Attention based models in End-to-End ASR

Exploration of Attention in ESPNET toolkit

Shreekantha Nadig

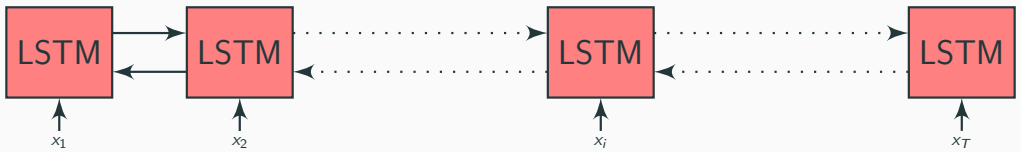
November 9, 2018

International Institute of Information Technology - Bangalore

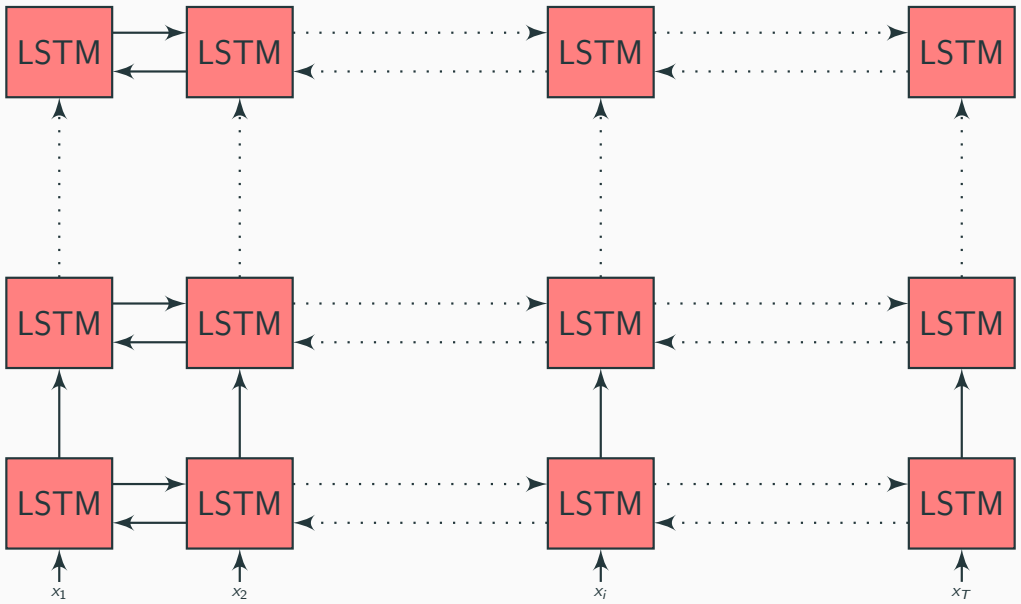
Table of contents

Introduction

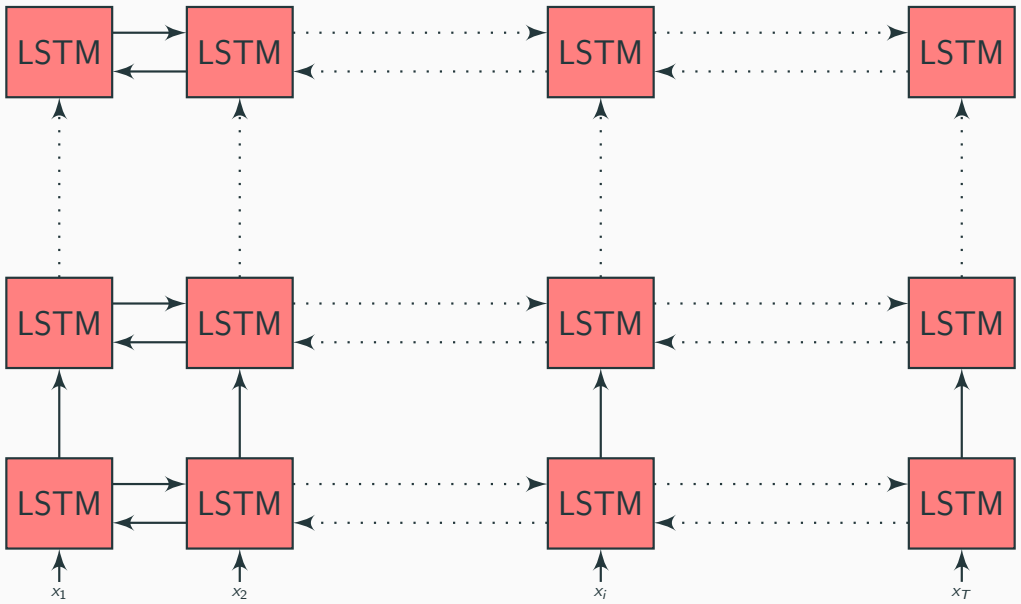
BLSTM



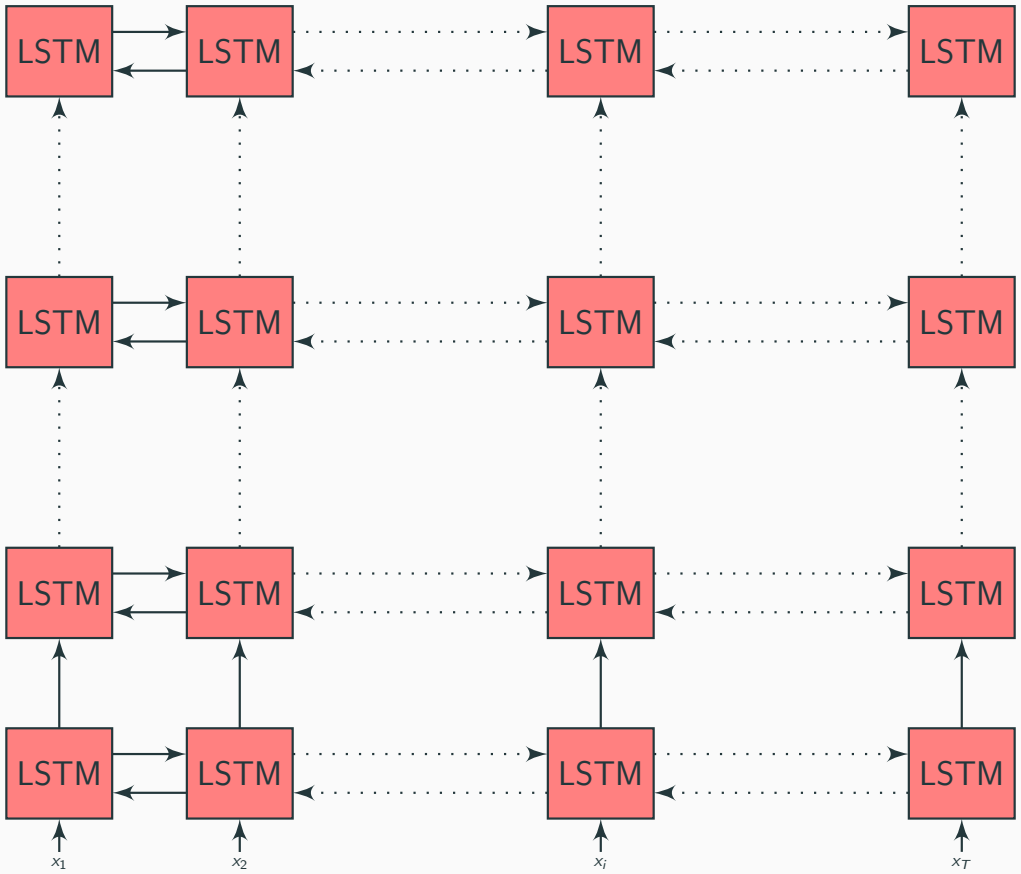
BLSTM


LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
x_1
x_2
x_i
x_T

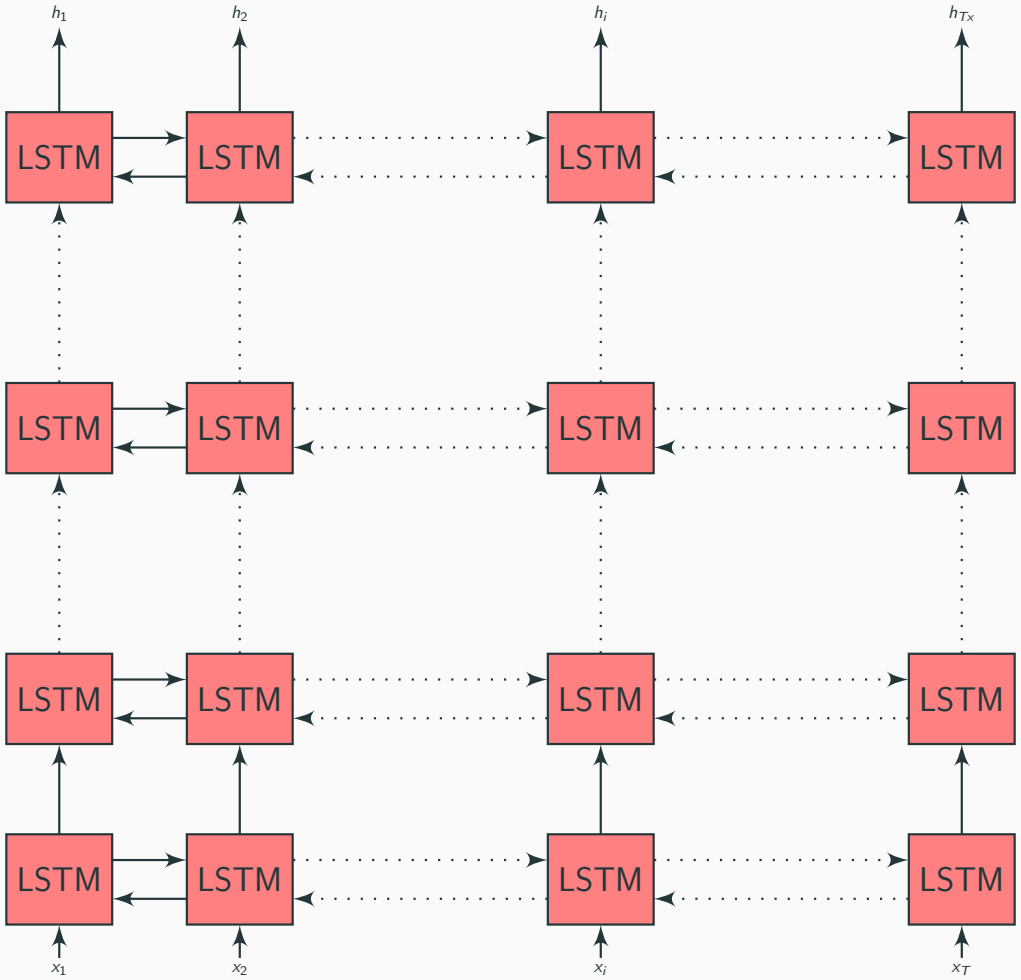
BLSTM


LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
x_1
x_2
x_i
x_T

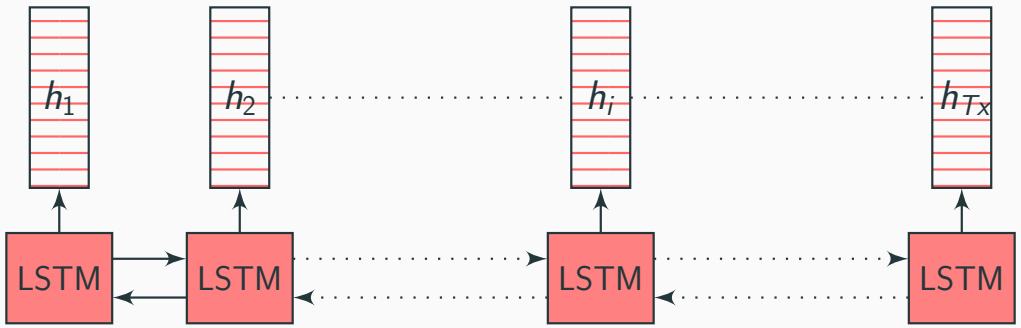
BLSTM


LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
x_1
x_2
x_i
x_T

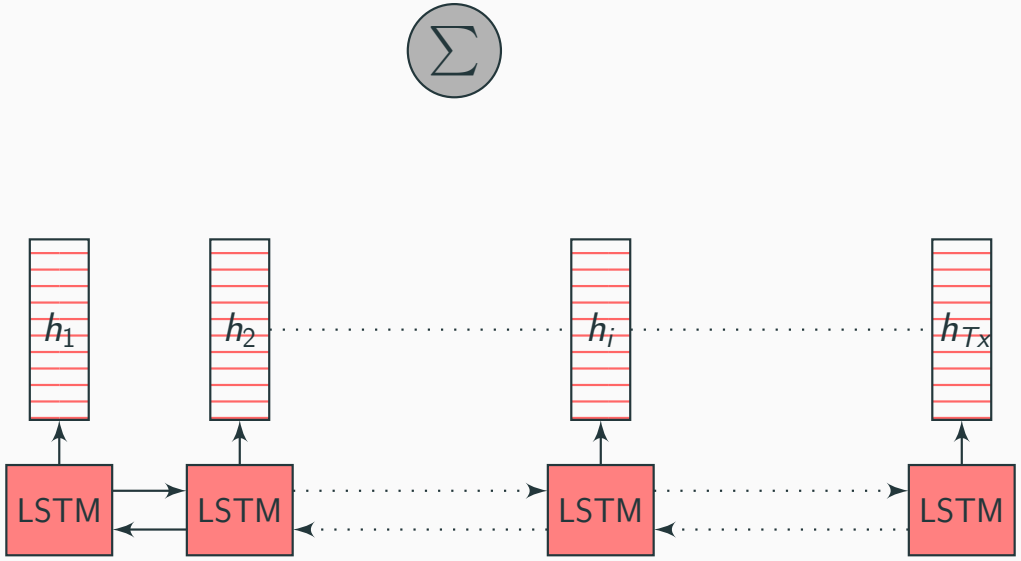
BLSTM



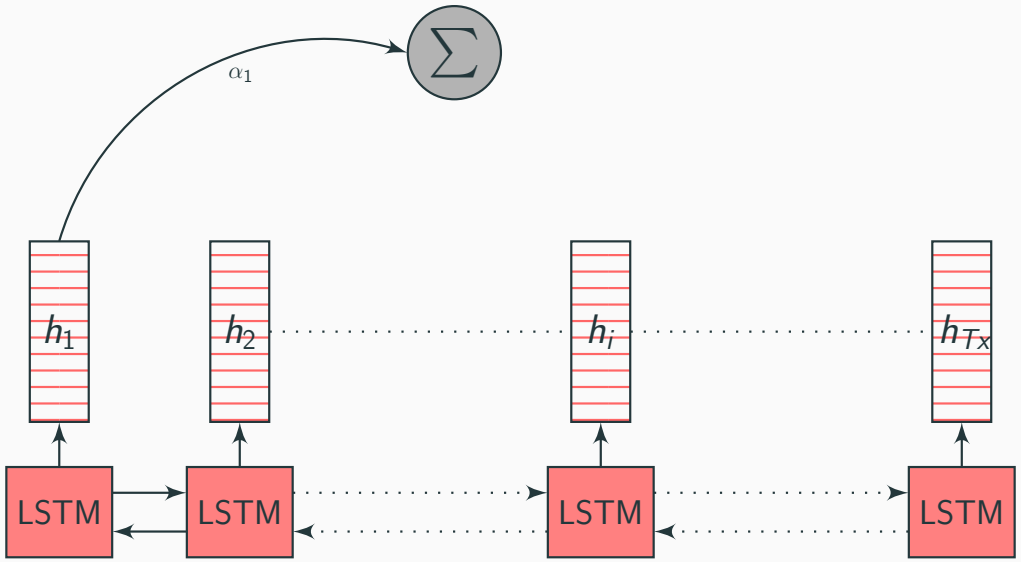
