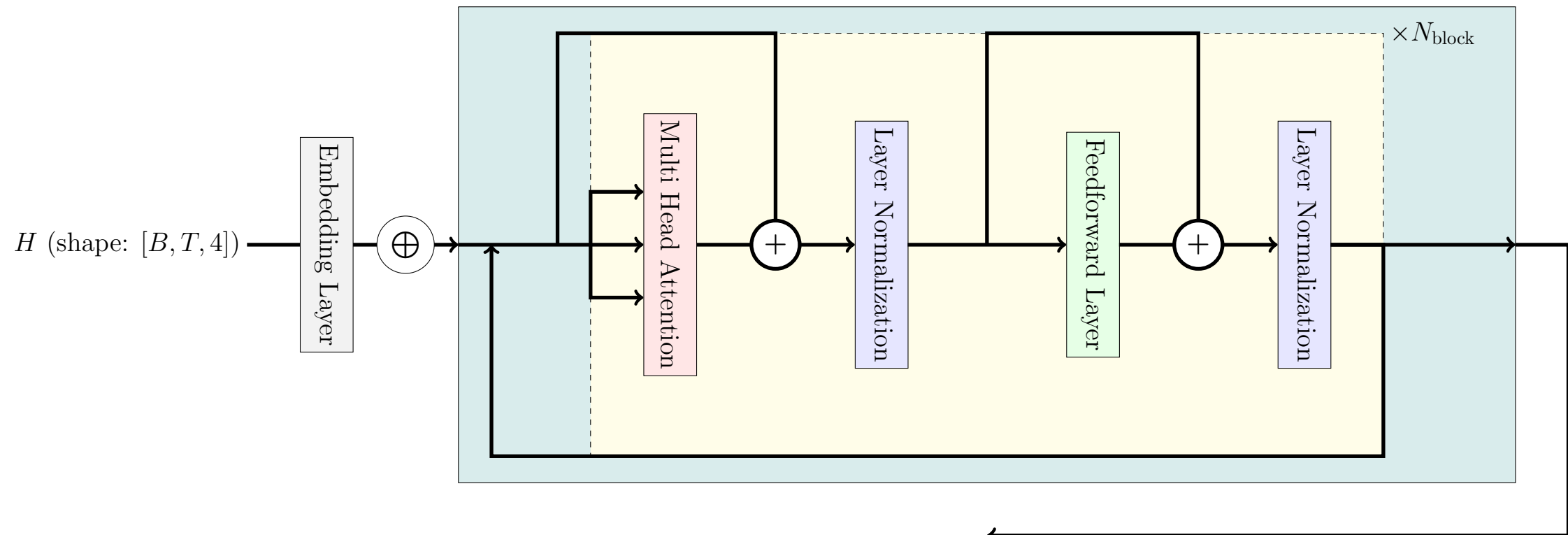
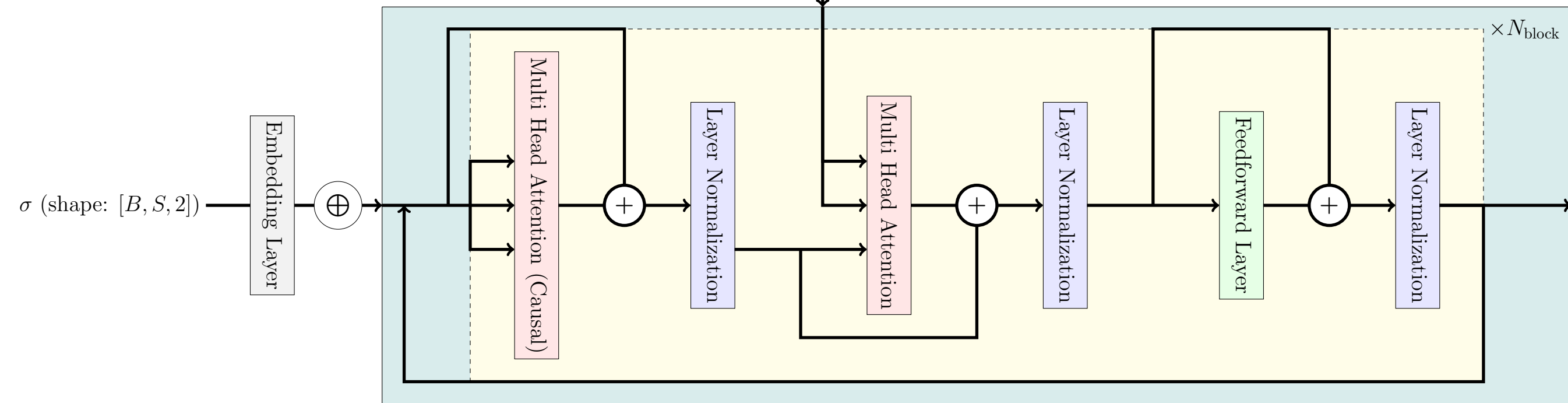


Transformer Encoder


 X (shape: $[B, T, E]$)

 Y (shape: $[B, S, E]$)

Dense Layer

Softmax

 P (shape: $[B, S, 2]$)

Transformer Decoder