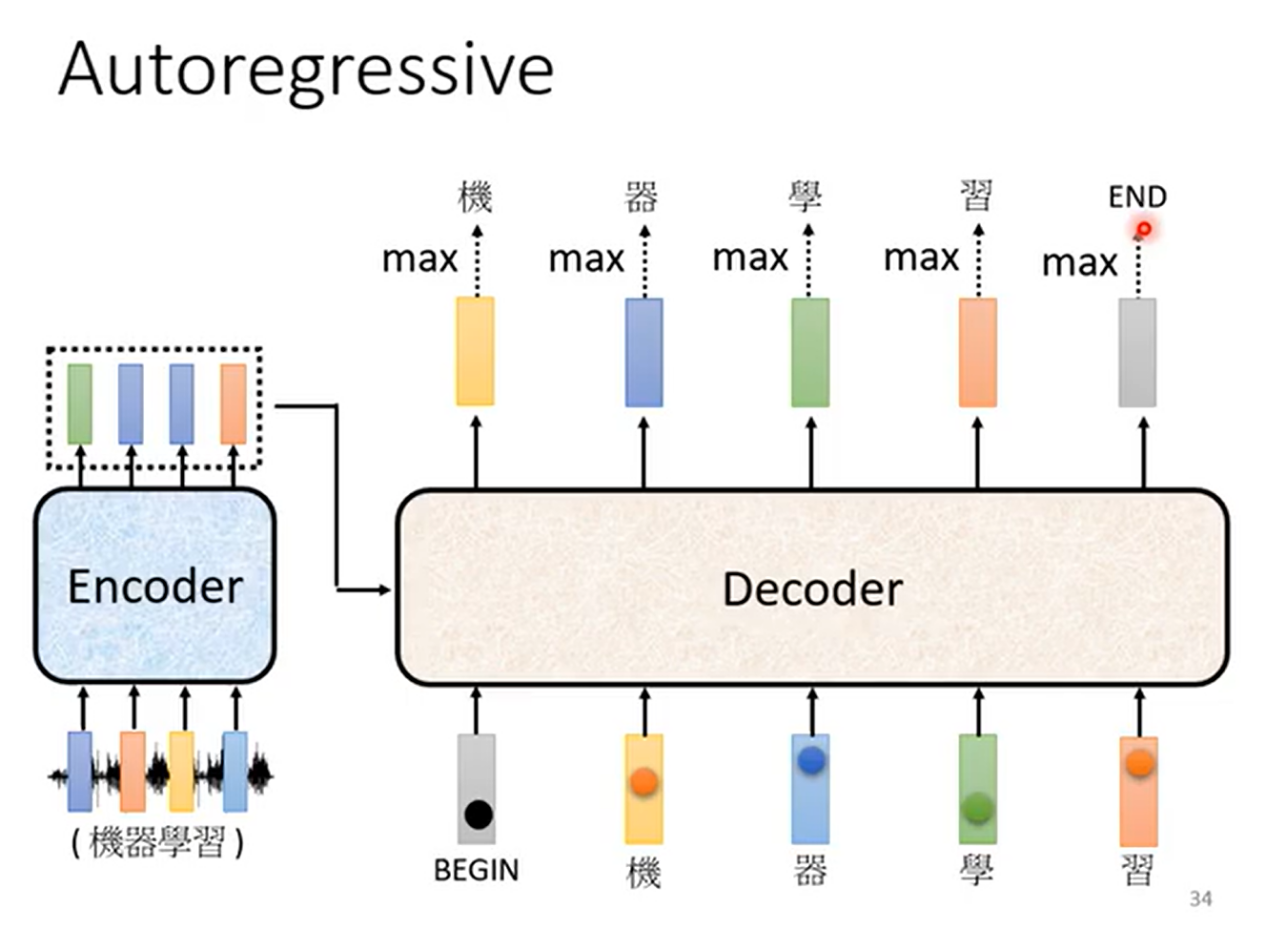
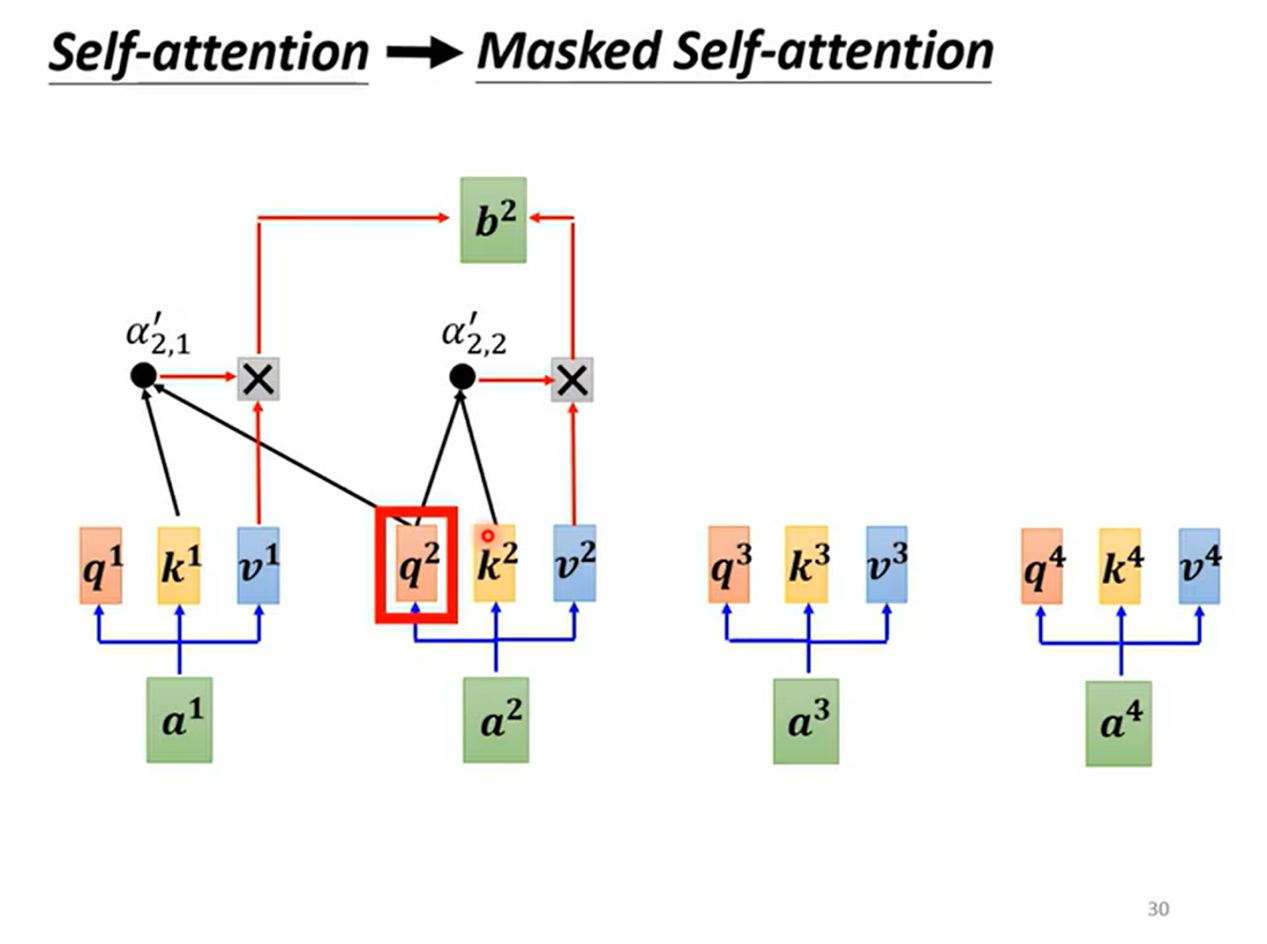