Introduction to Attention



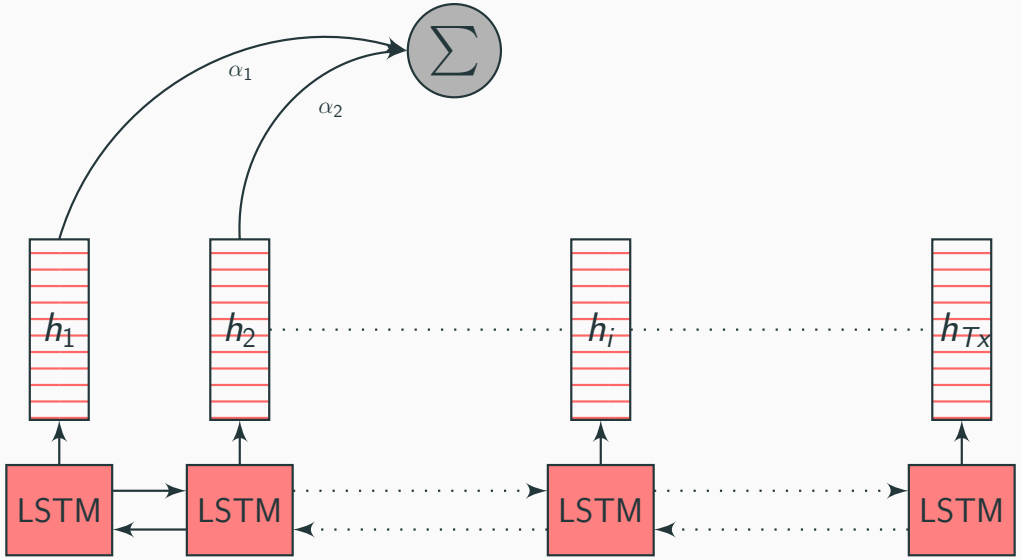
Introduction to Attention



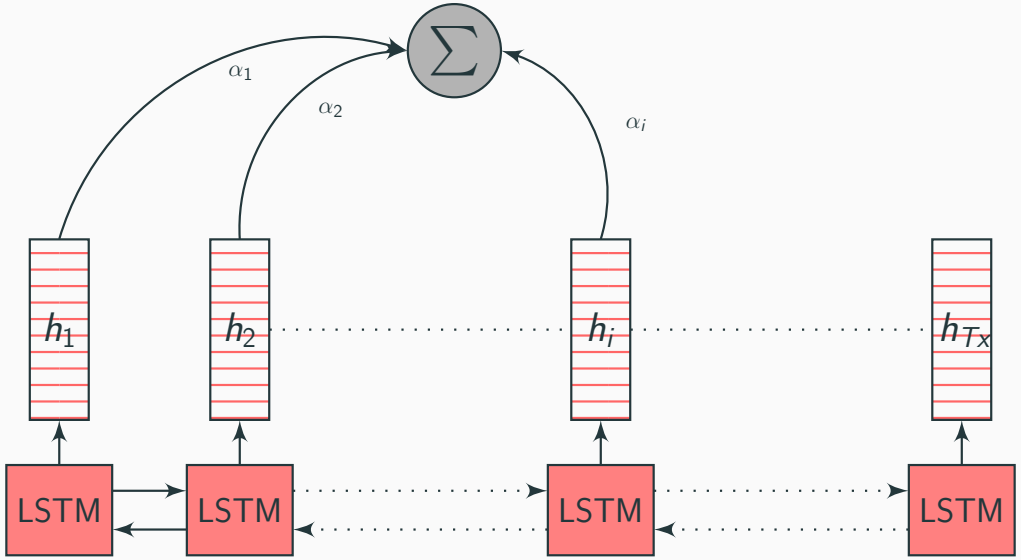
Introduction to Attention



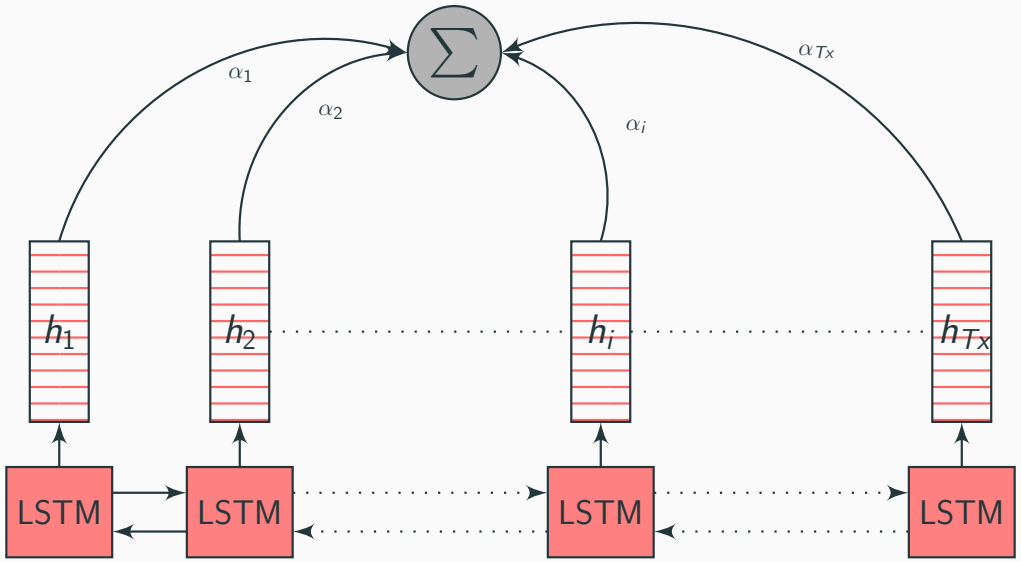
Introduction to Attention



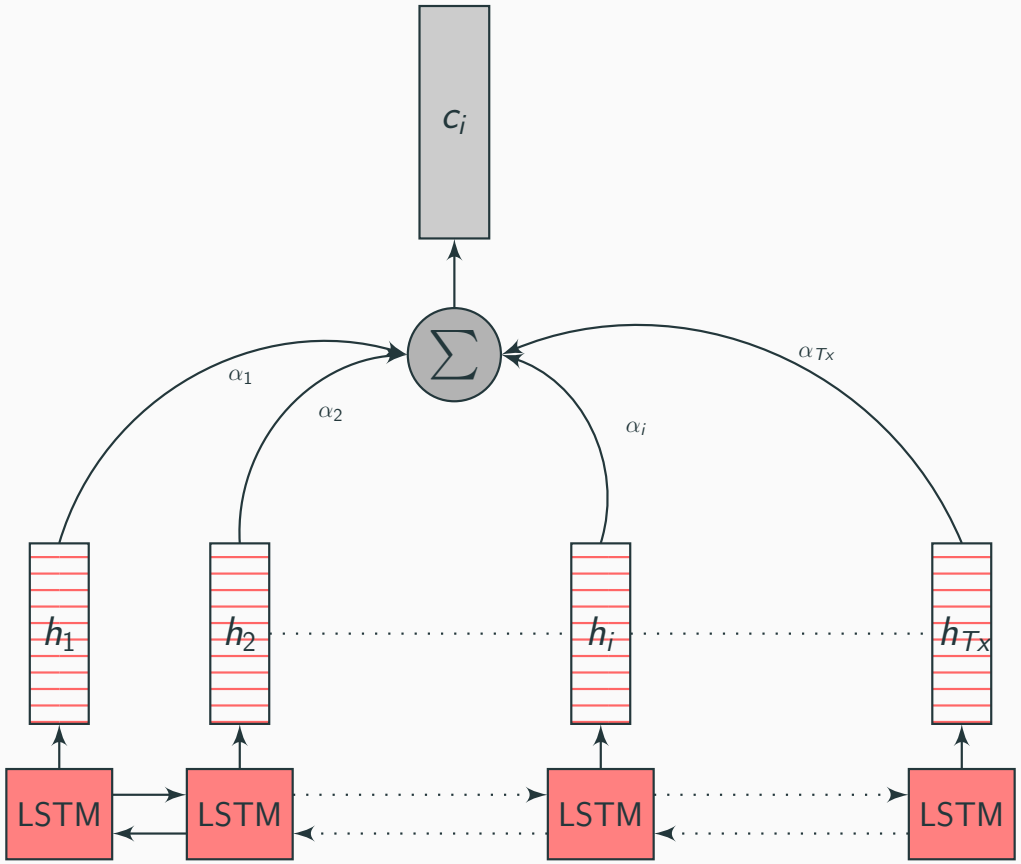
Introduction to Attention



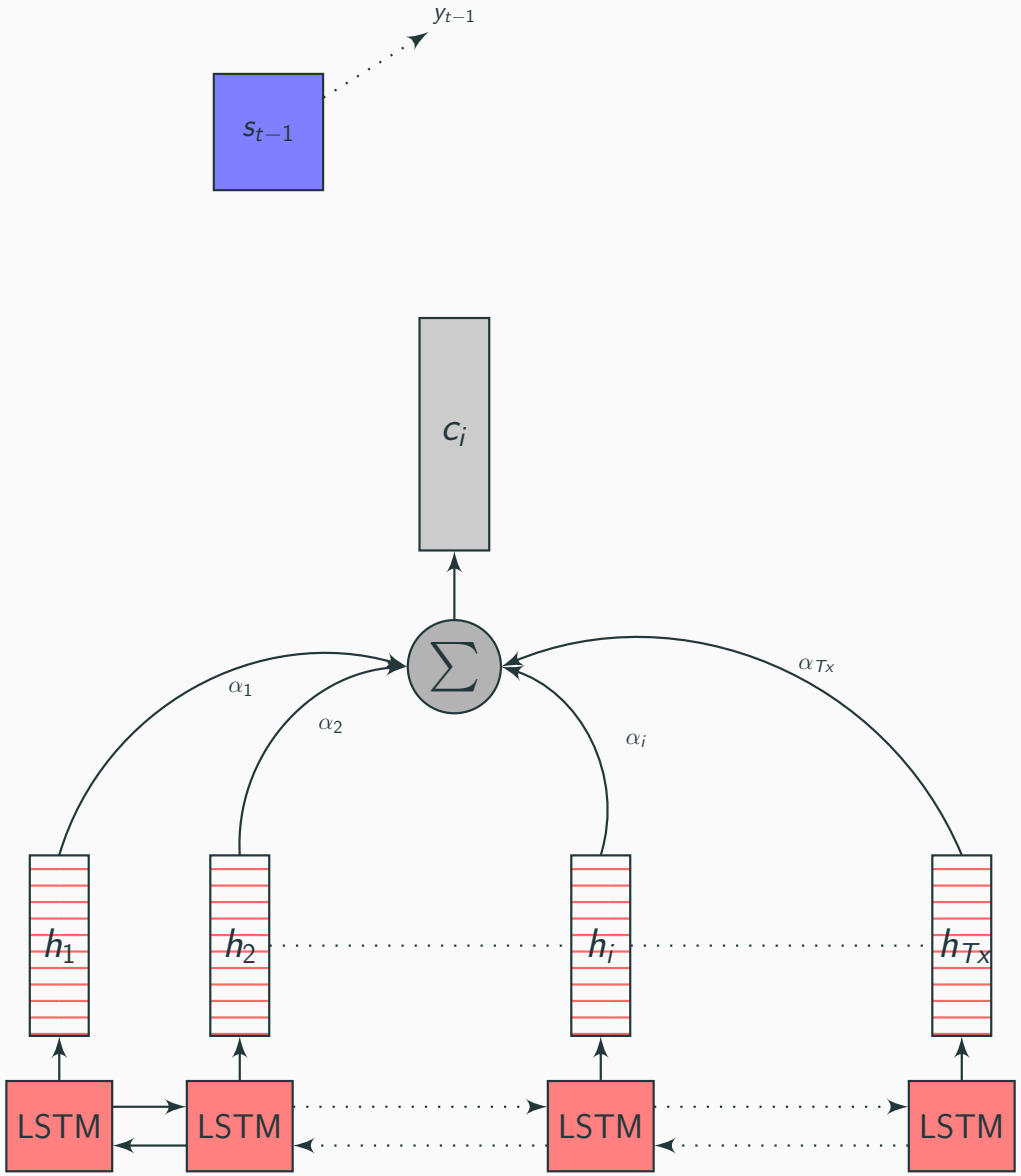
Introduction to Attention



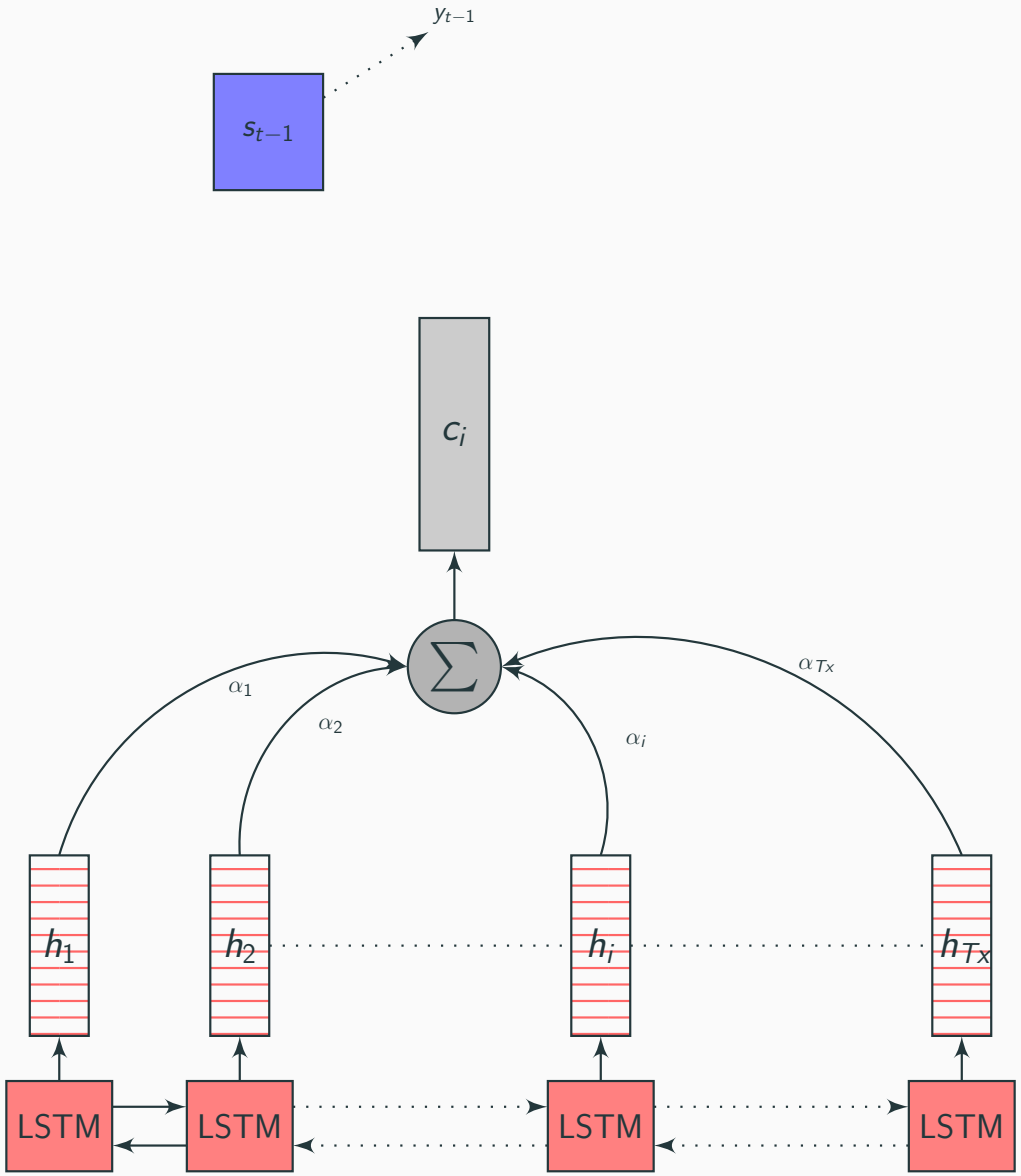
Introduction to Attention



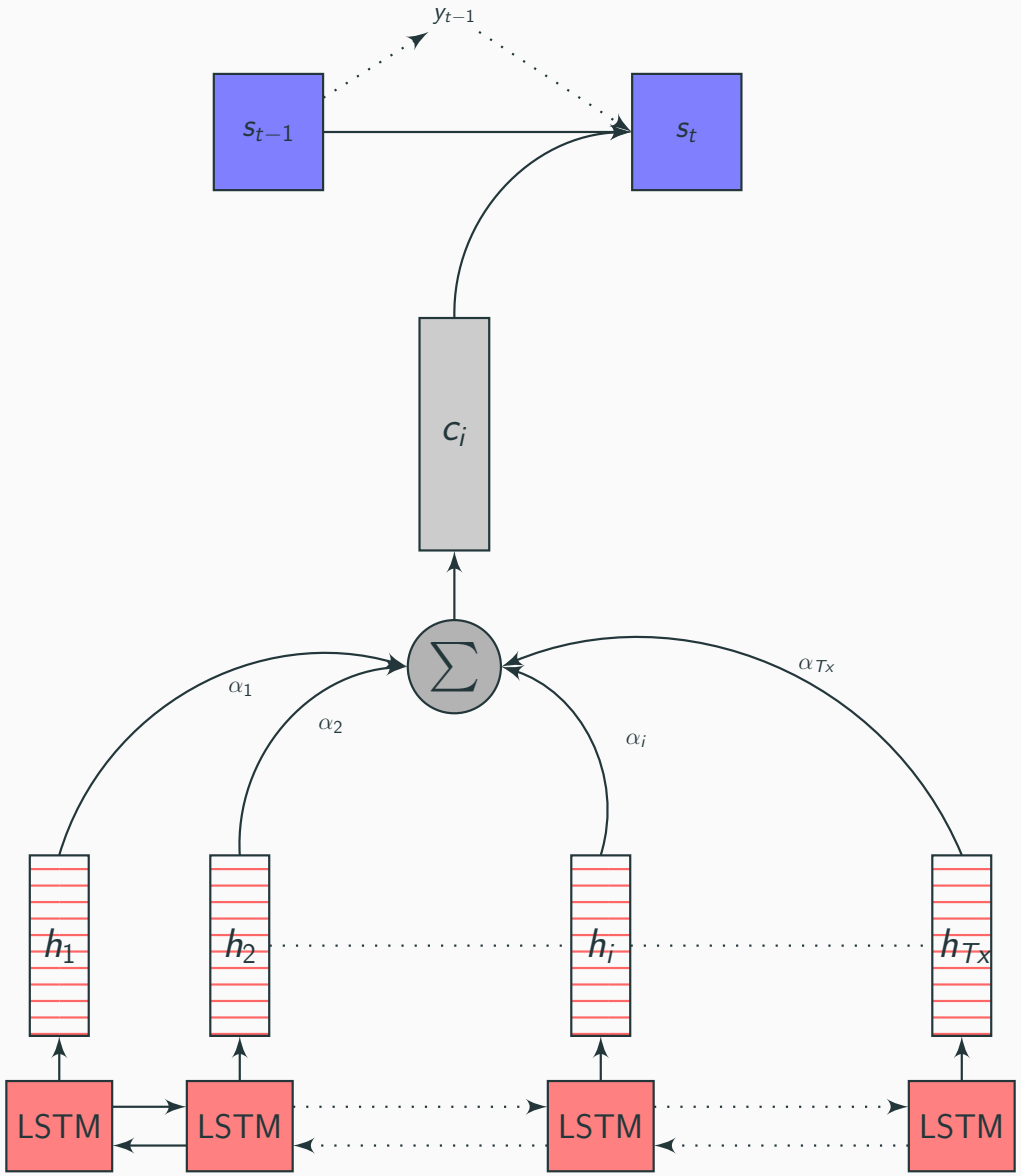
Introduction to Attention



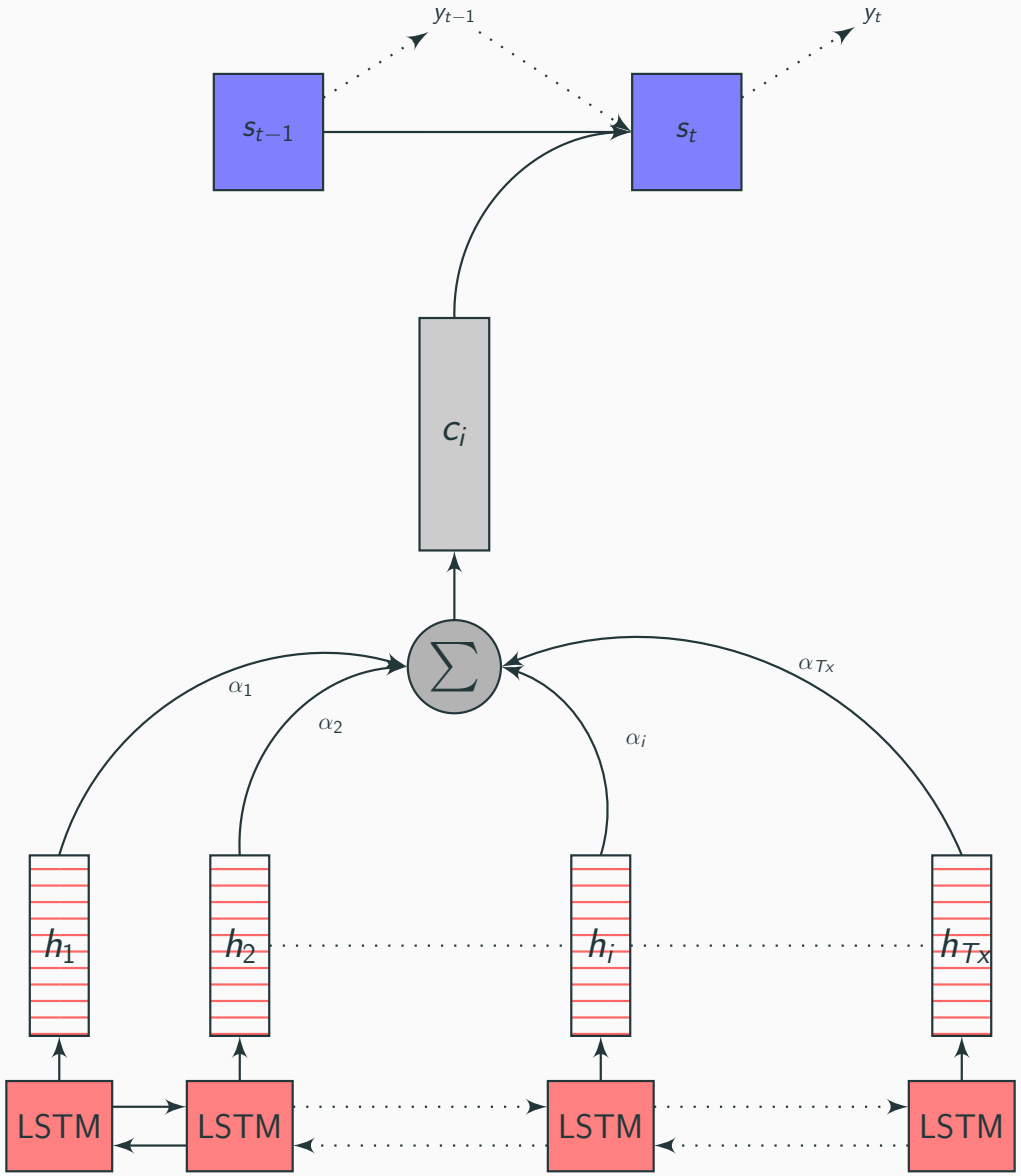
Introduction to Attention



Introduction to Attention

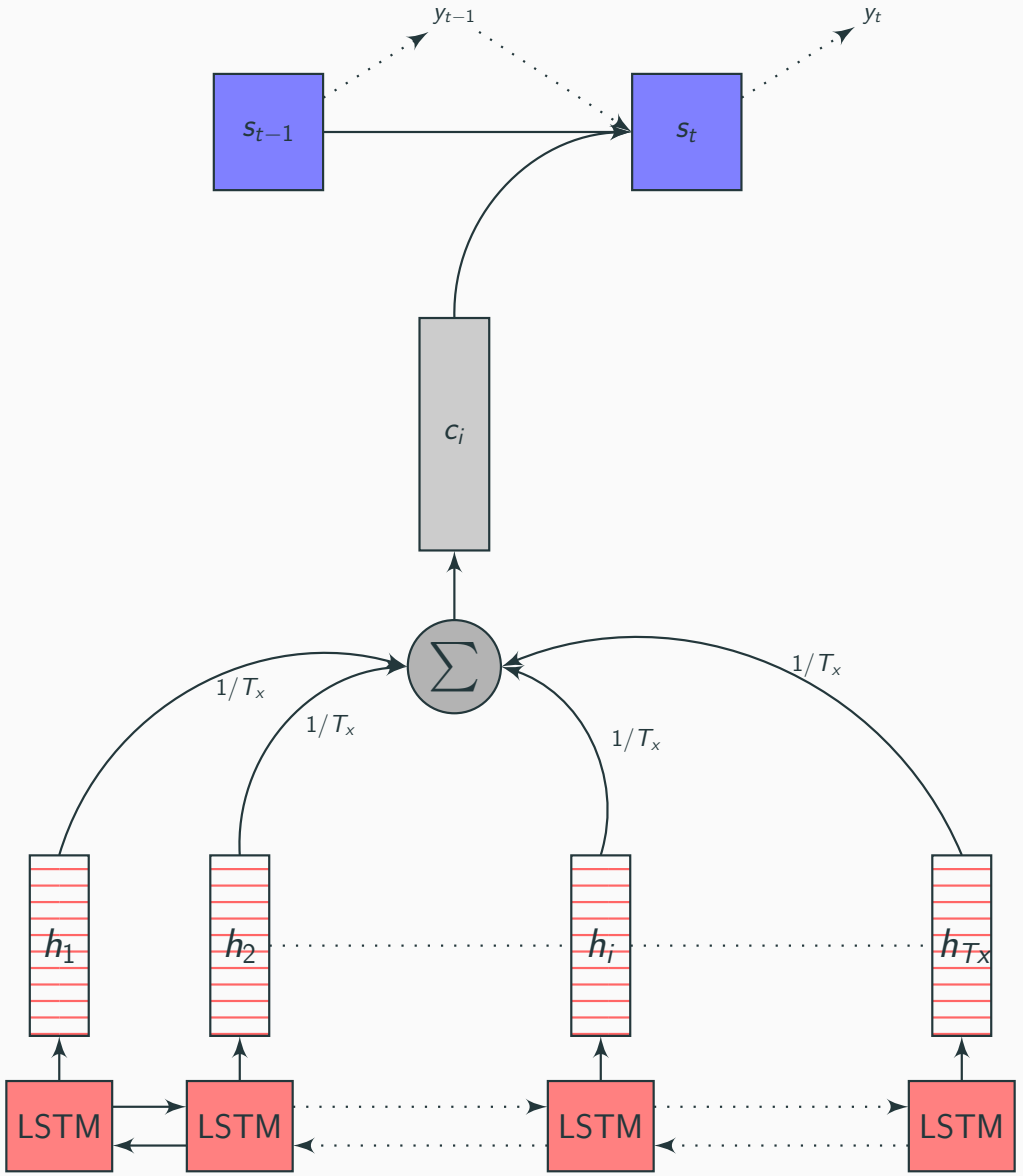


Introduction to Attention

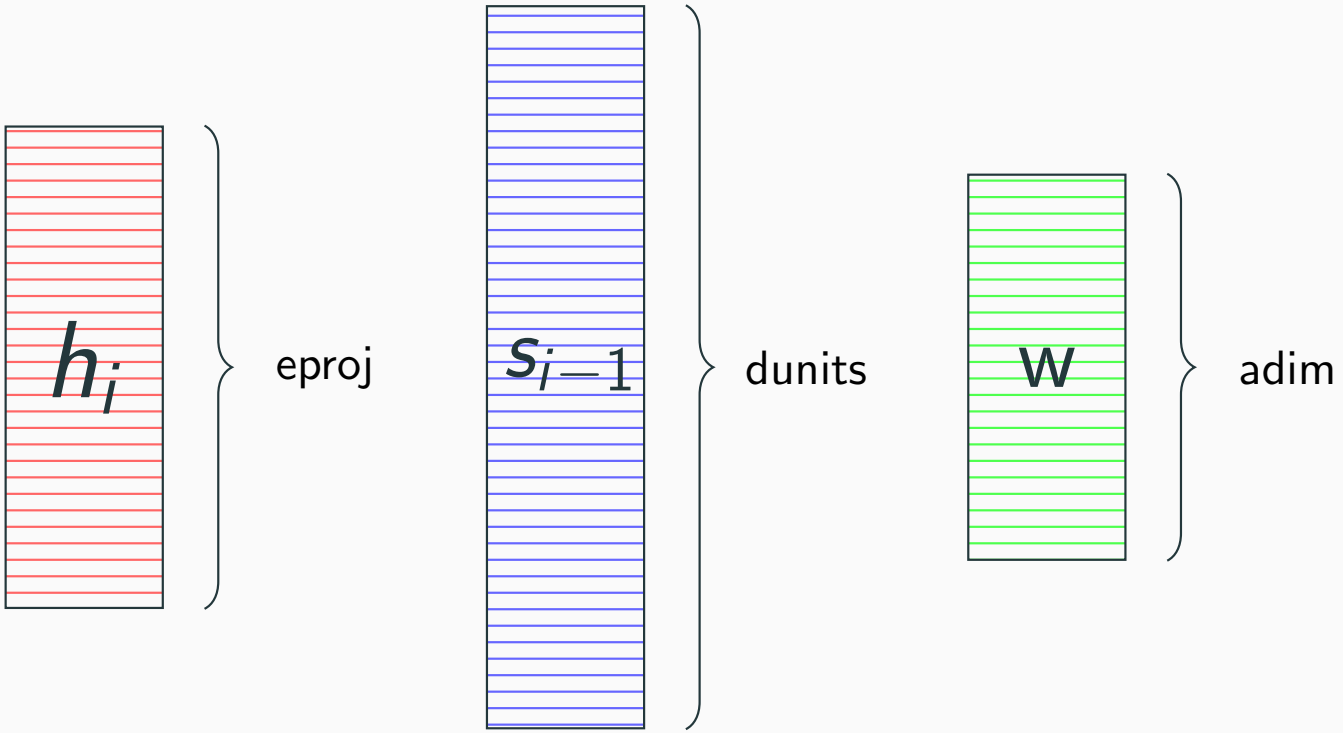


No Attention [Equal Attention?]

No Attention [Equal Attention?]

