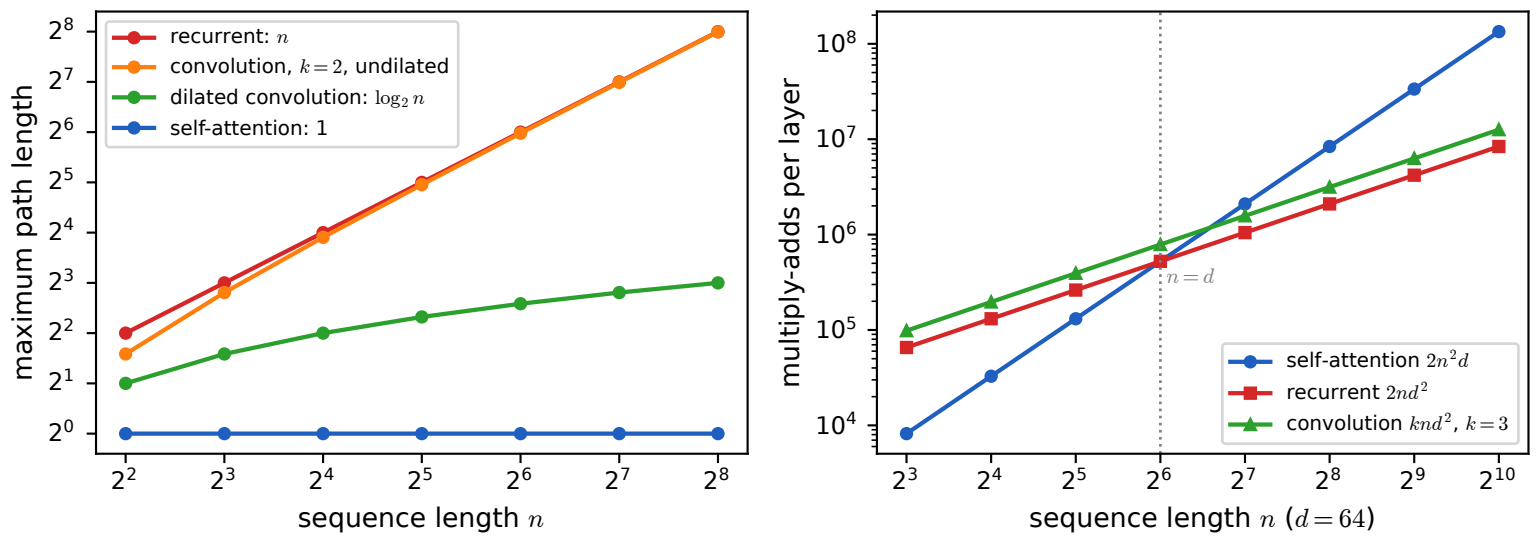

recurrent: n
convolution, $k=2$, undilated
dilated convolution: $\log_2 n$
self-attention: 1
maximum path length
sequence length n
multiply-adds per layer
$n=d$
self-attention $2n^2d$
recurrent $2nd^2$
convolution knd^2, $k=3$
sequence length n ($d=64$)