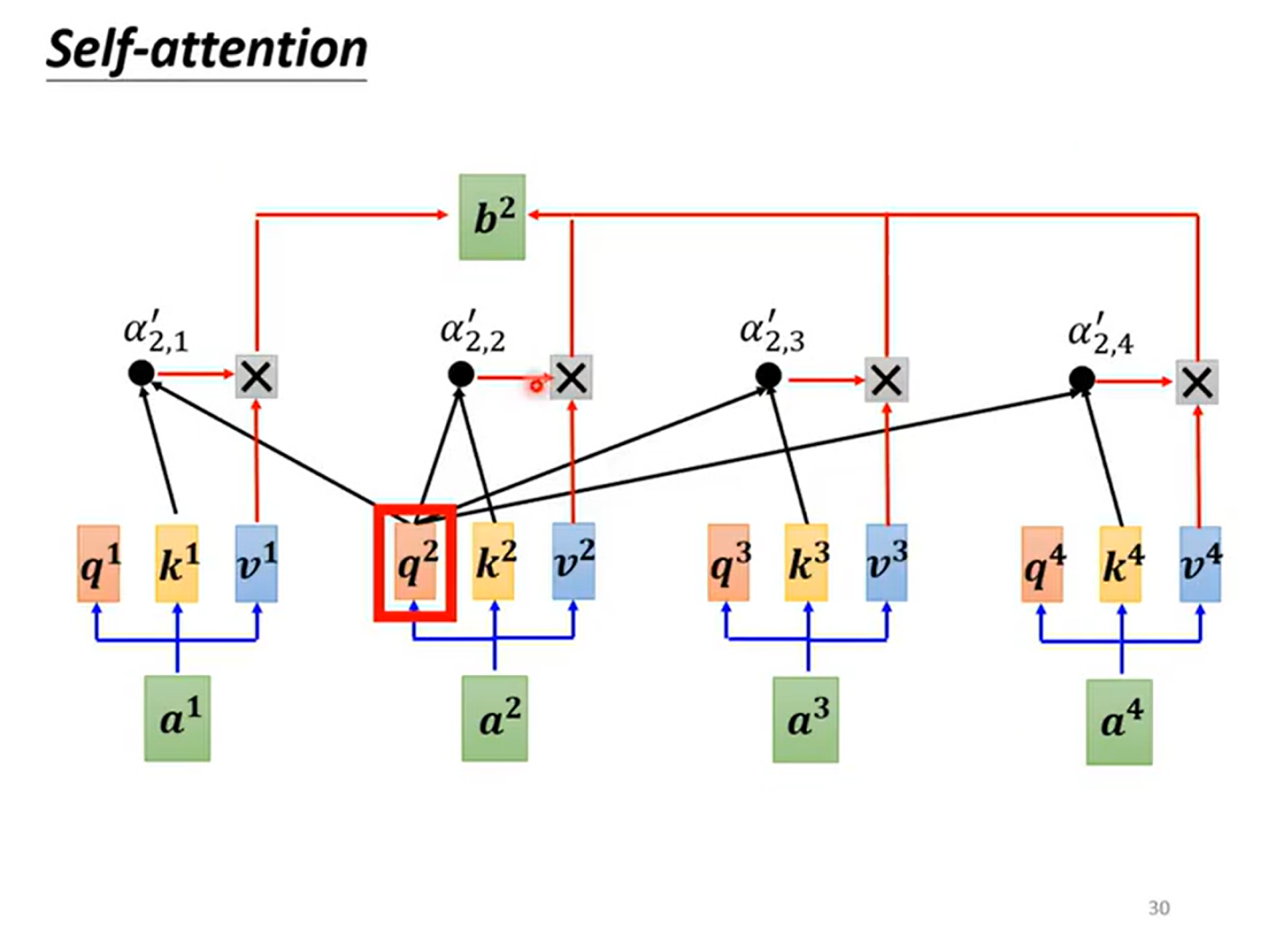
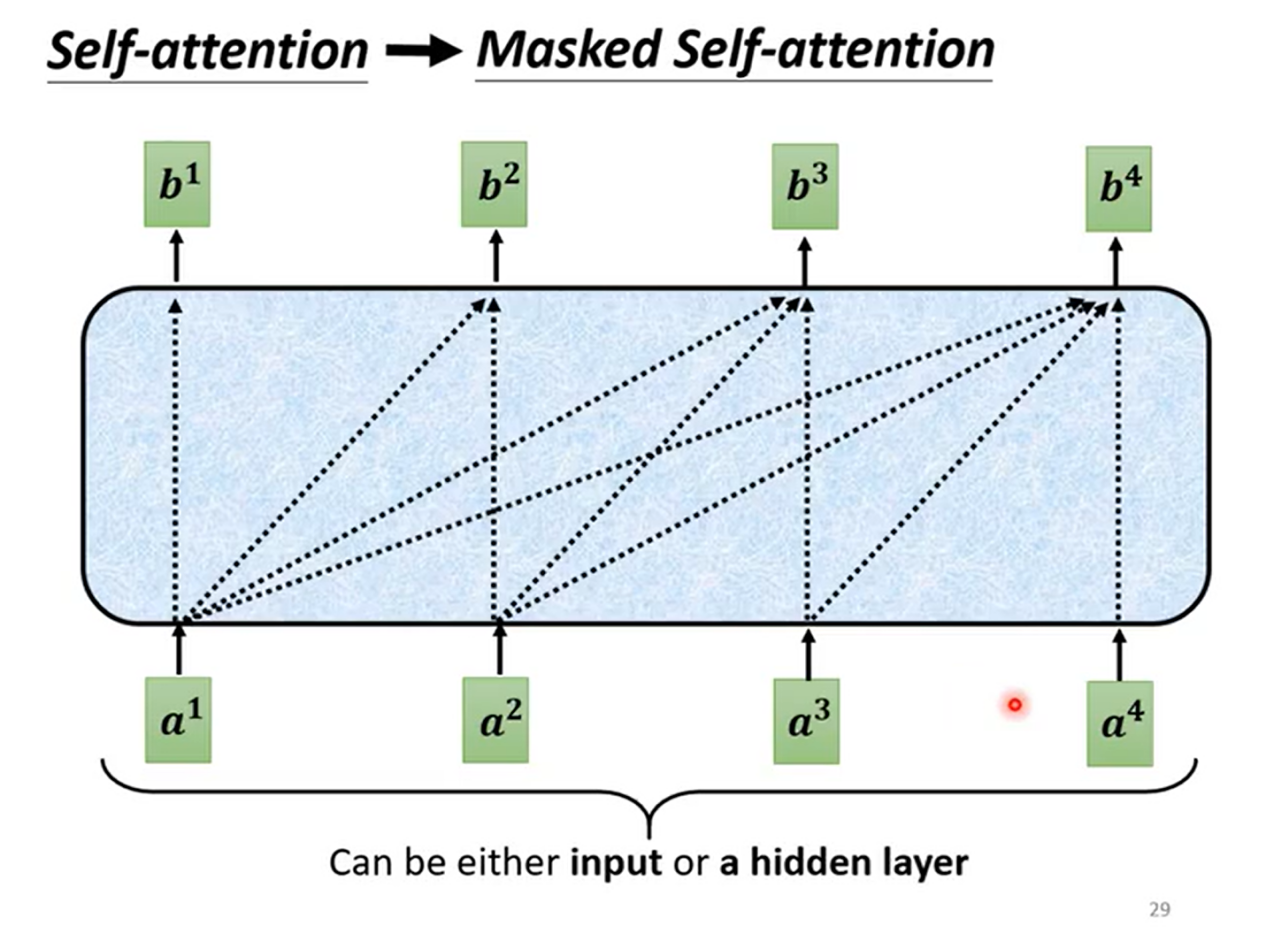