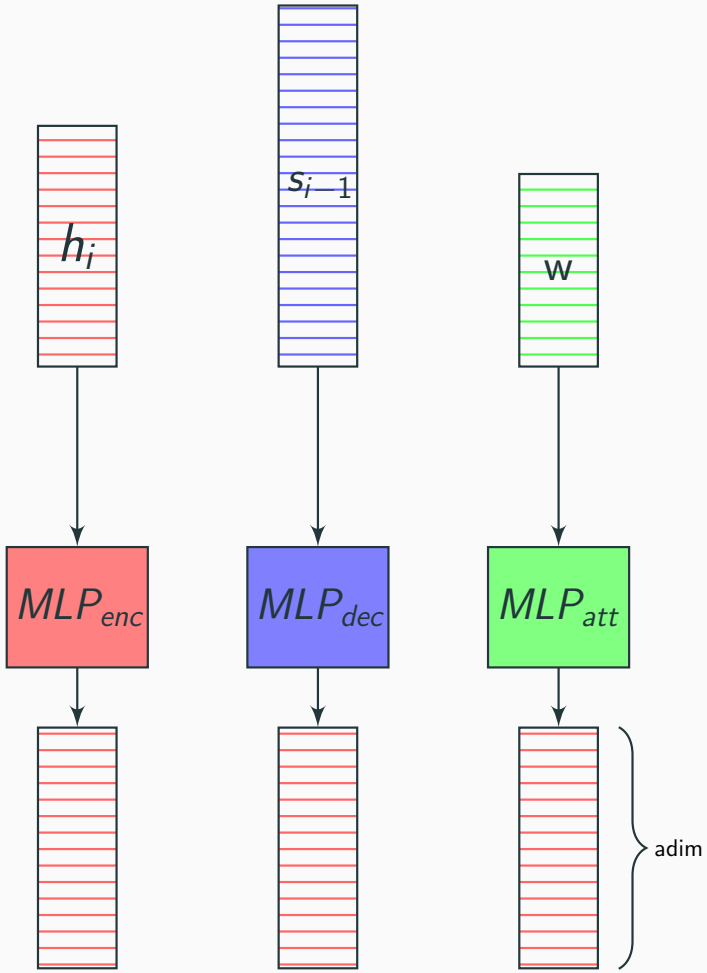


Dimensions of representations



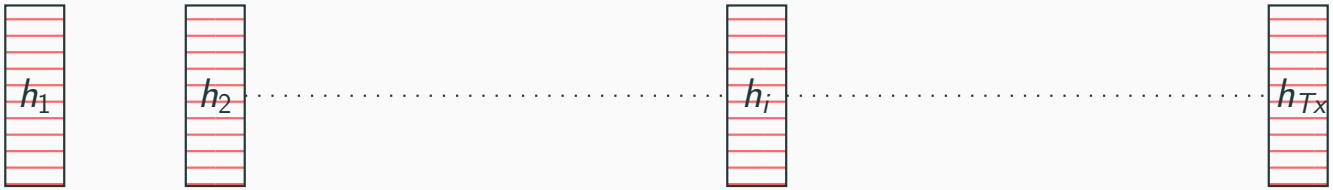
Mostly *eproj* $\neq$ *dunits* $\neq$ *adim*

