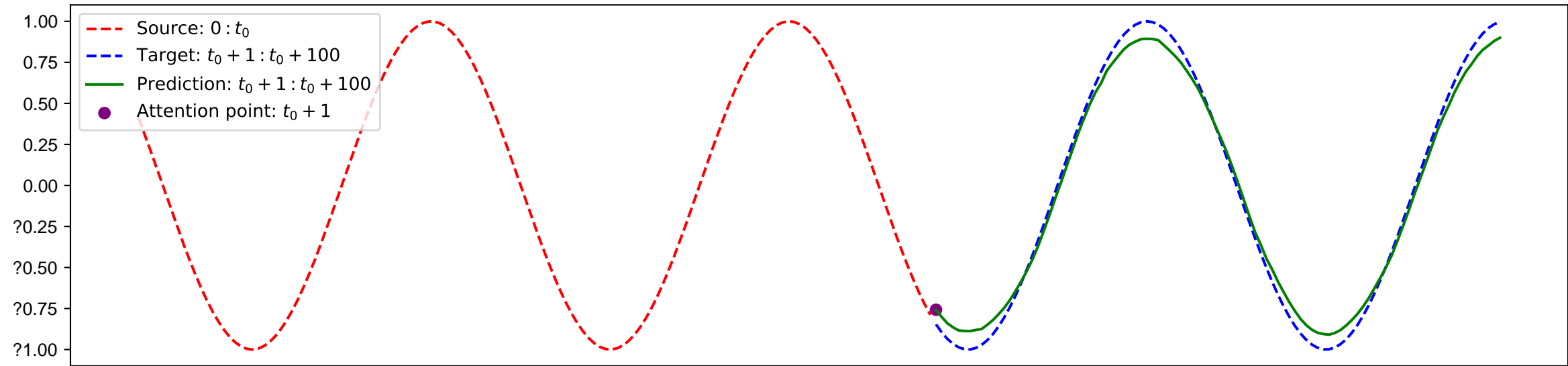n-steps non-overlapping, total epochs: 256

Attention pattens for $t_0 + 1$