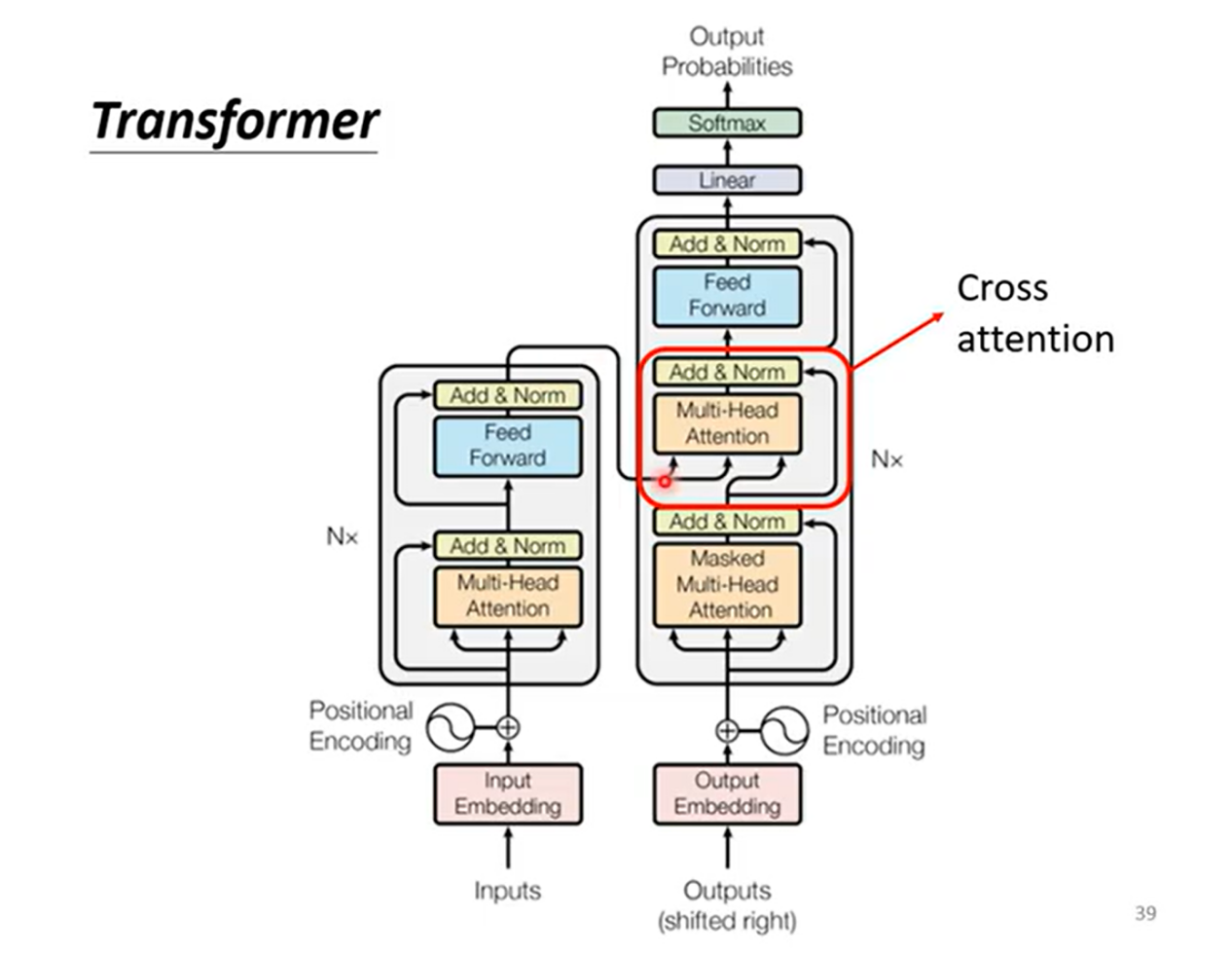
Transformer:



去除decoder里面灰色的部分，encoder和decoder基本没区别, decoder的那个多头注意力的部分多了masked







Cross attention是encoder和decoder的桥梁，两个输入来自encoder，一个输入来自decoder；实际就是用encoder的k,v 和 decoder的q当充作qkv