Matching the dimensions of representations

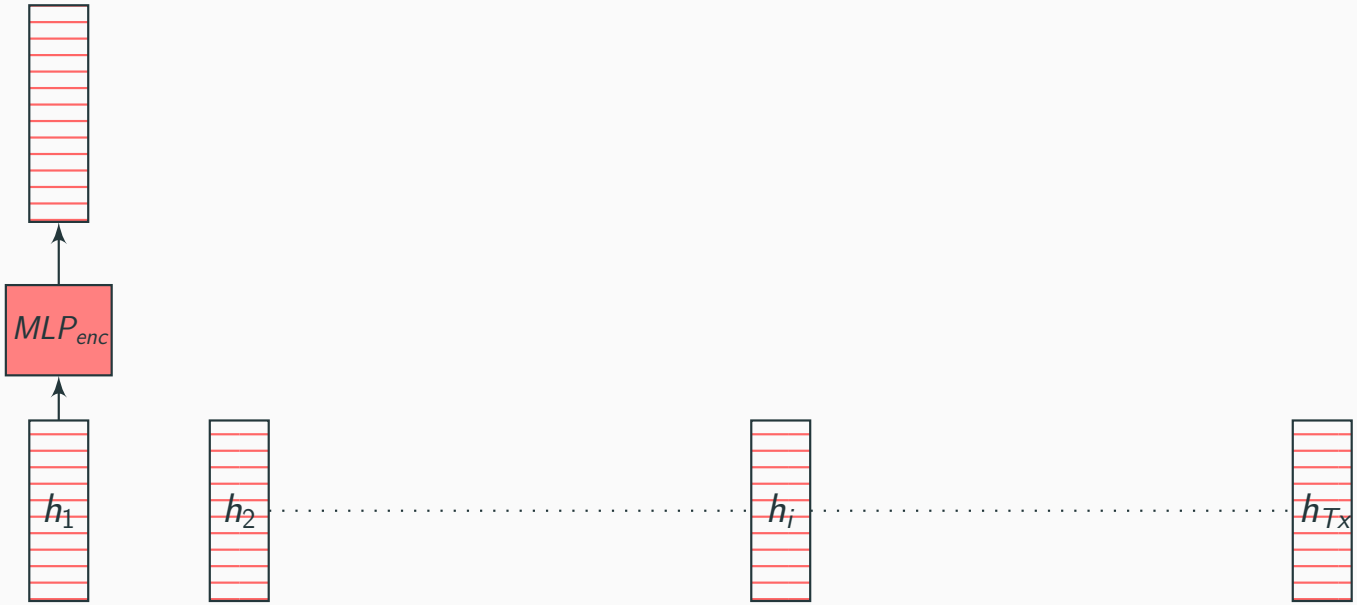


Dot product Attention

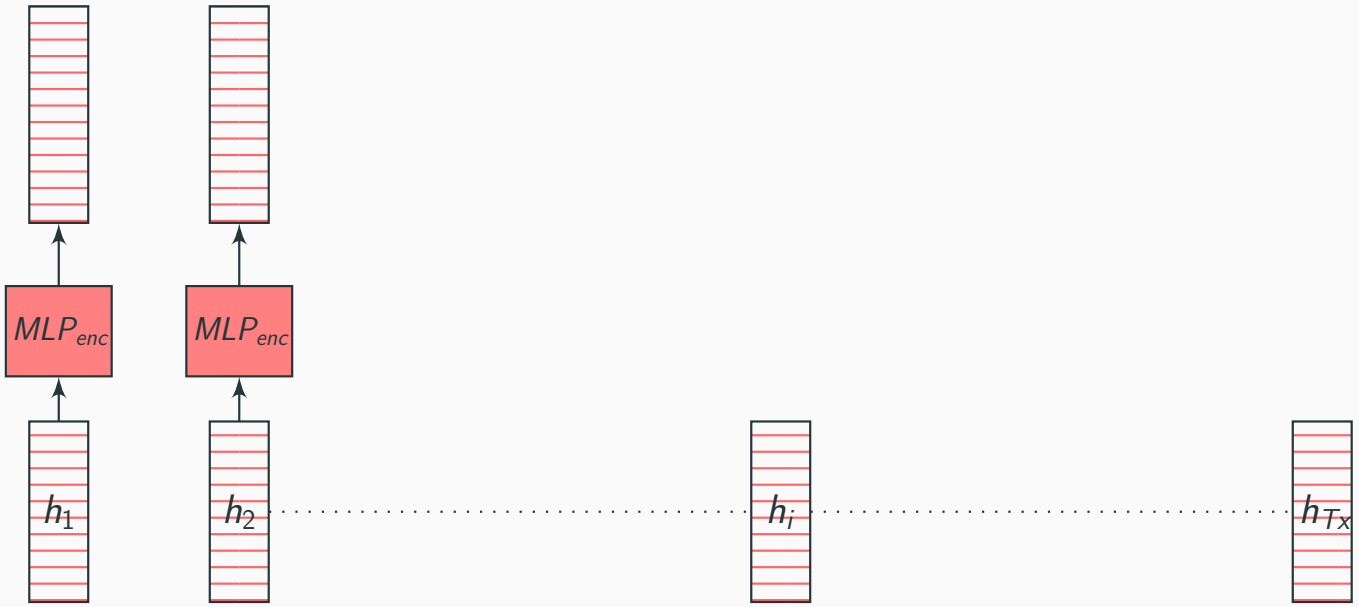
Dot product Attention



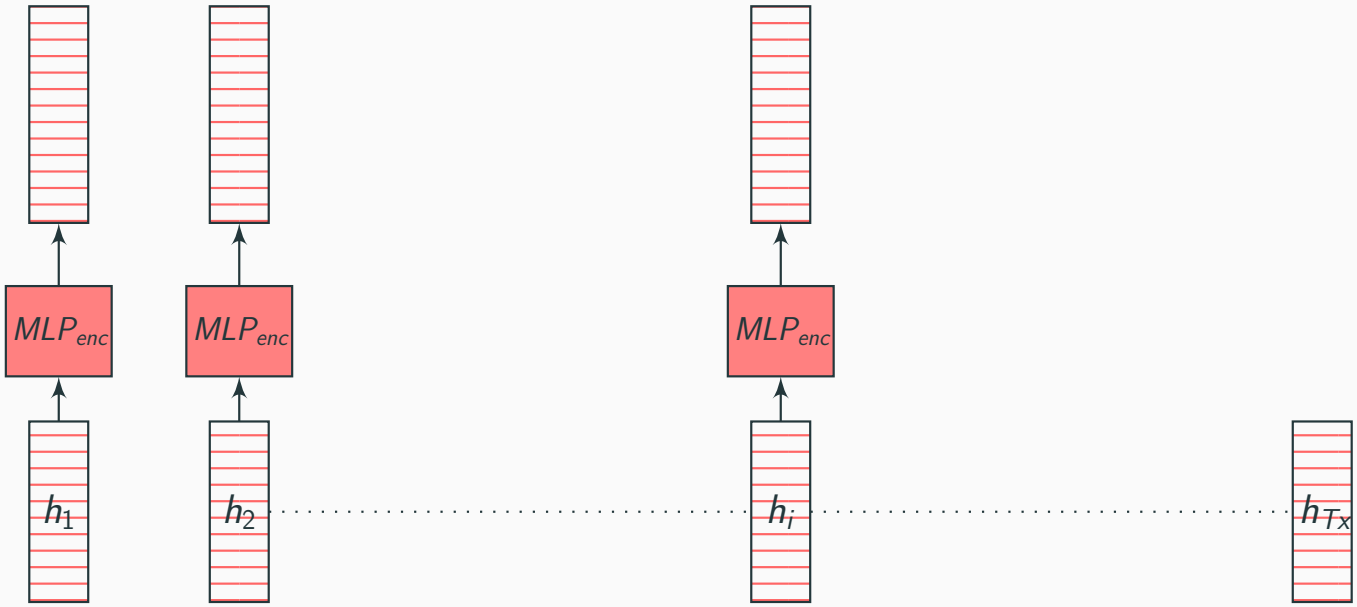
Dot product Attention



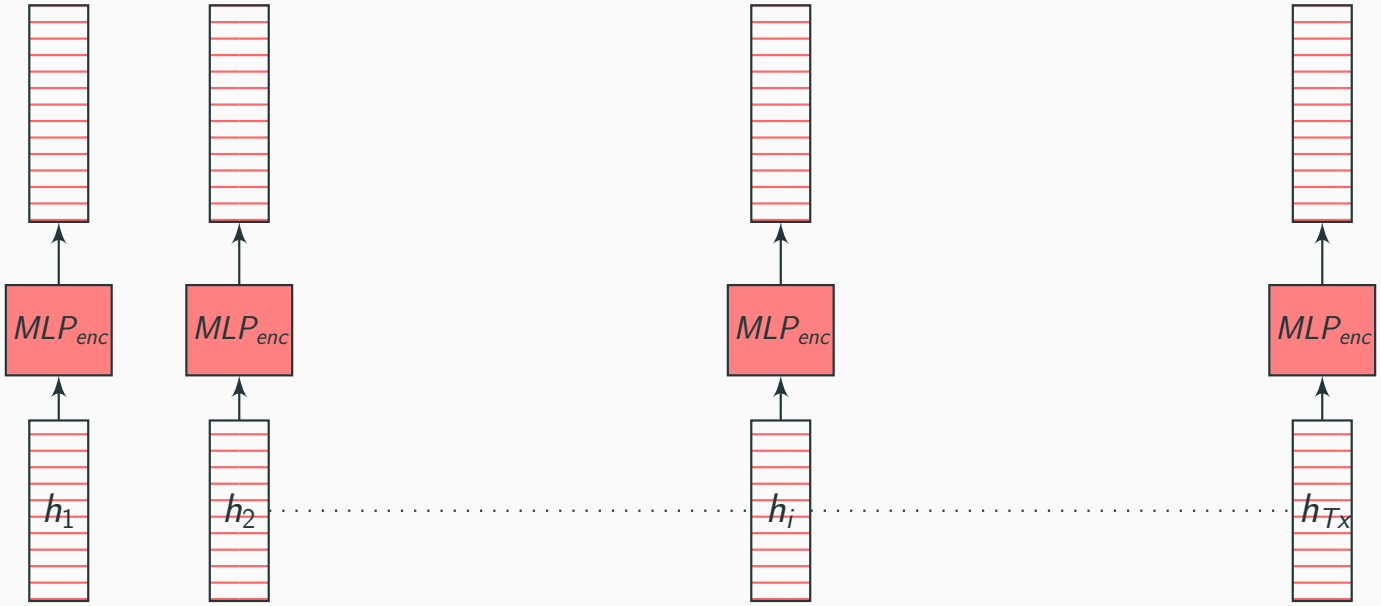
Dot product Attention



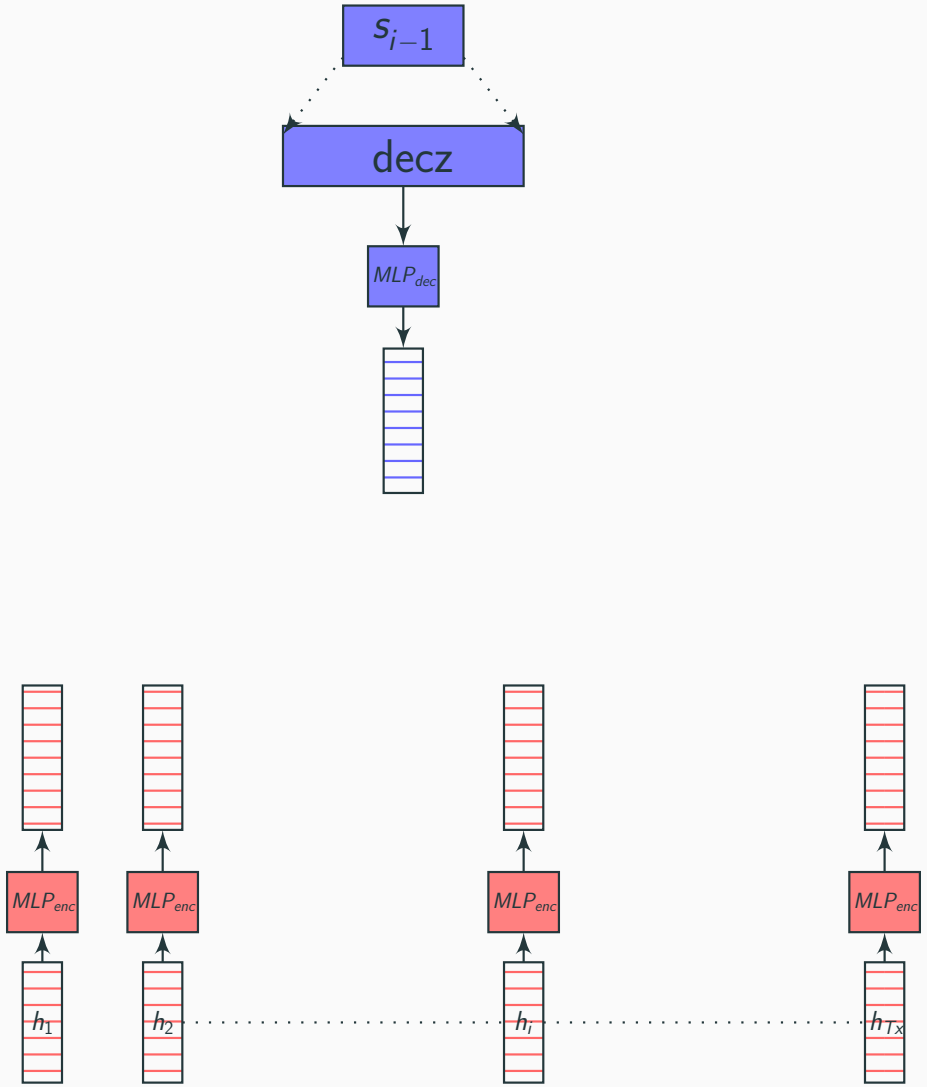
Dot product Attention



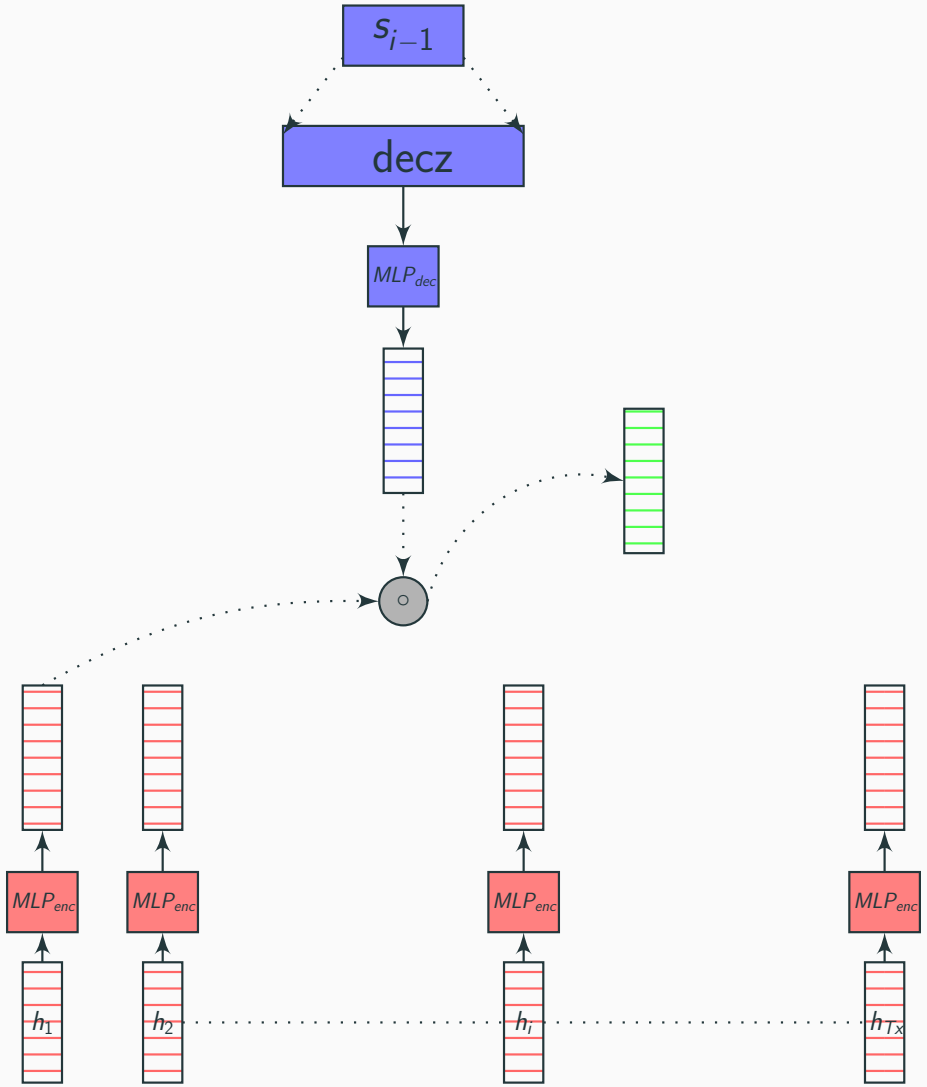
Dot product Attention



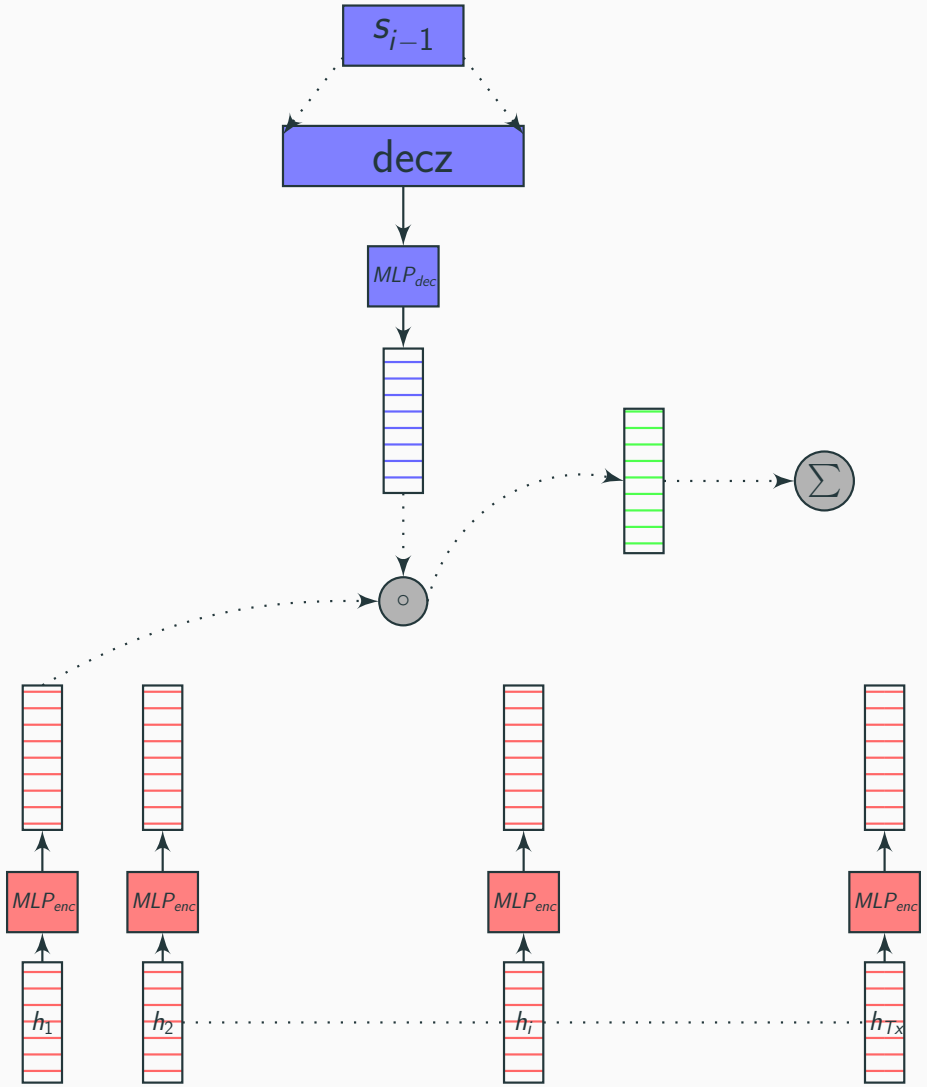
Dot product Attention



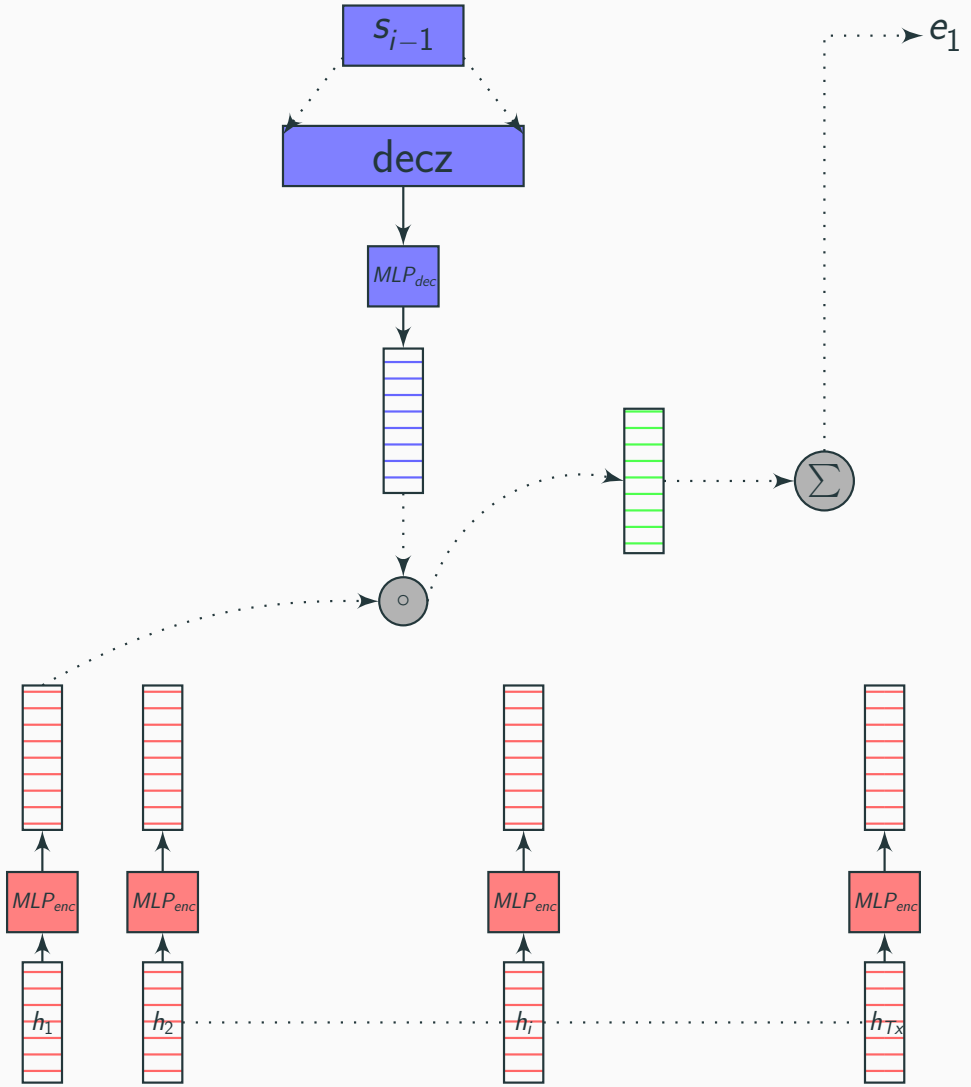
Dot product Attention



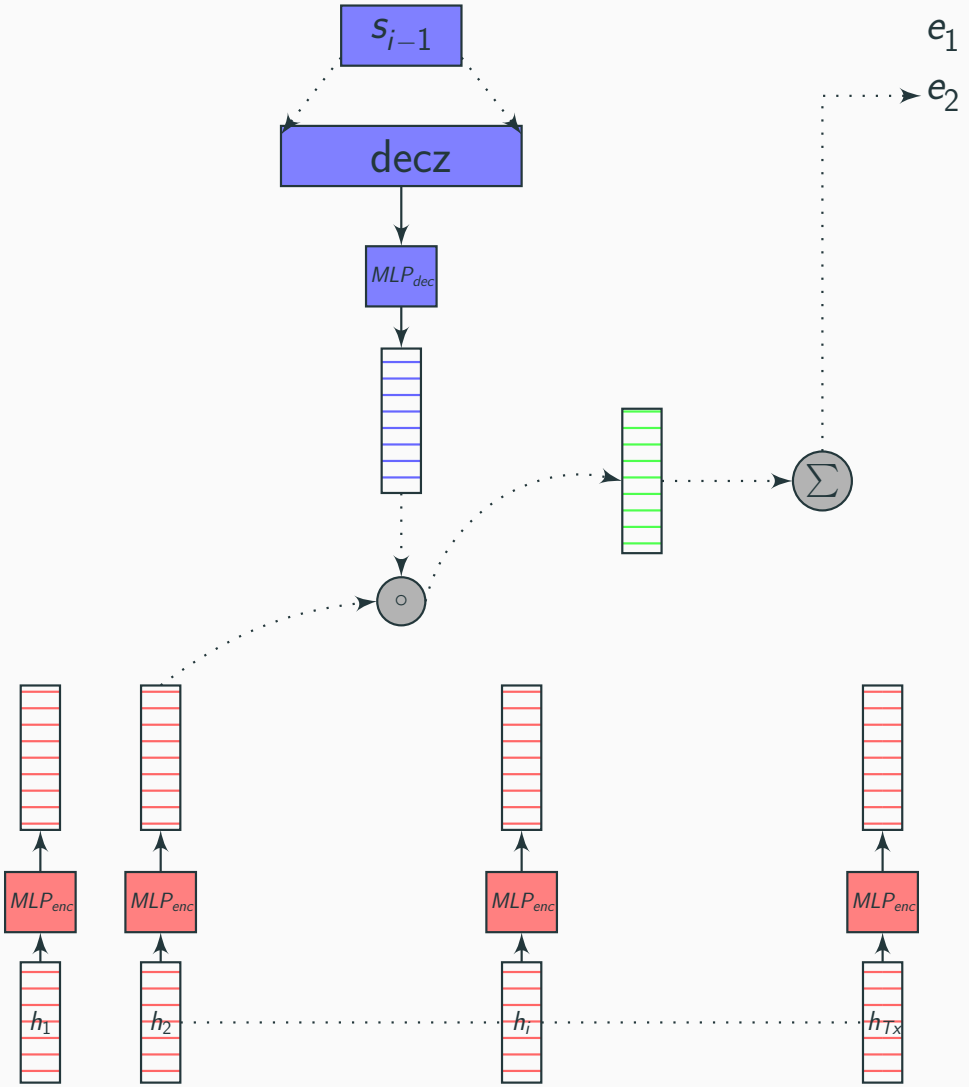
Dot product Attention



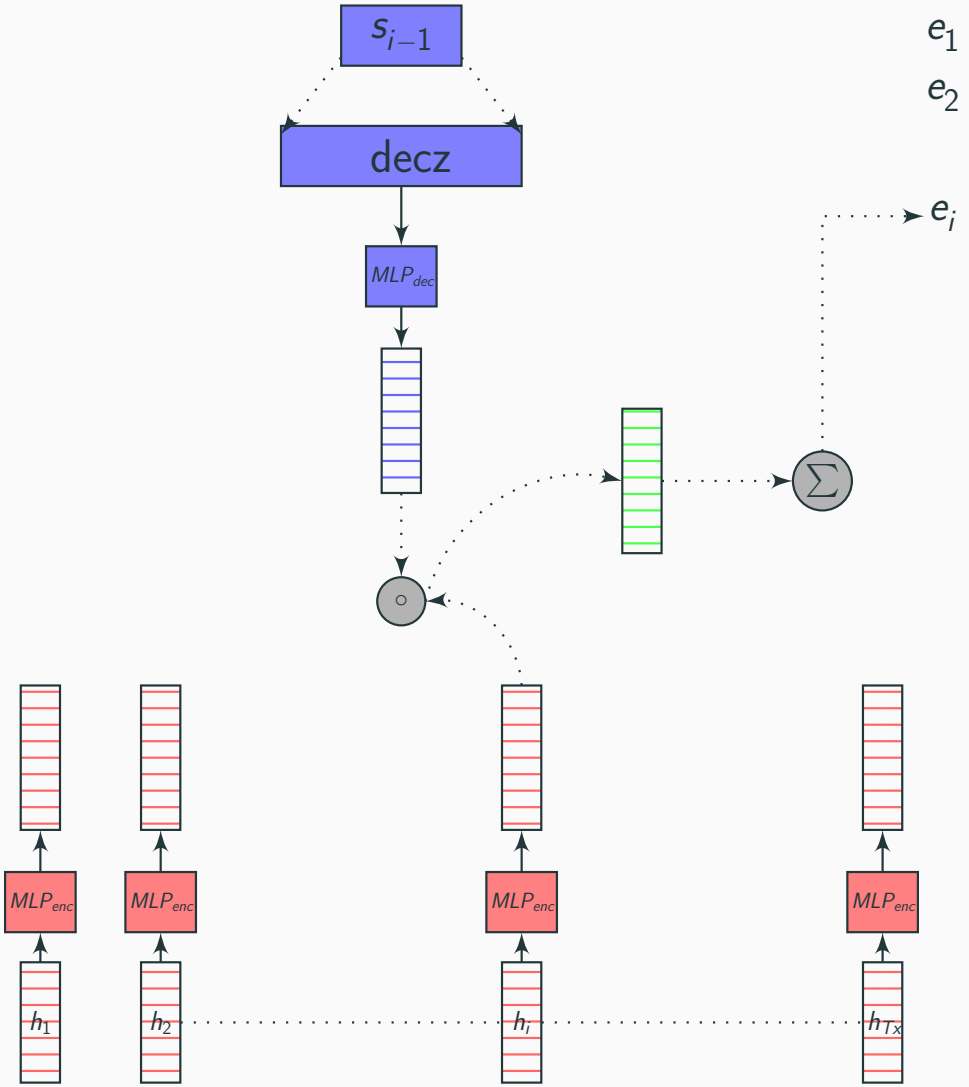
Dot product Attention



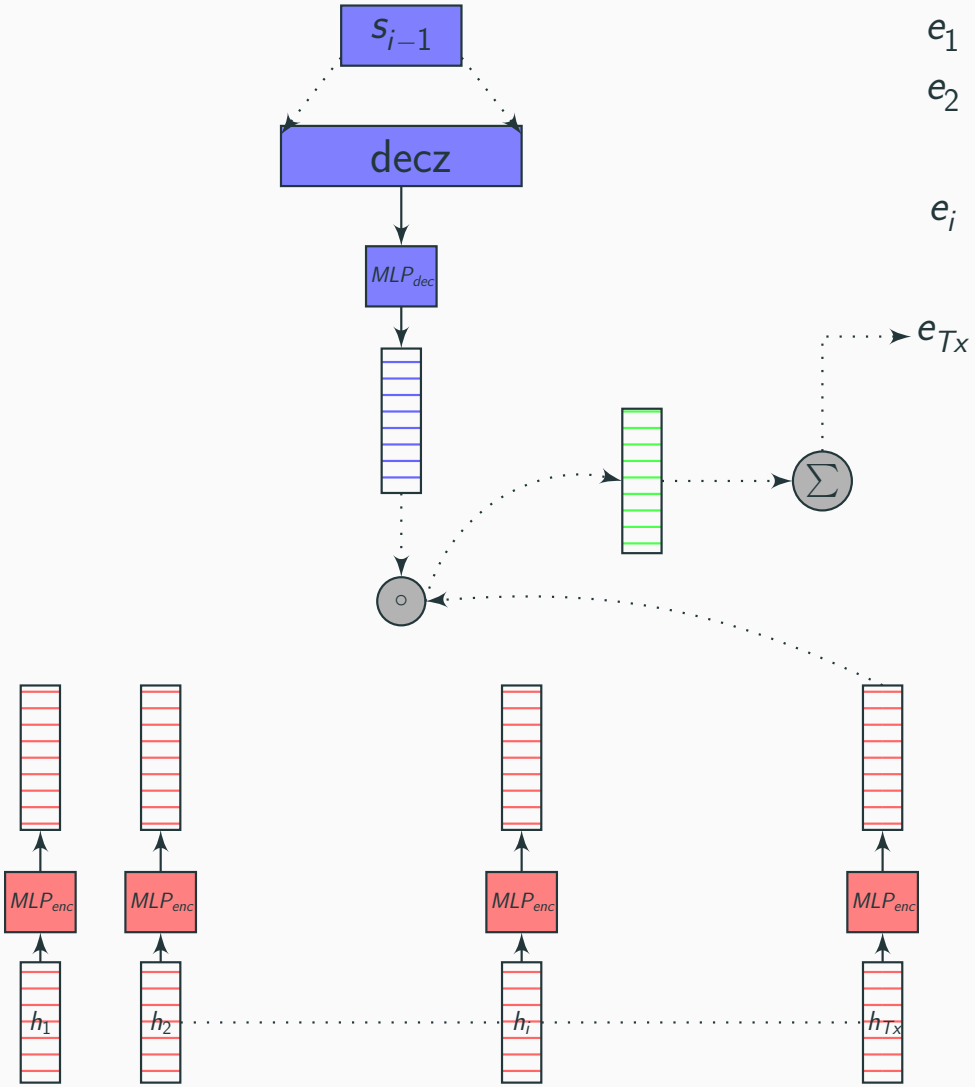
Dot product Attention



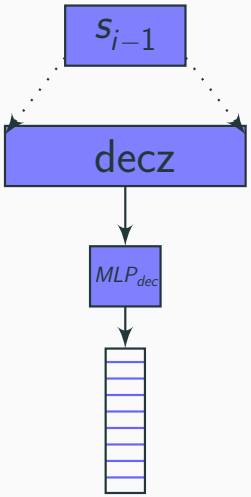
Dot product Attention

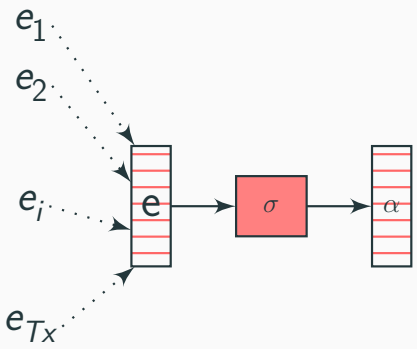


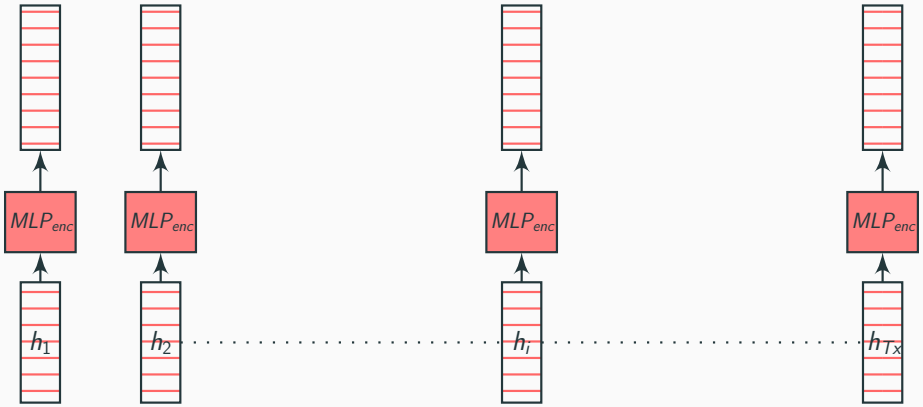
Dot product Attention



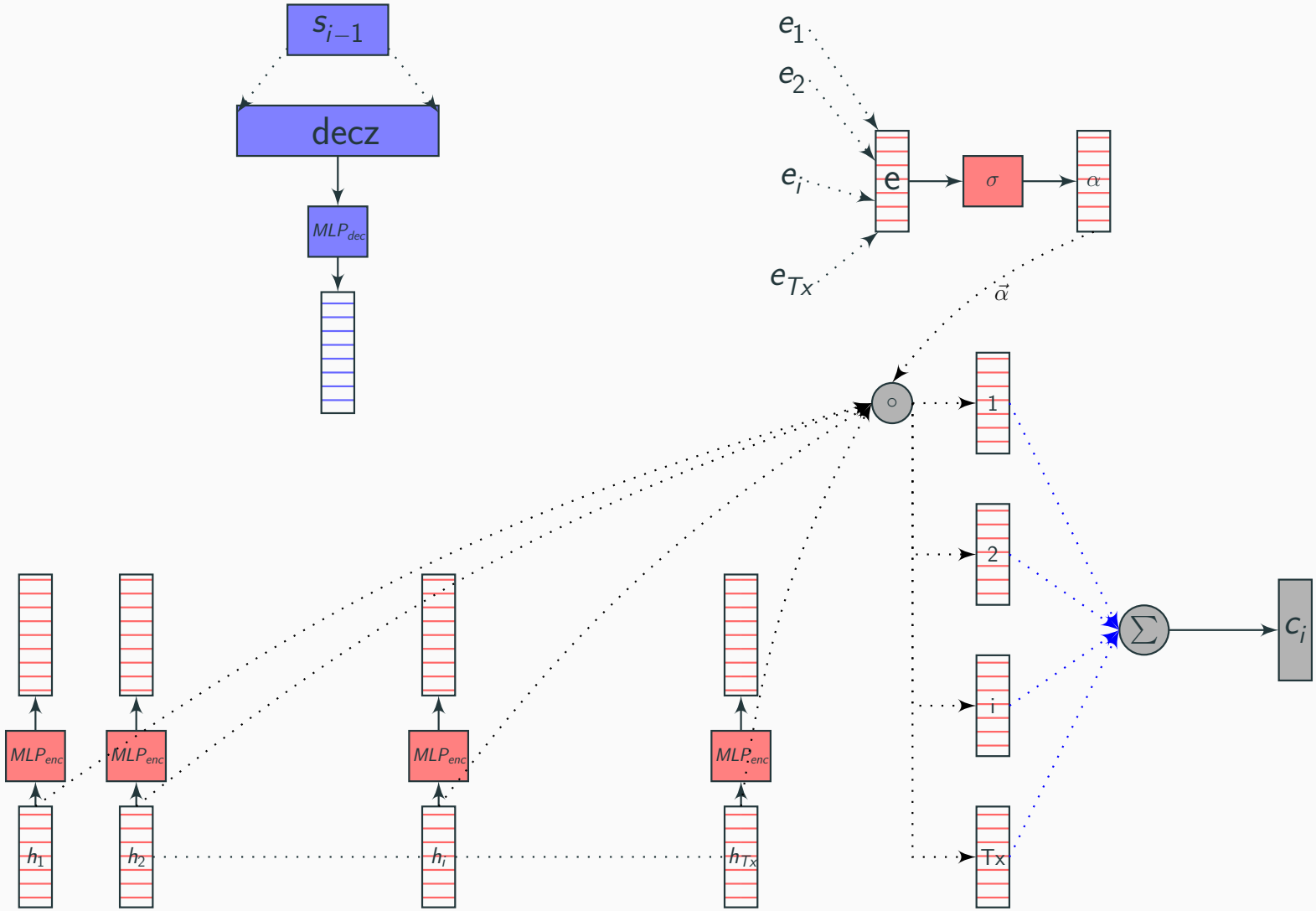
Dot product Attention



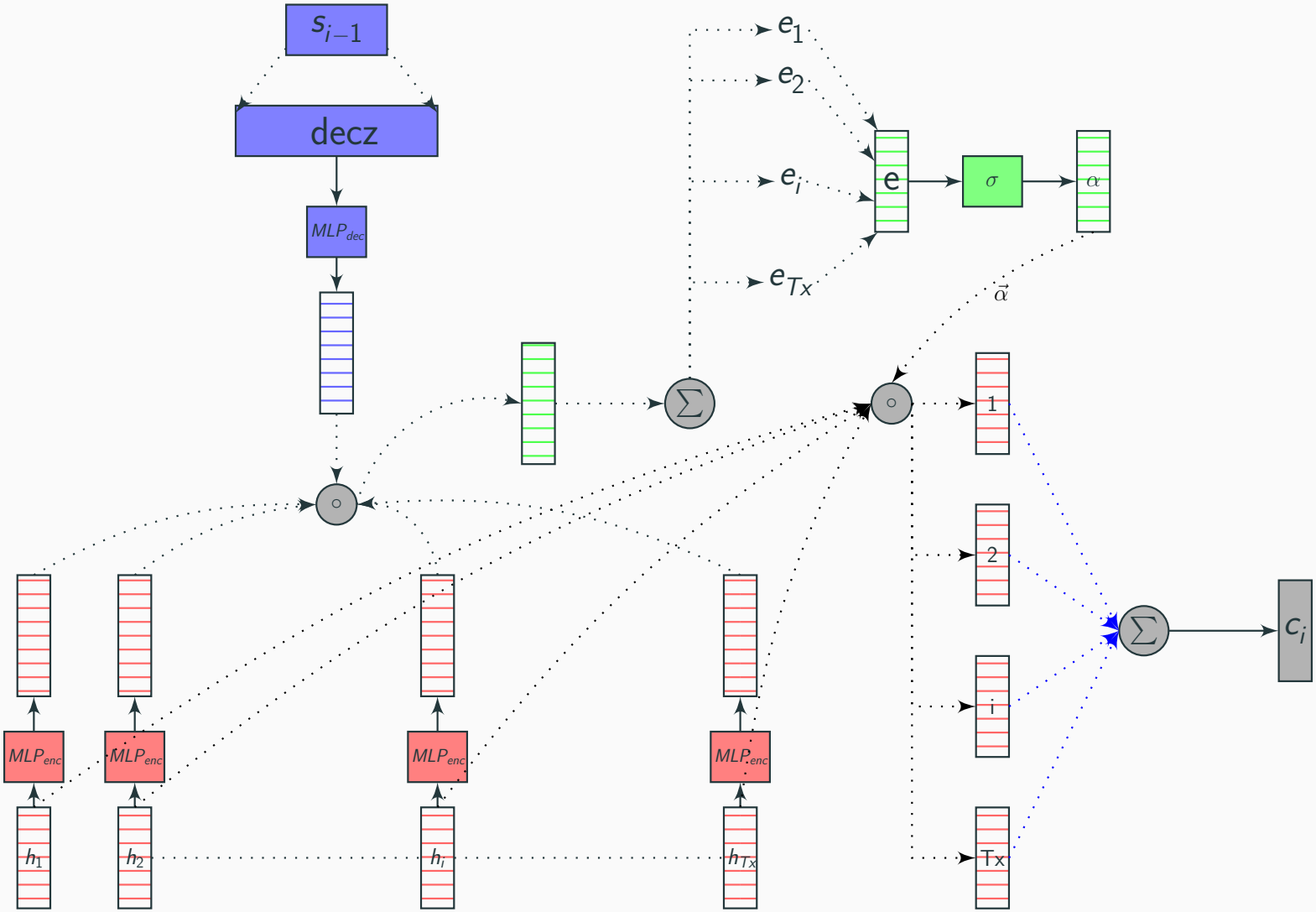




Dot product Attention

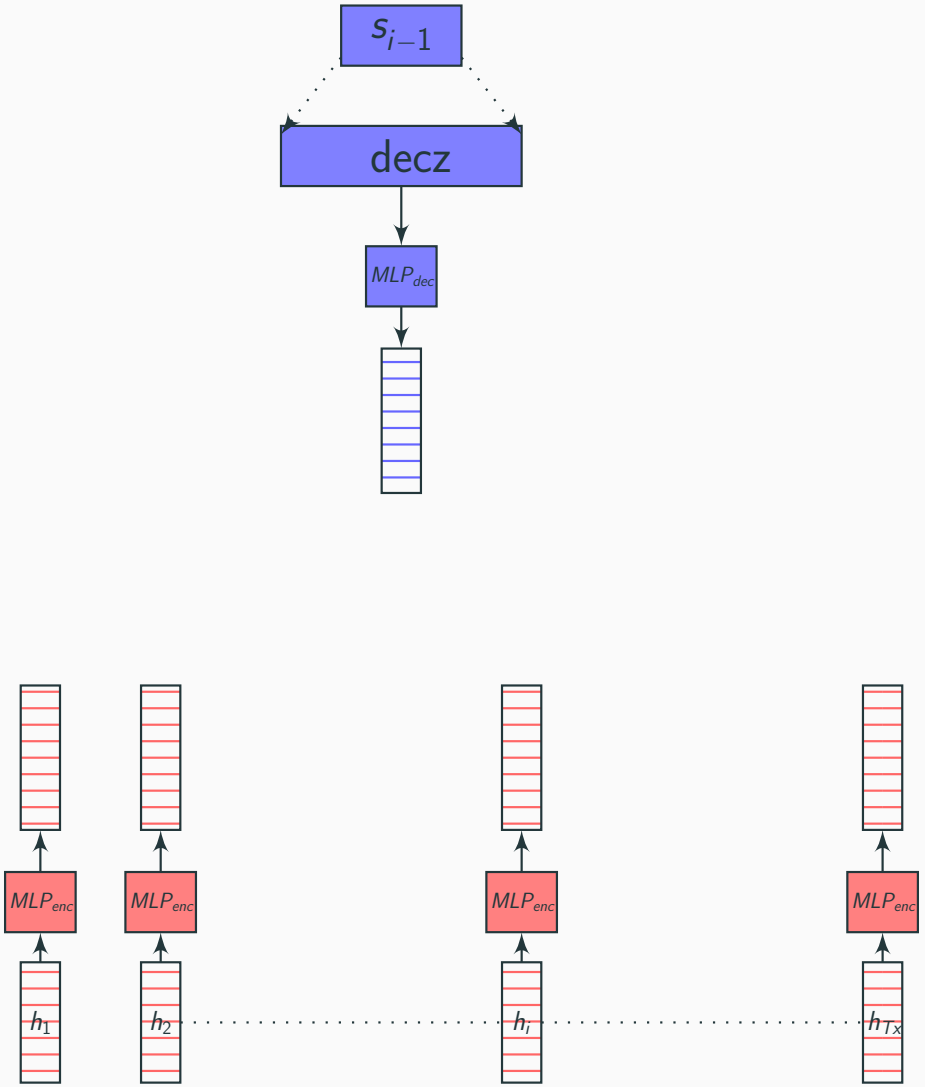


Dot product Attention - Full picture

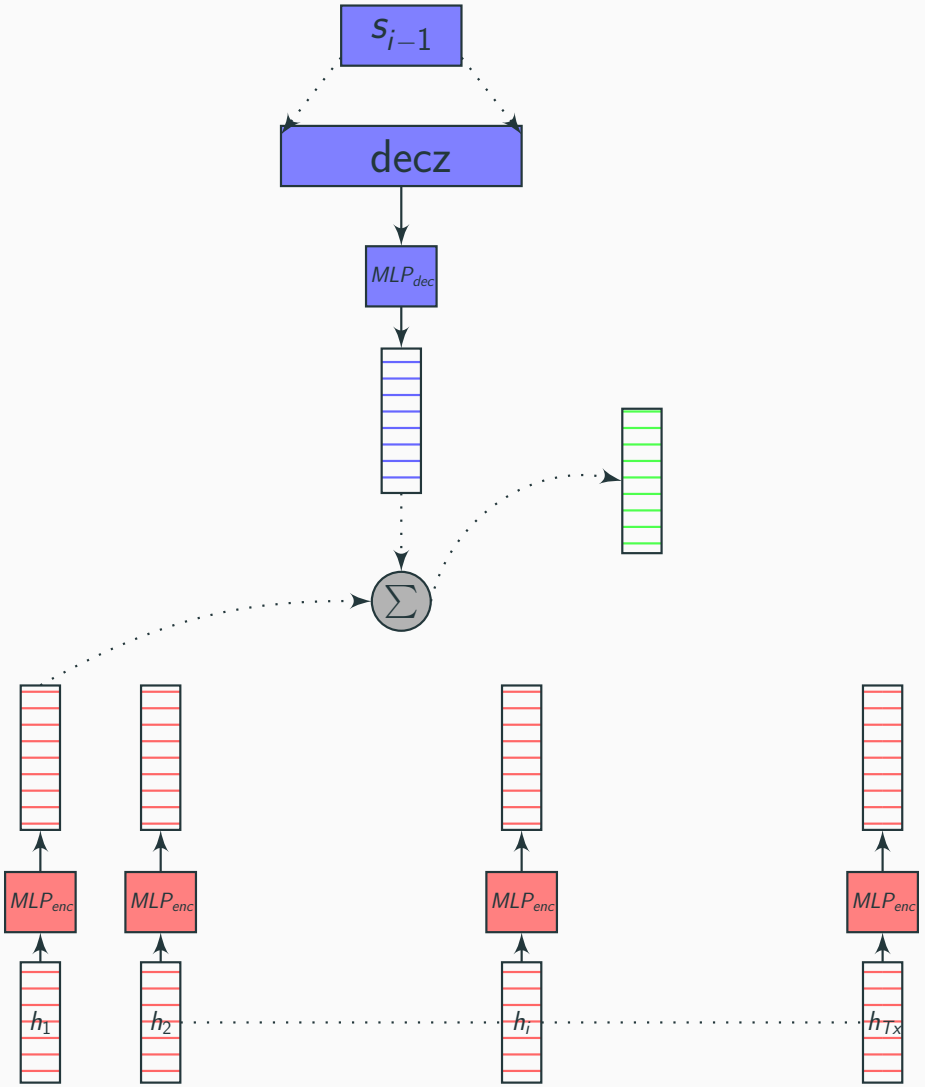


Additive Attention

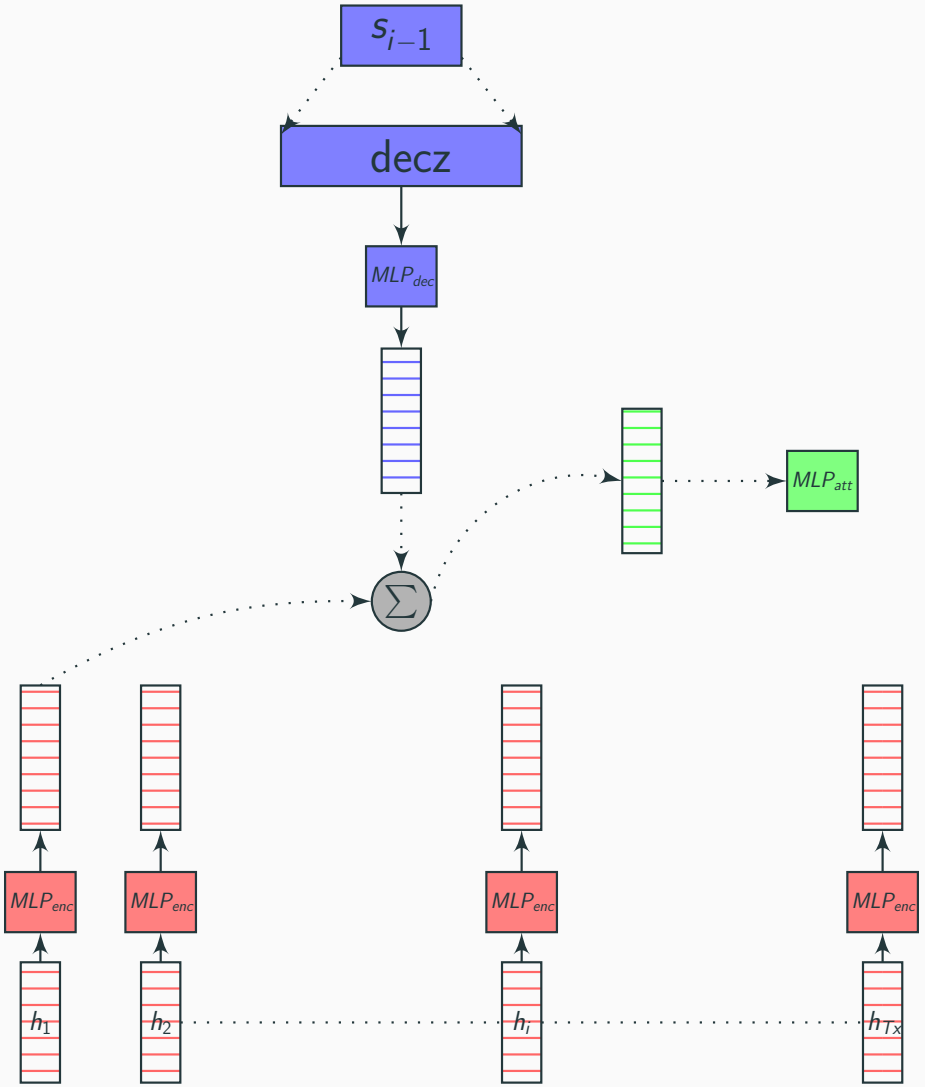
Additive Attention



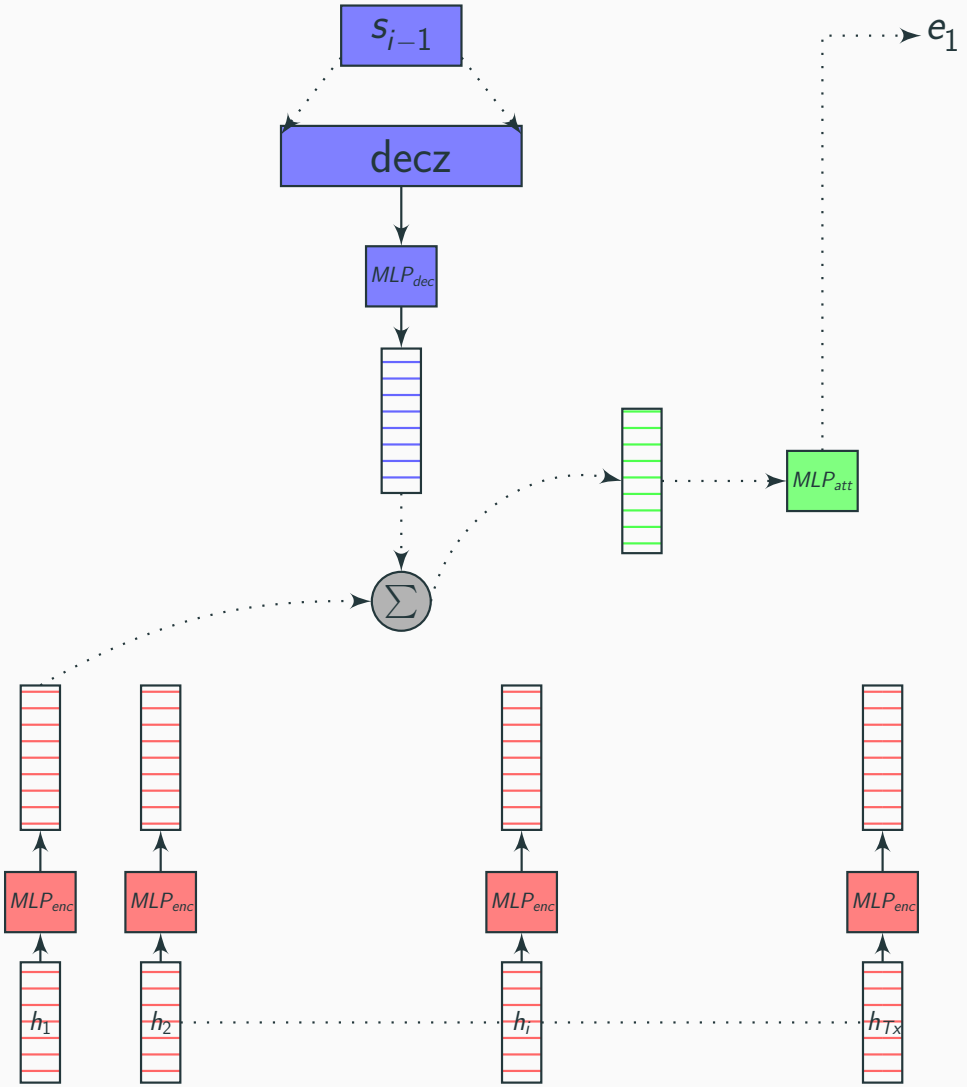
Additive Attention



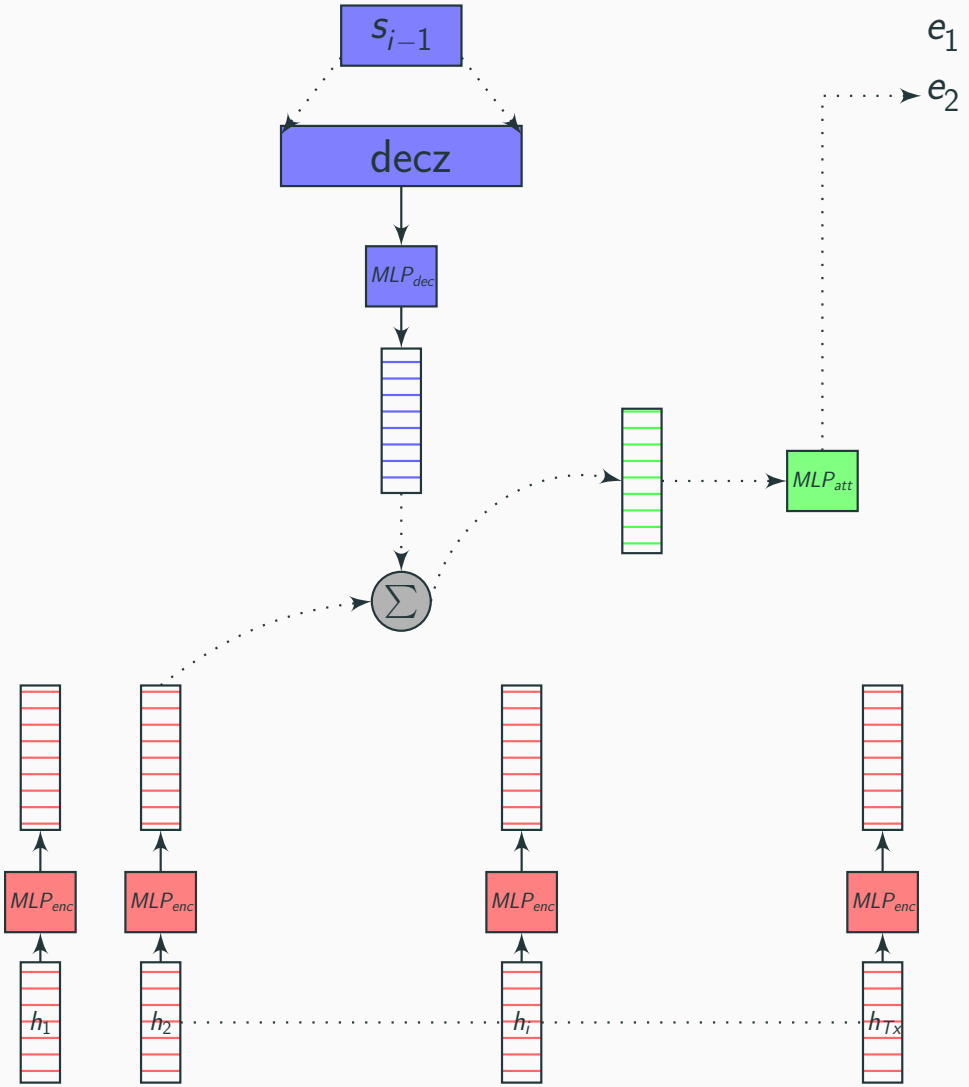
Additive Attention



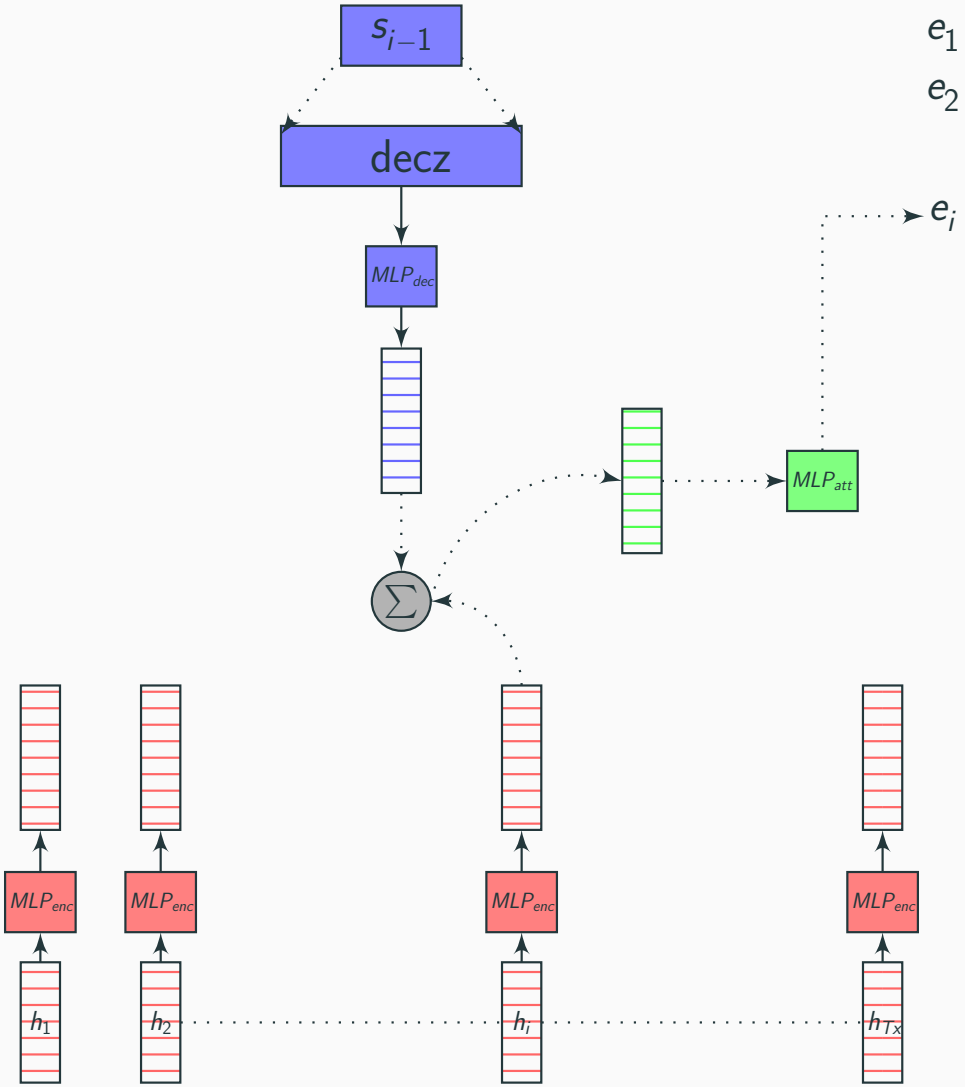
Additive Attention



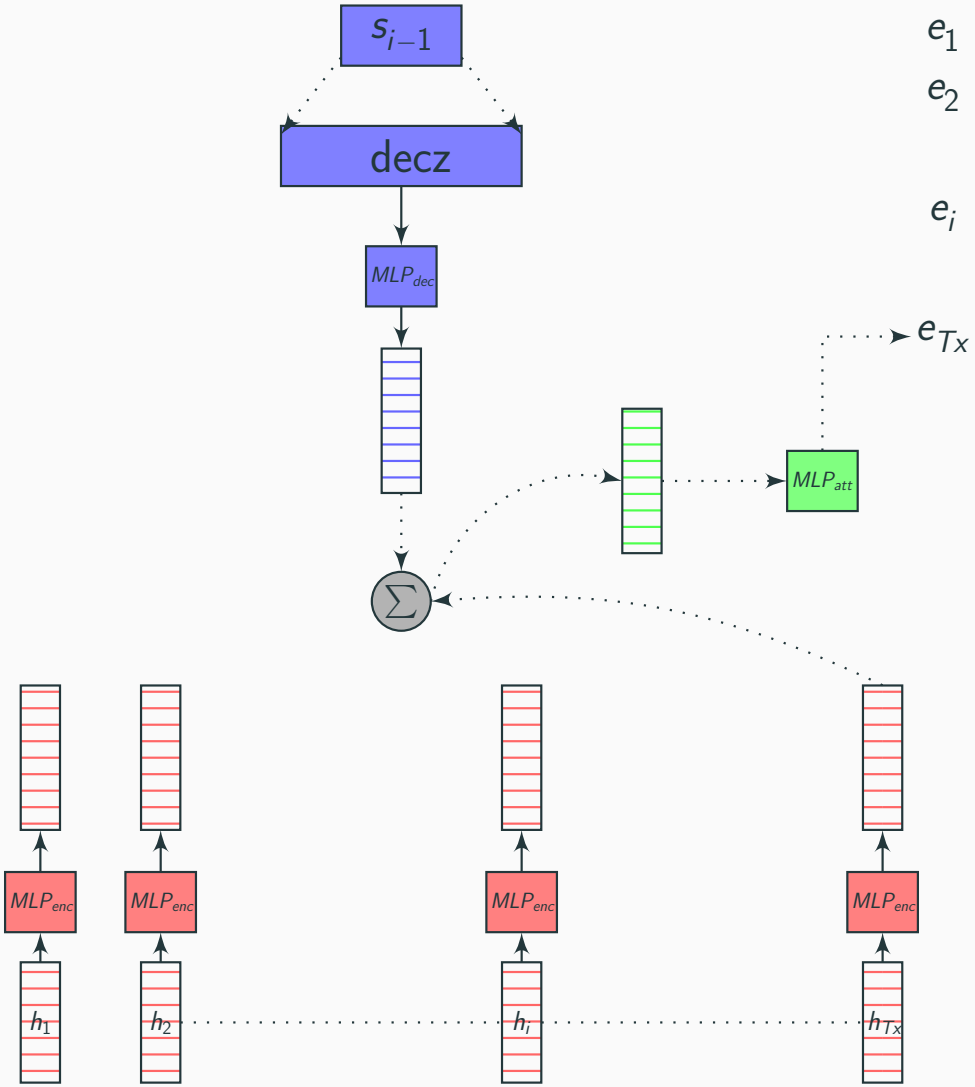
Additive Attention



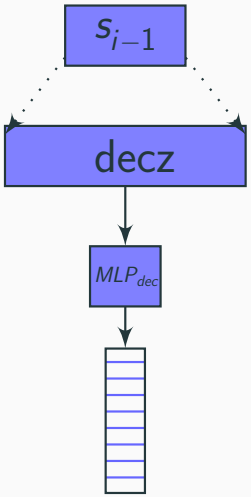
Additive Attention

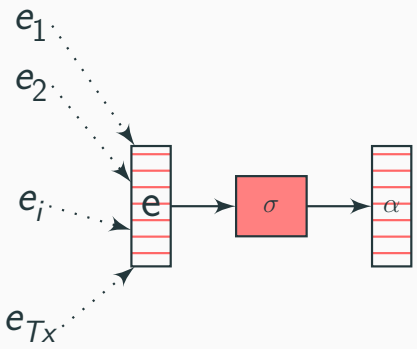


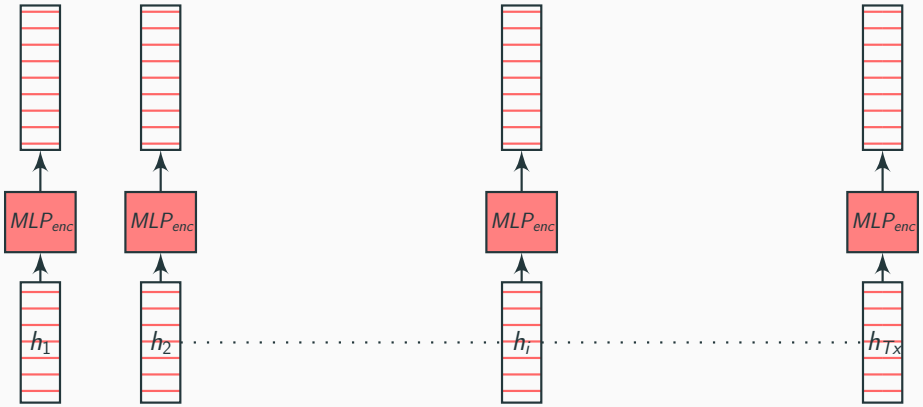
Additive Attention



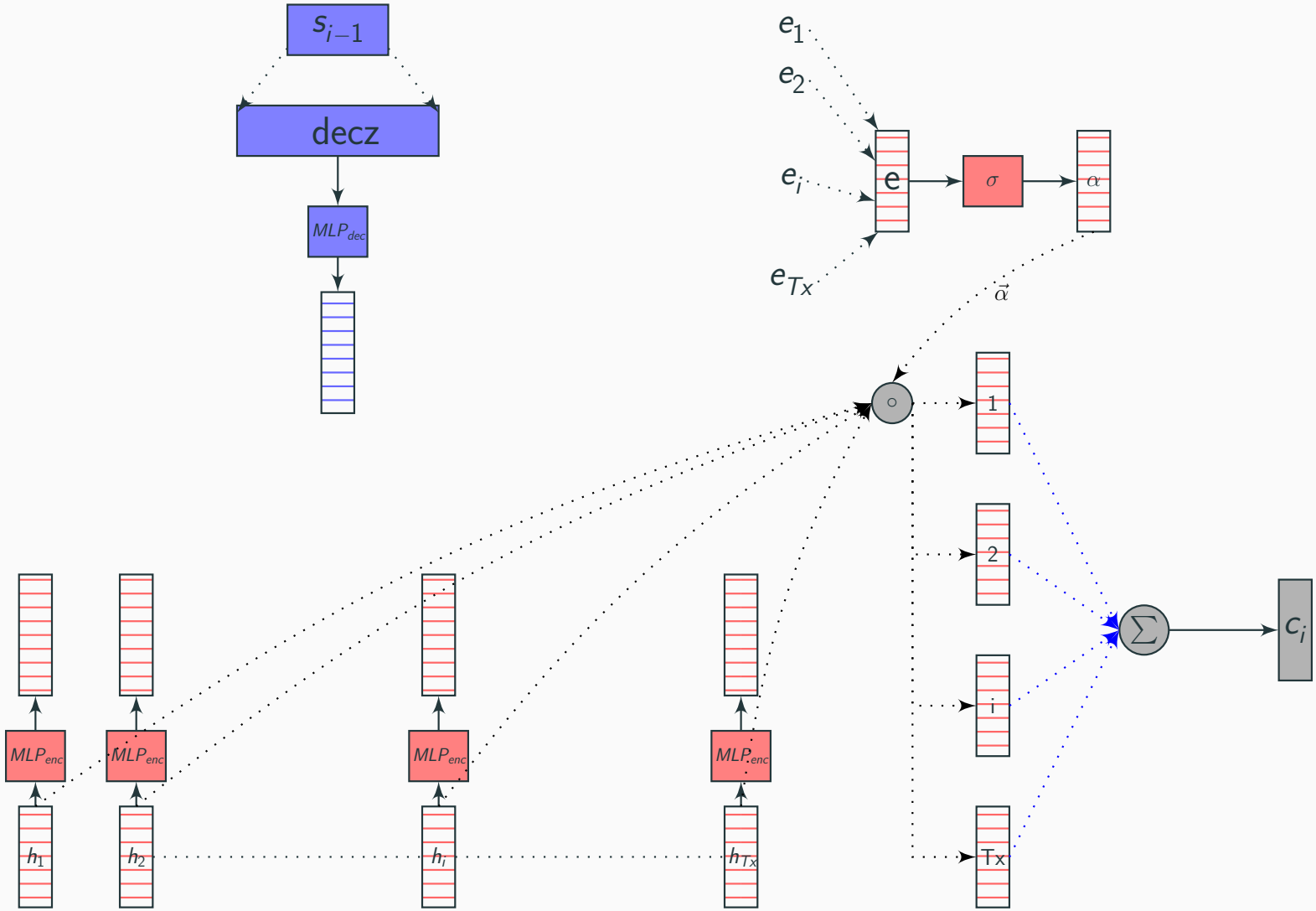
Additive Attention



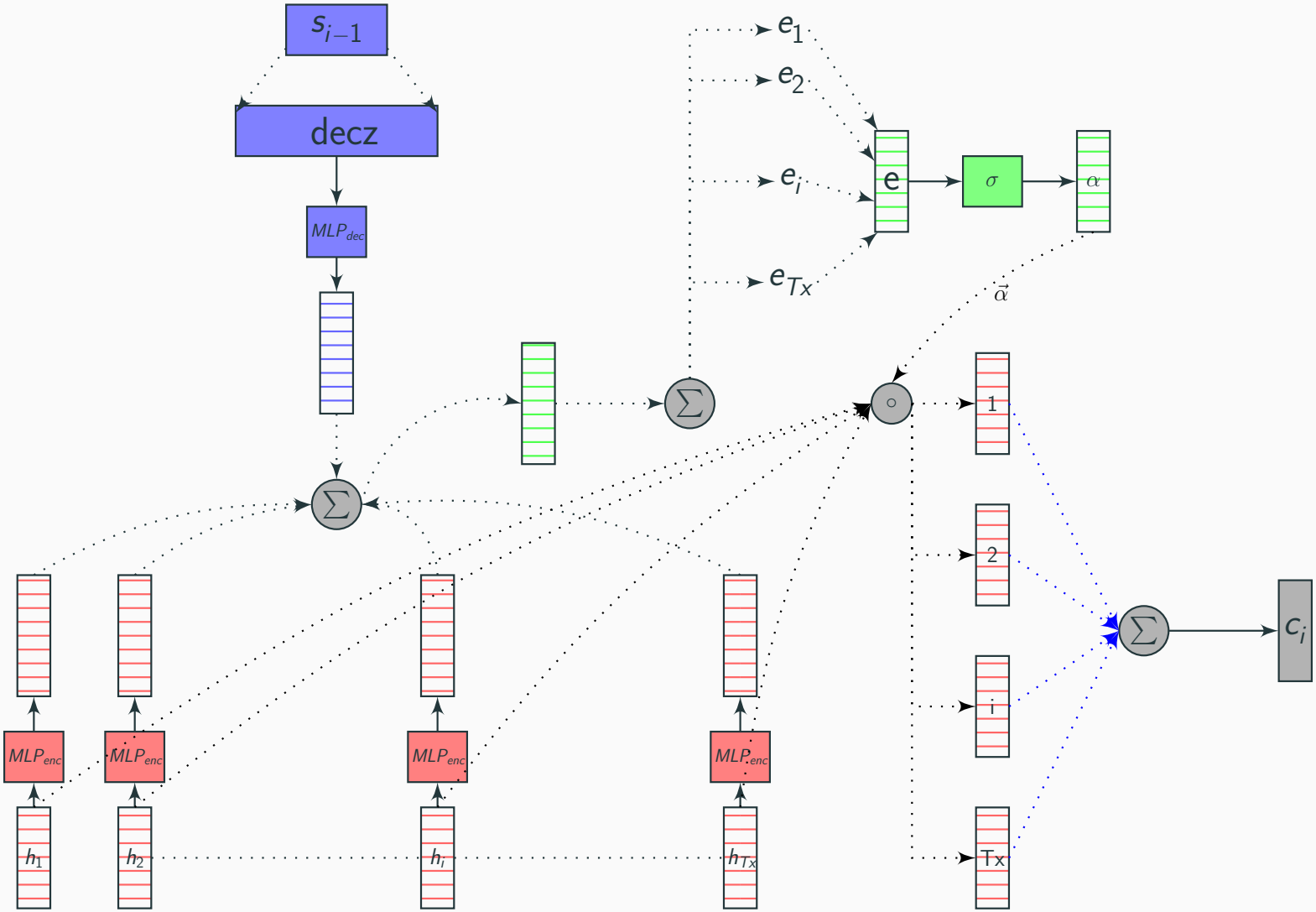




Additive Attention

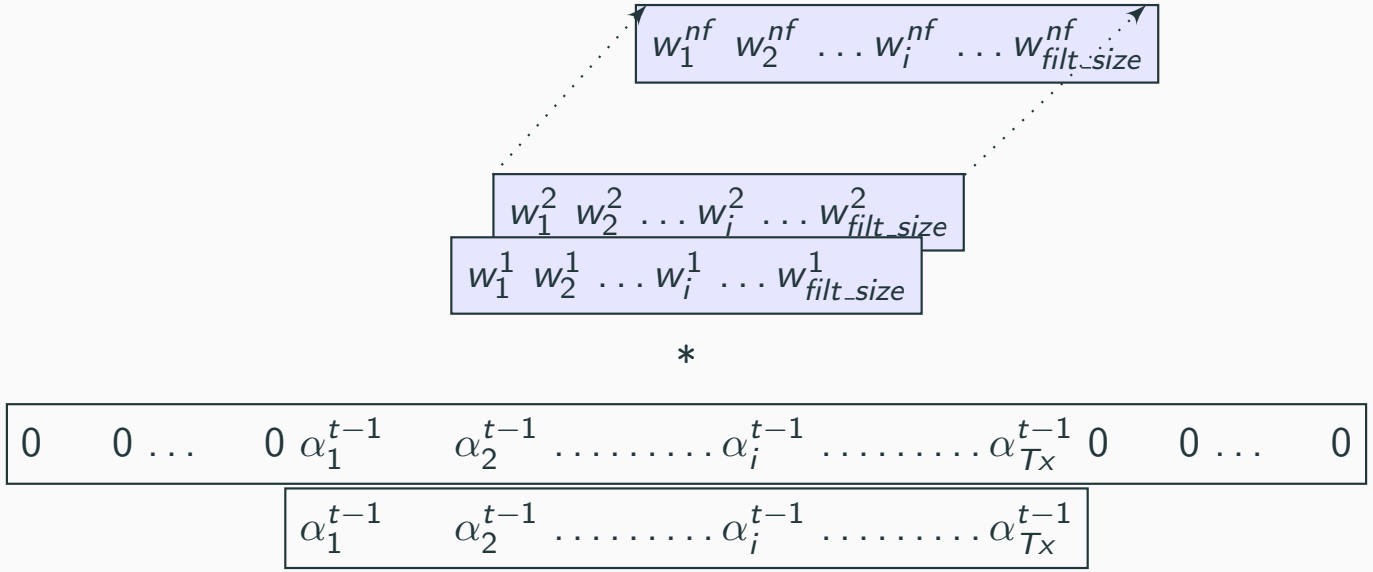


Additive Attention - Full picture

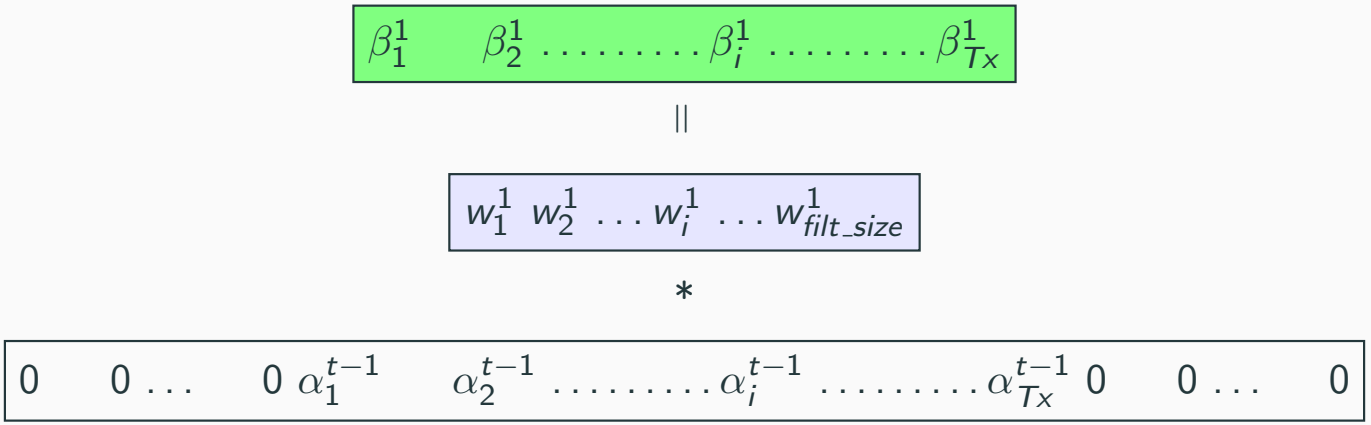


Location Aware Attention

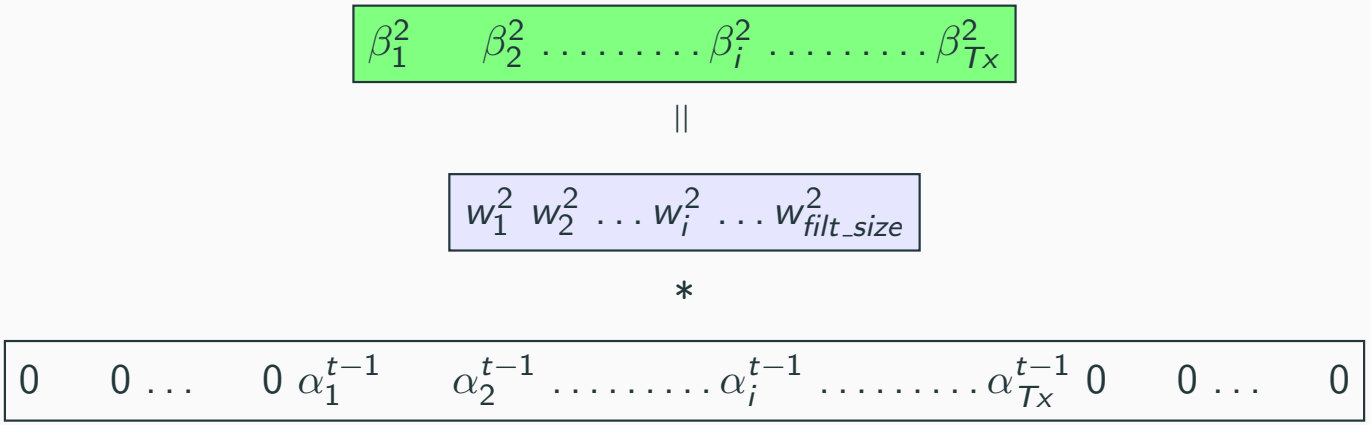
Location Aware Attention



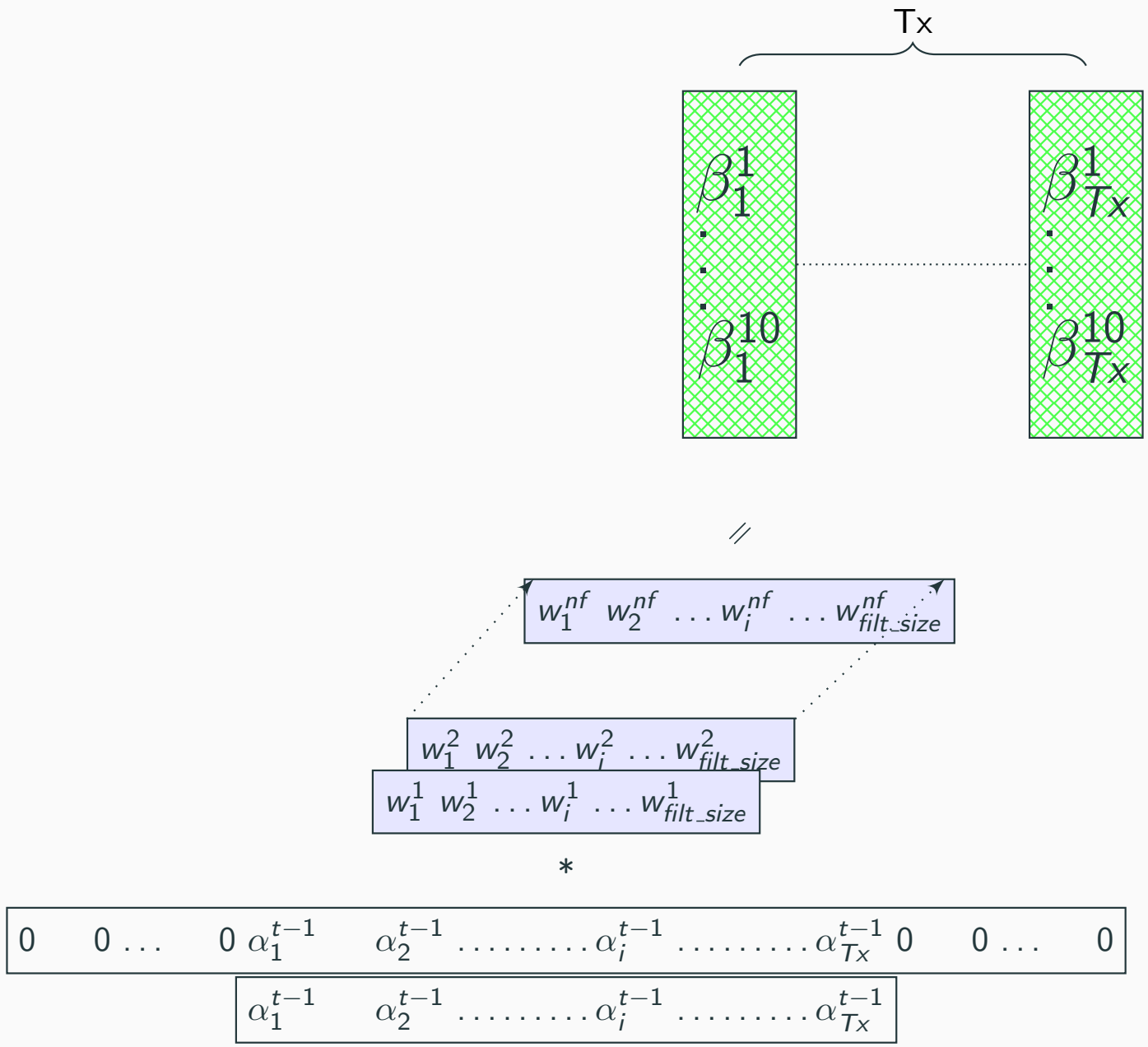
Location Aware Attention



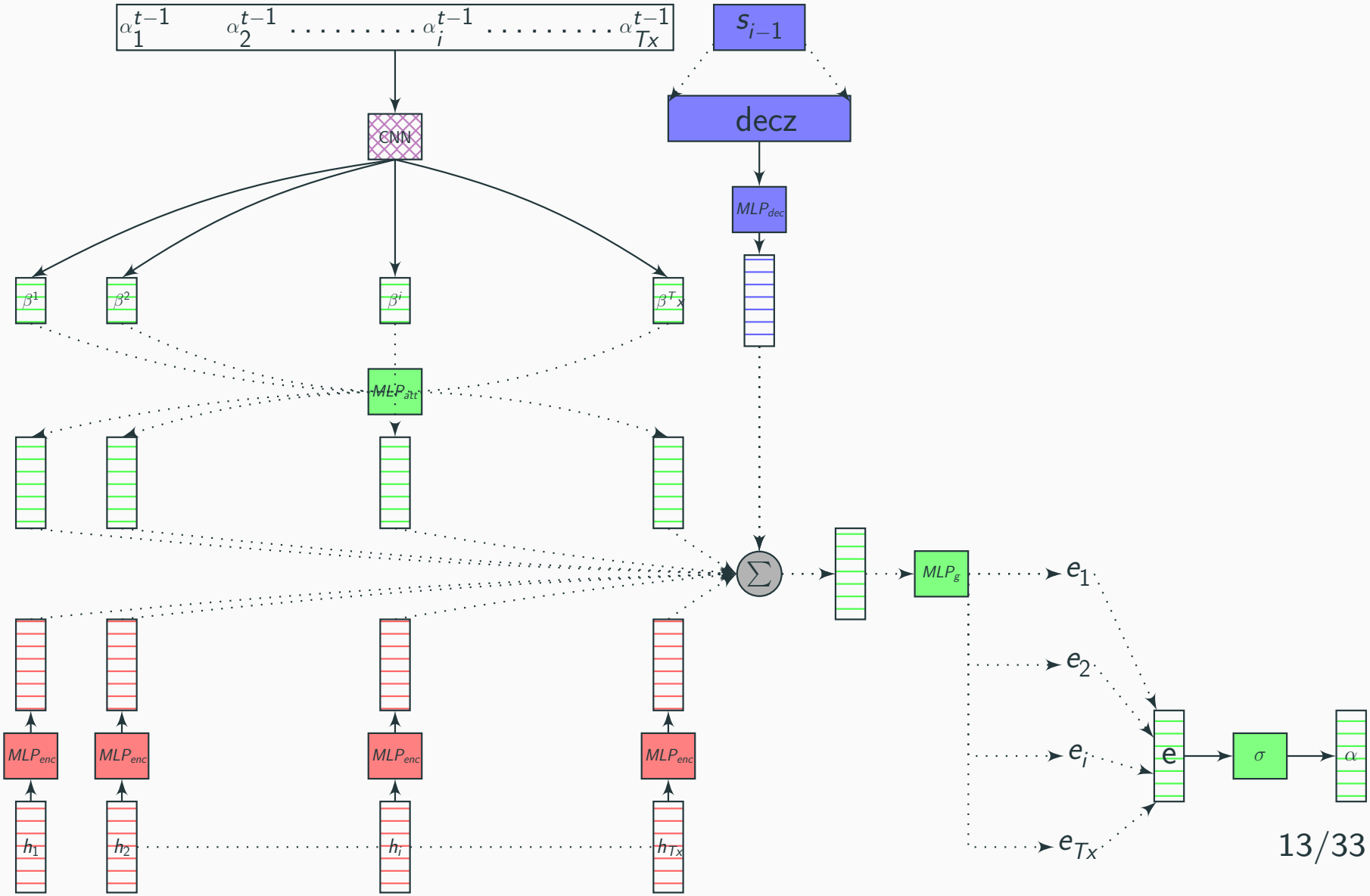
Location Aware Attention



Location Aware Attention

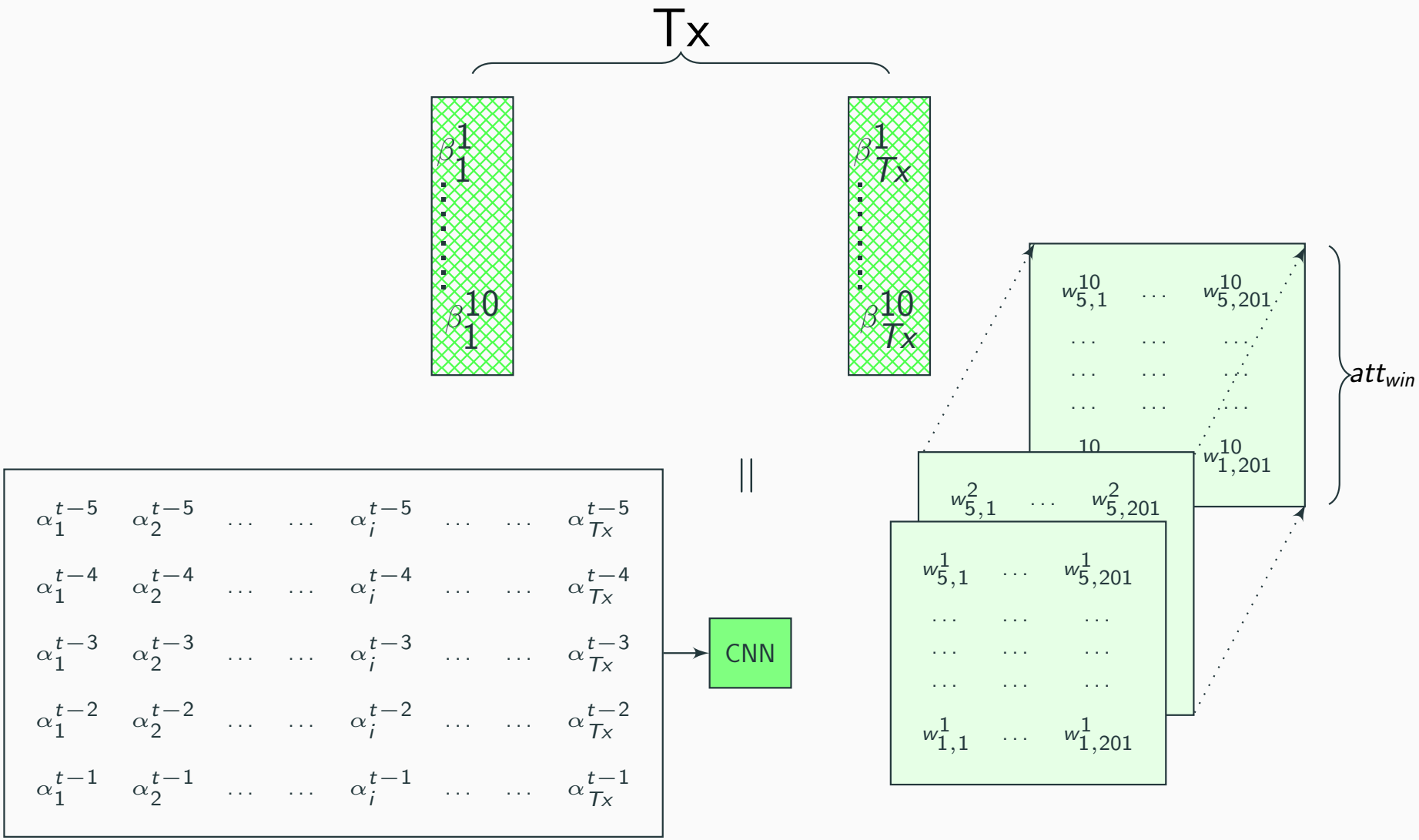


Location Aware Attention - Full picture

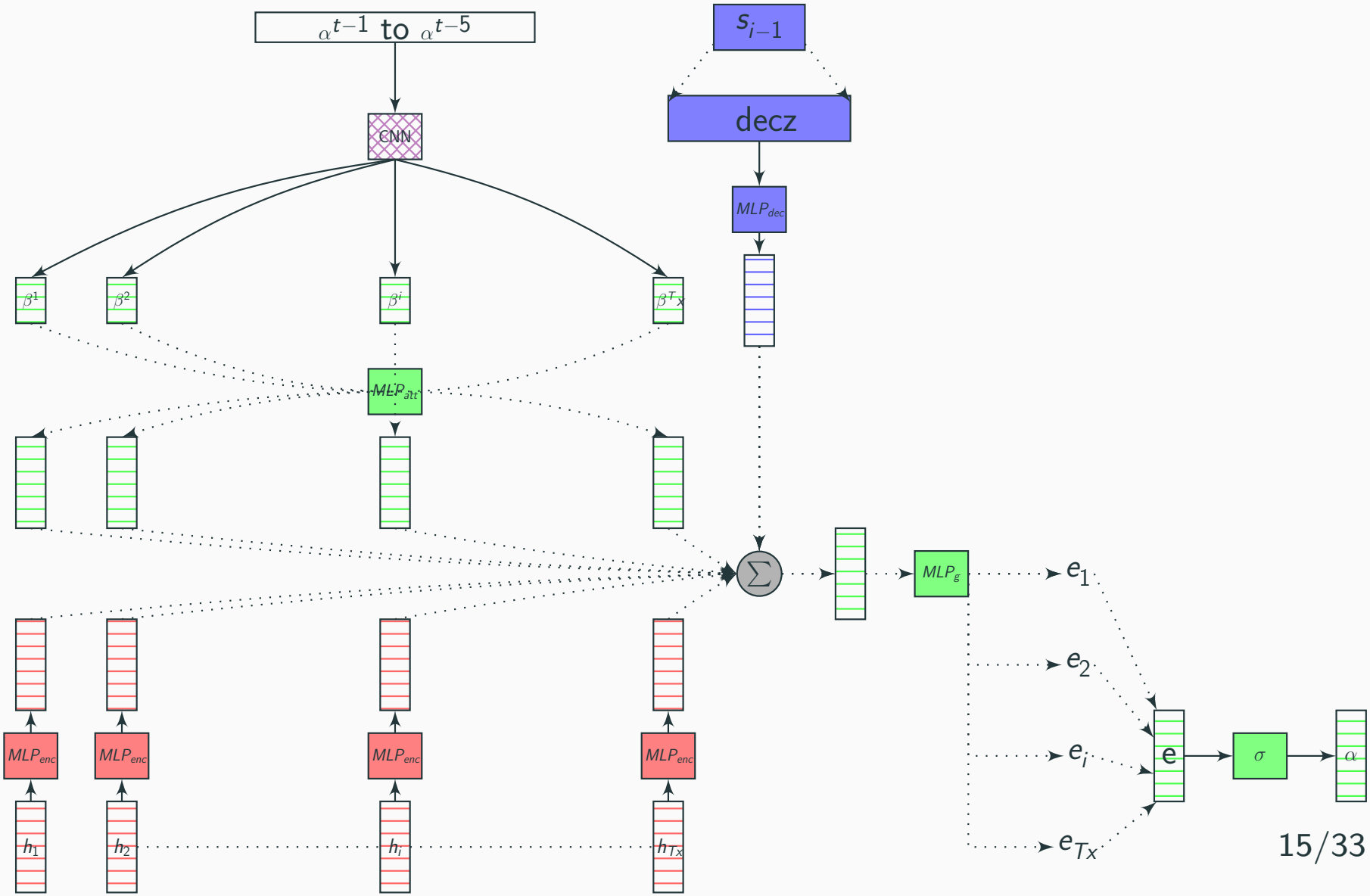


2D Location Aware Attention

2D Location Aware Attention

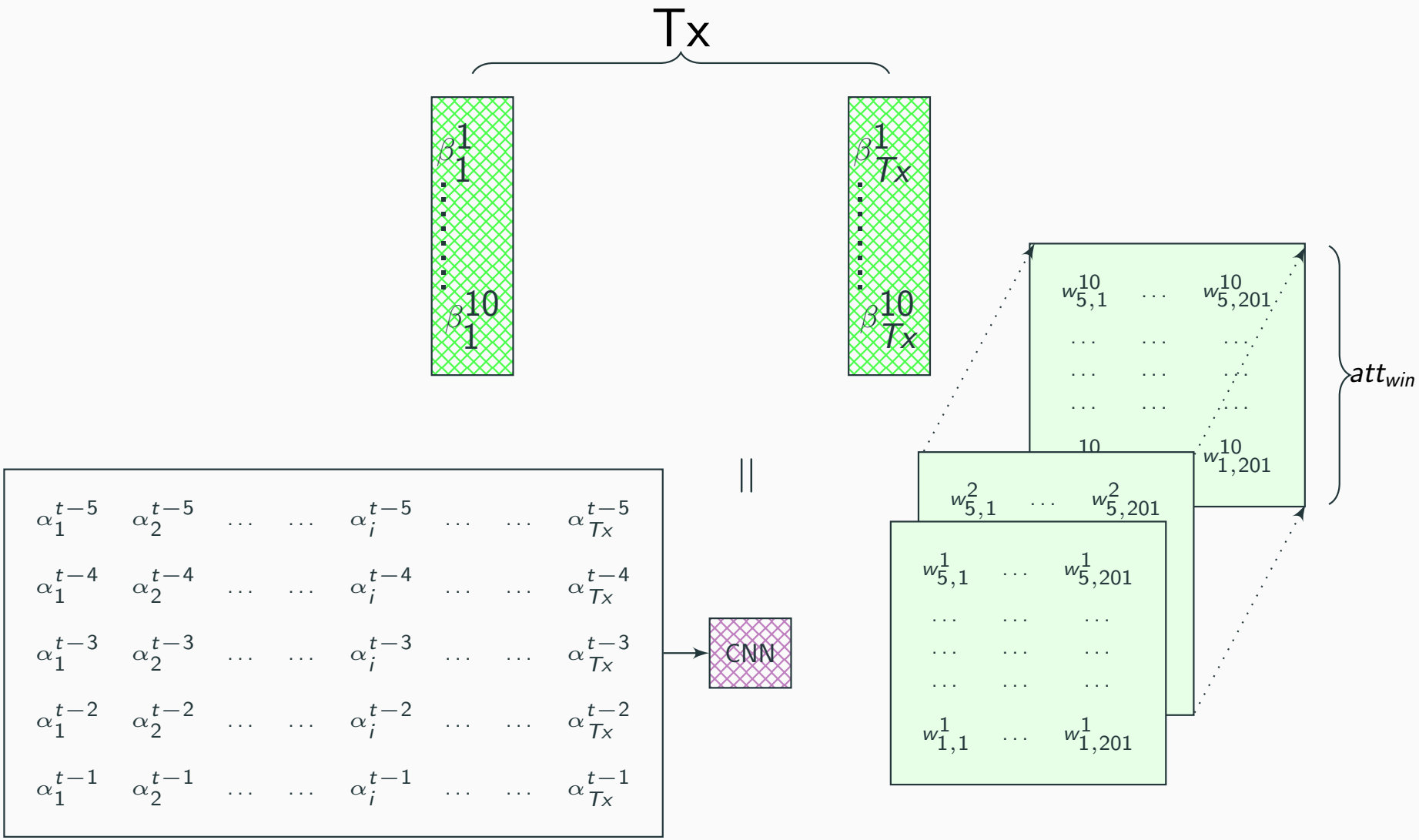


2D Location Aware Attention - Full picture

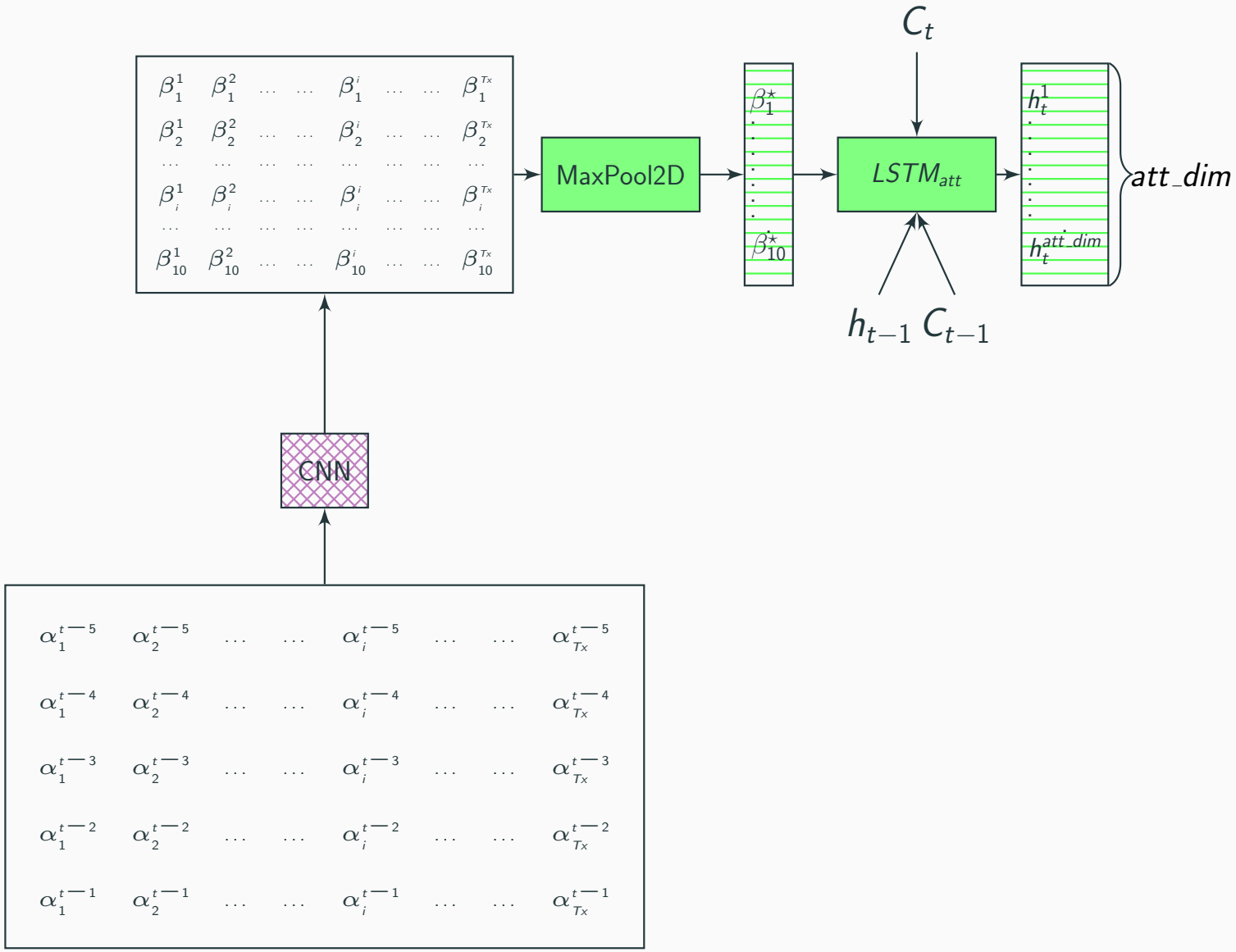


Location Aware Recurrent Attention

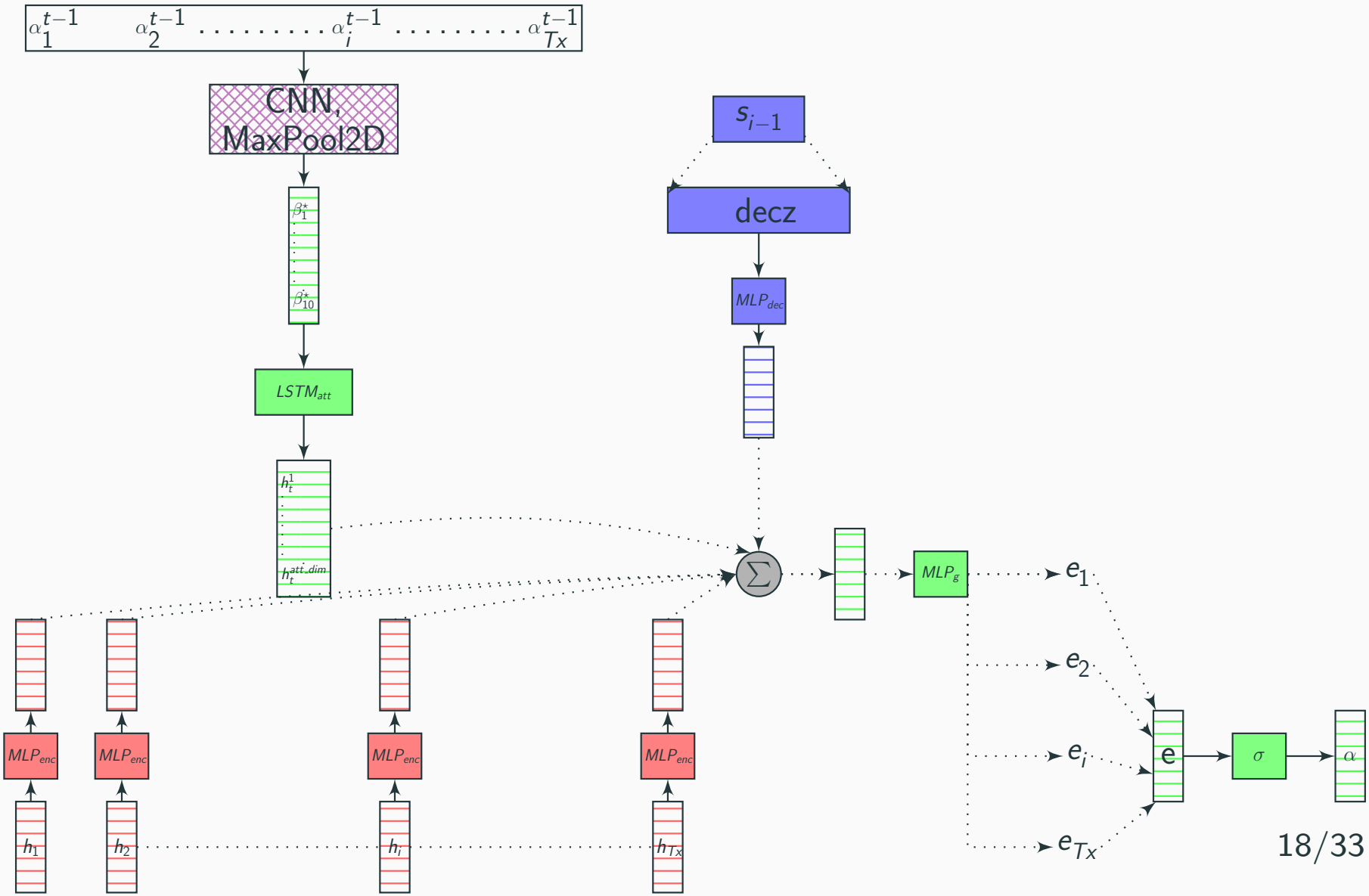
Location Aware Recurrent Attention



Location Aware Recurrent Attention - weights



Location Aware Recurrent Attention - Full picture

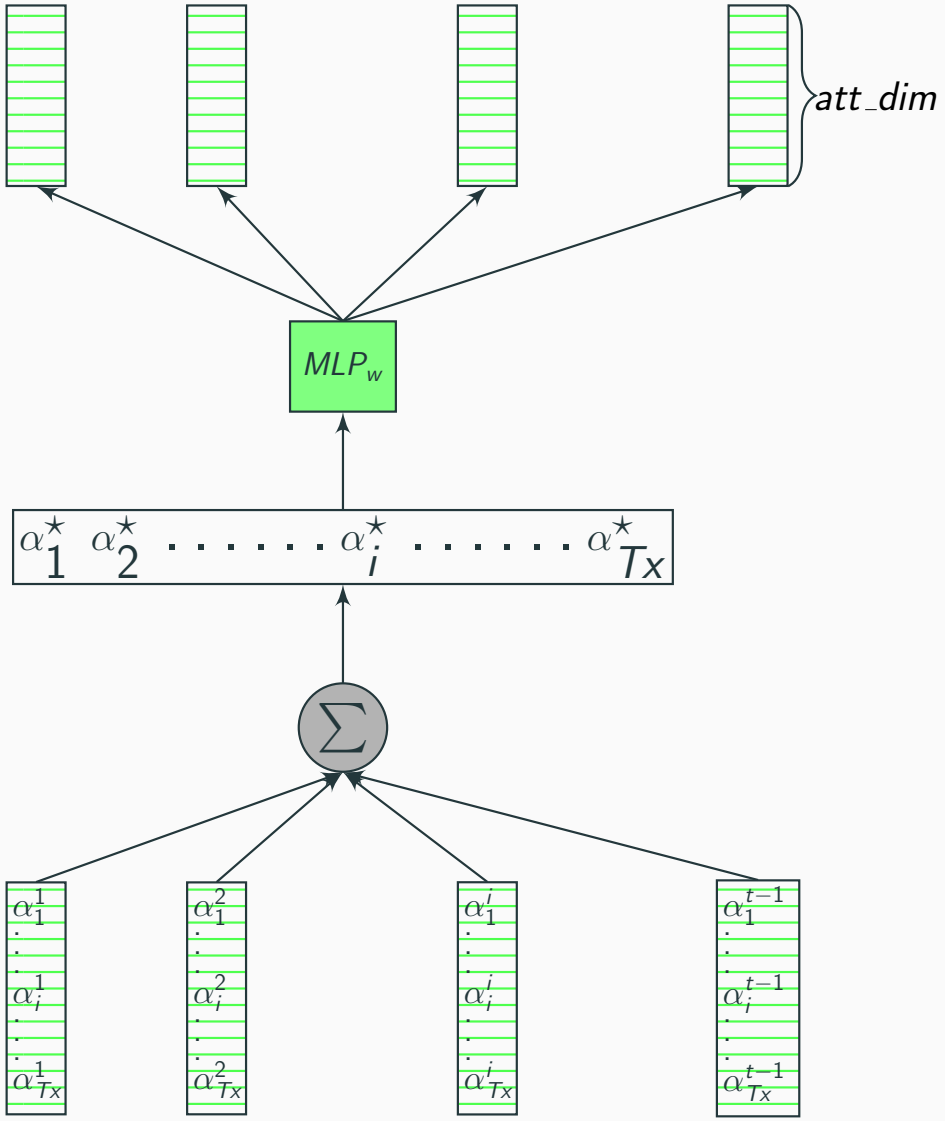


Coverage mechanism Attention

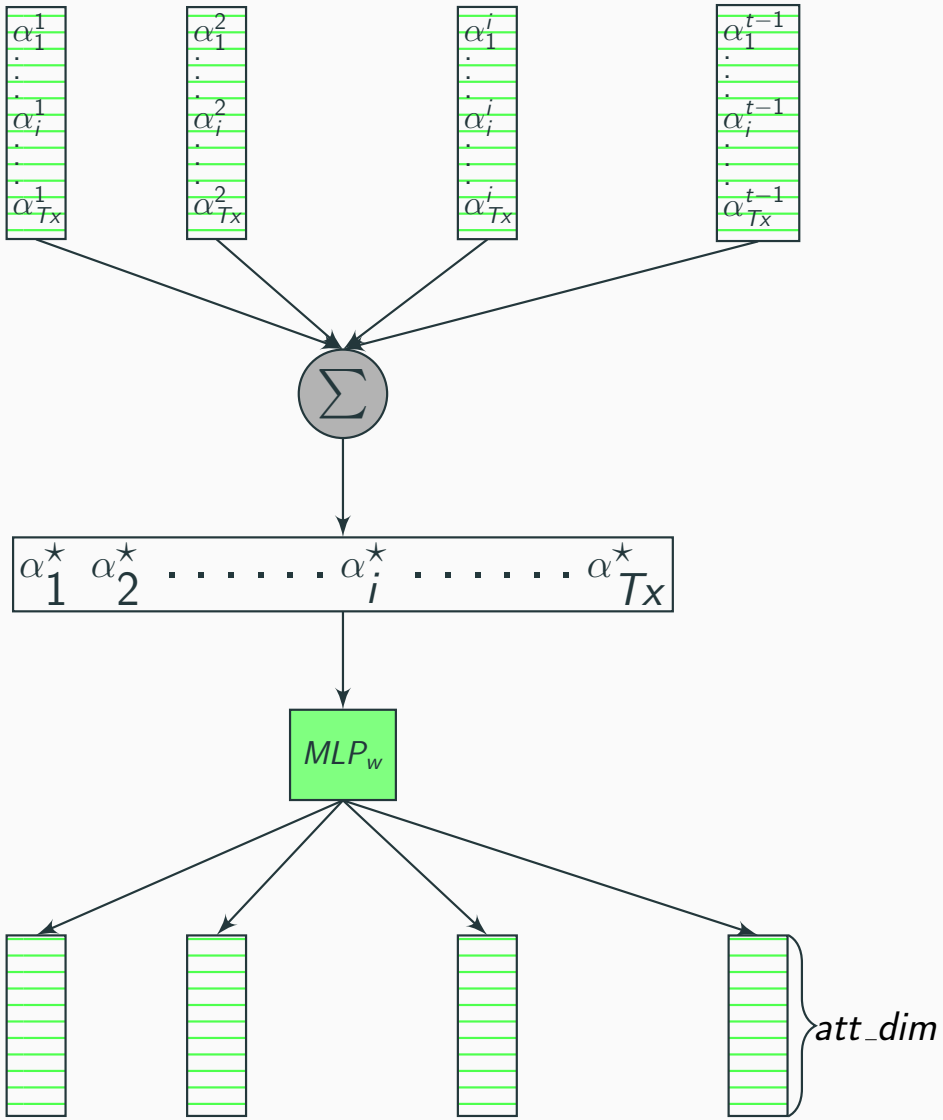
Coverage mechanism Attention

- Text summarization - Seq-to-Seq models
- Not reliable in producing factual details correctly
- Extend the standard seq-to-seq attention models
  - Hybrid pointer-generator network
  - Coverage

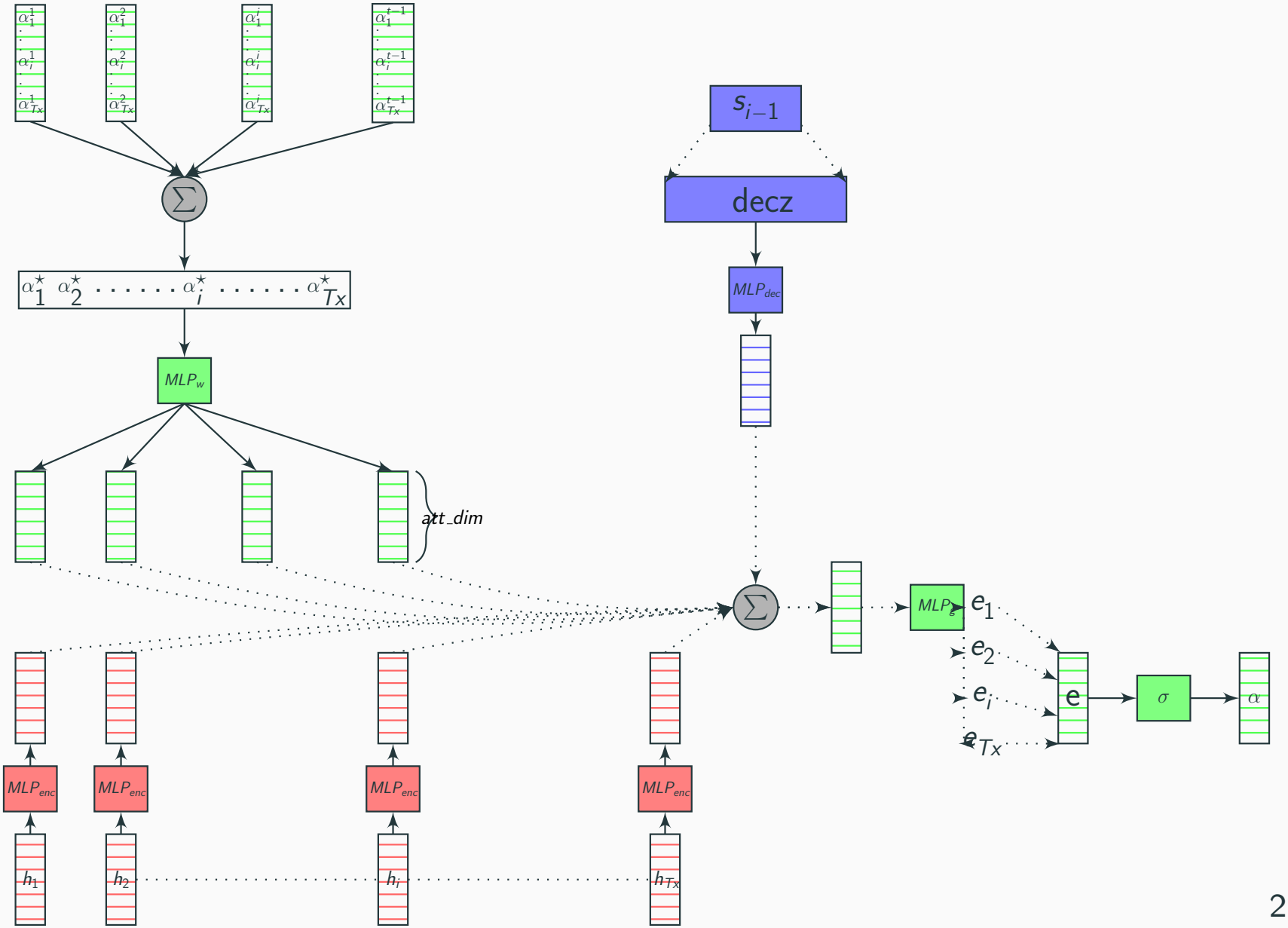
Coverage mechanism Attention



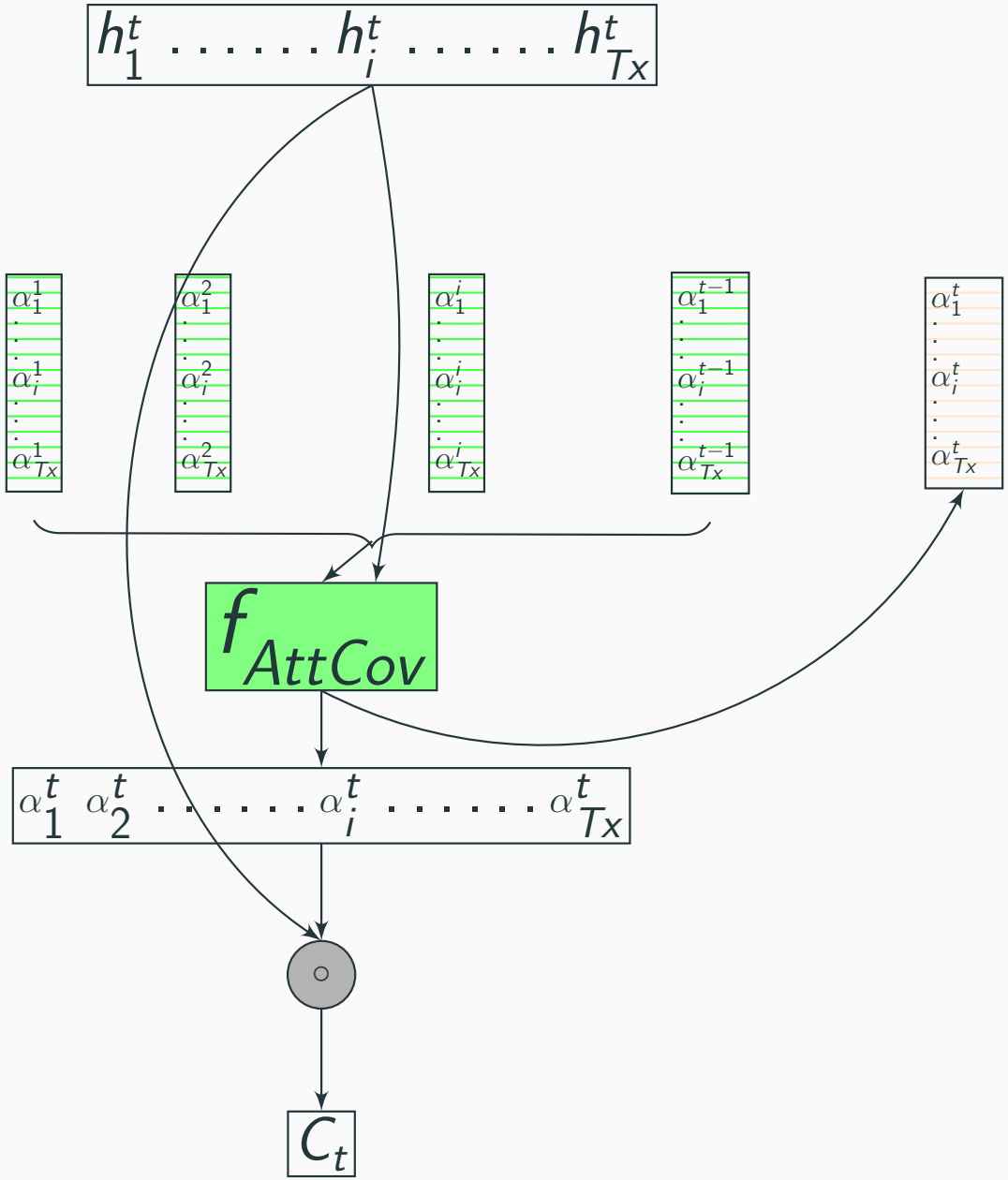
Coverage mechanism Attention



Coverage mechanism Attention

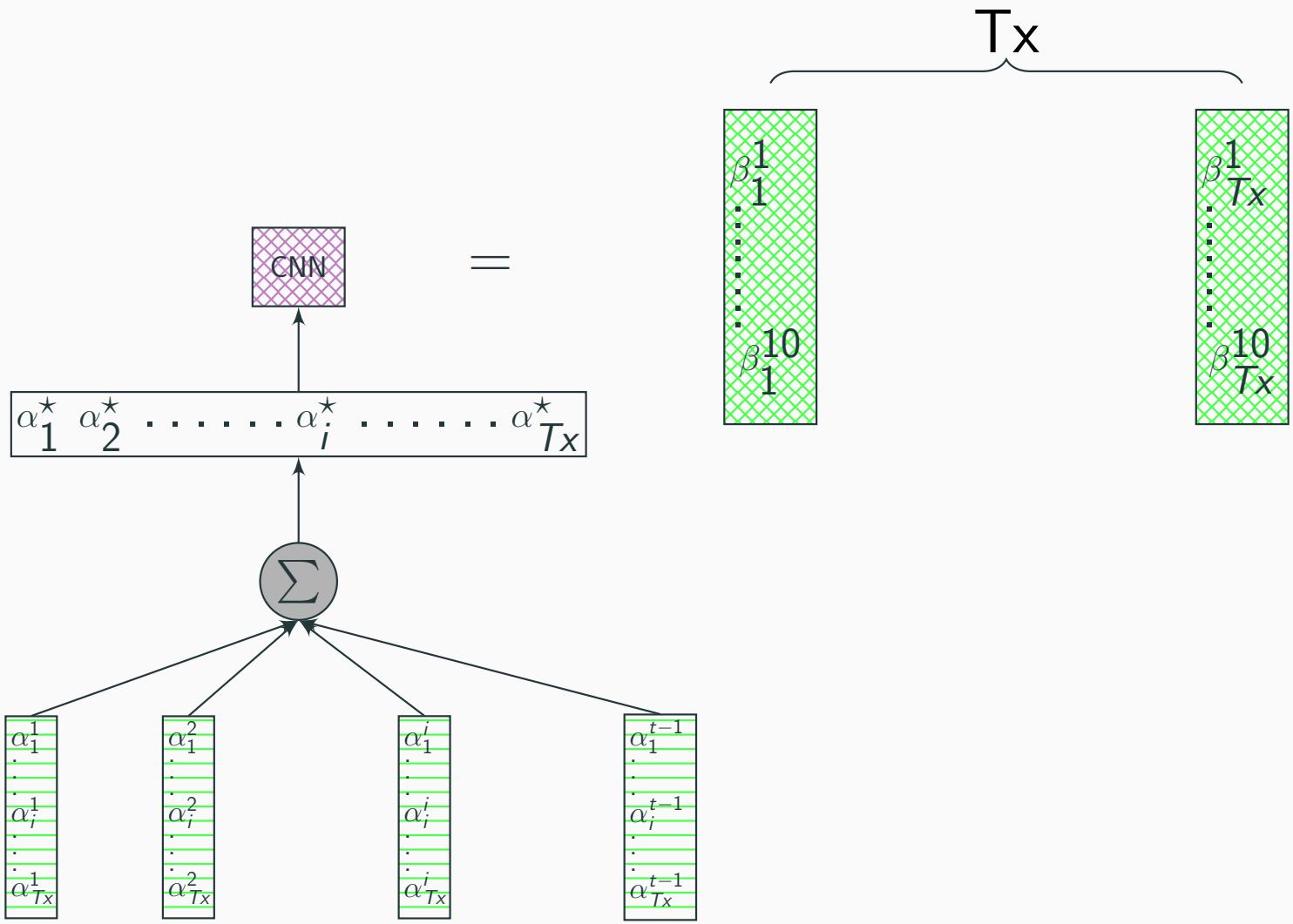


Coverage mechanism Attention

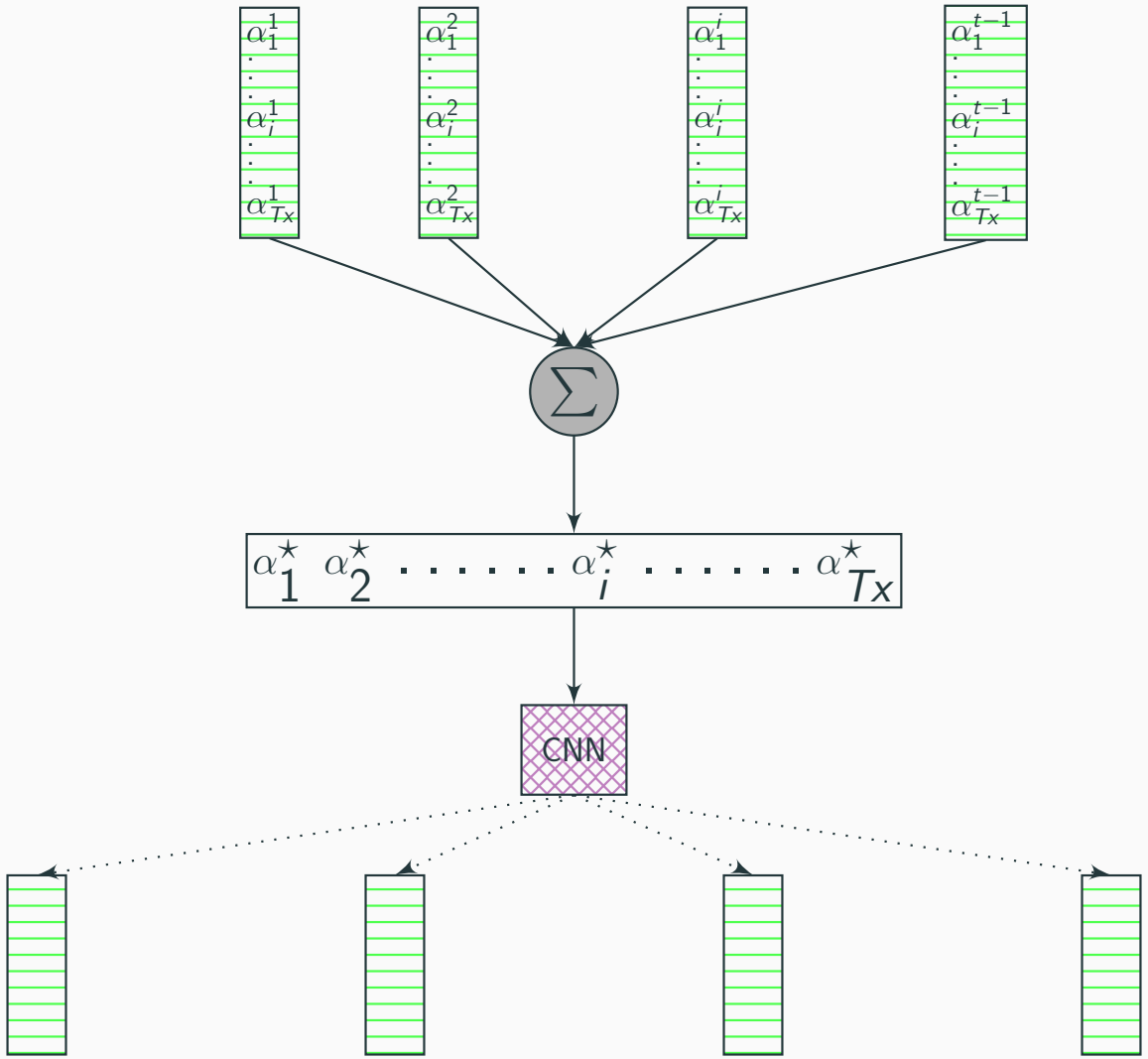


Coverage mechanism location aware Attention

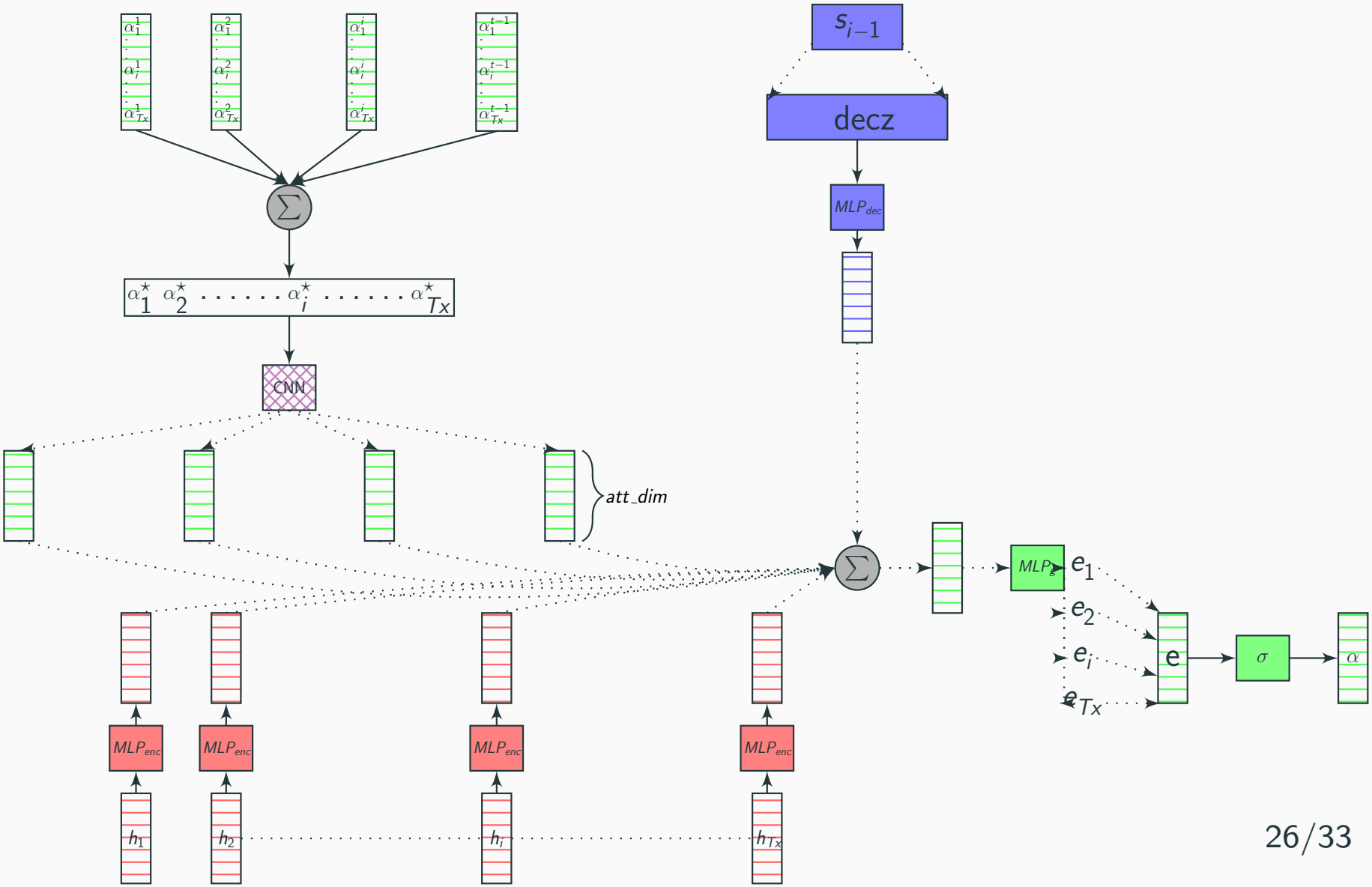
Coverage mechanism location aware Attention



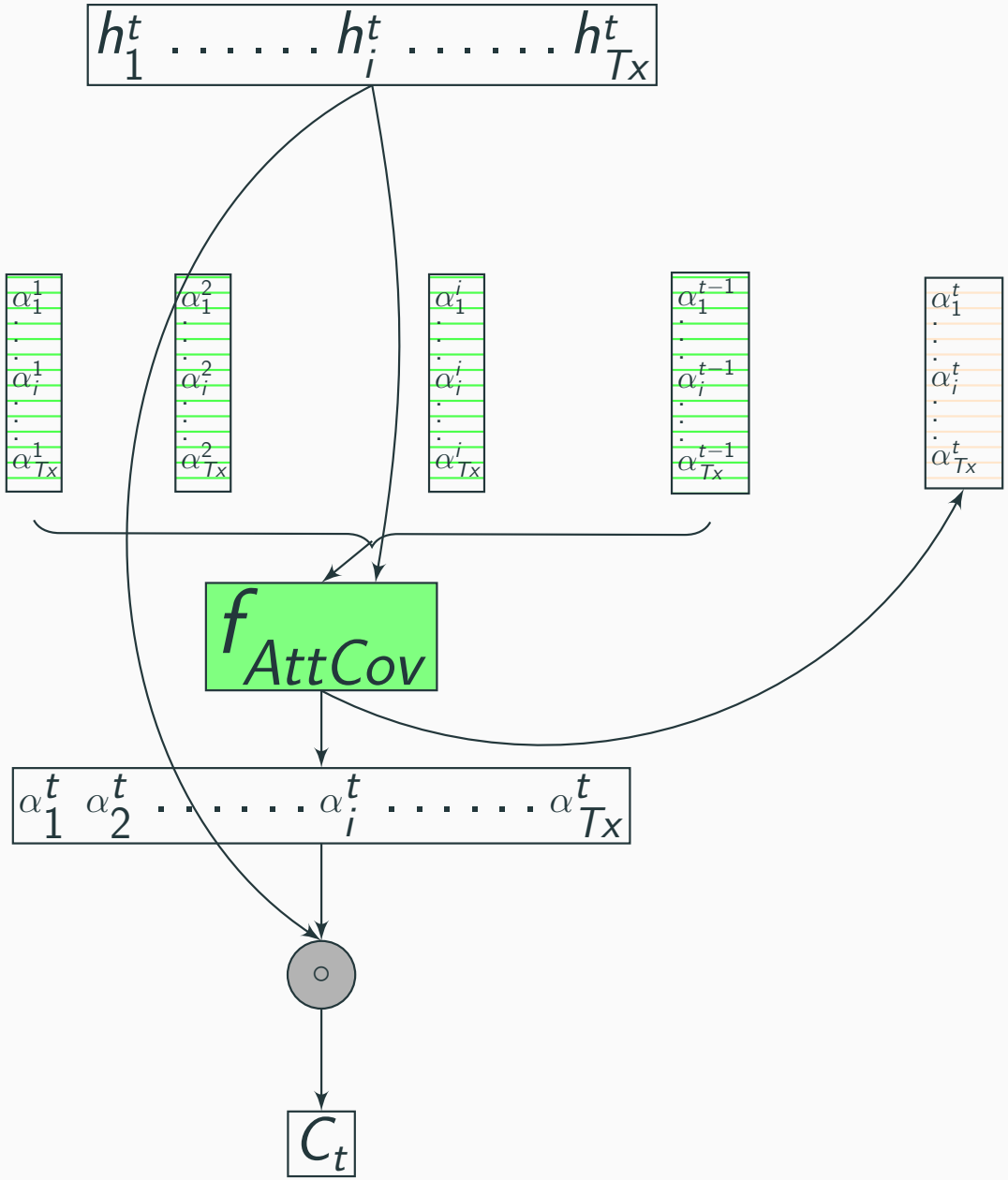
Coverage mechanism location aware Attention



Coverage mechanism location aware Attention

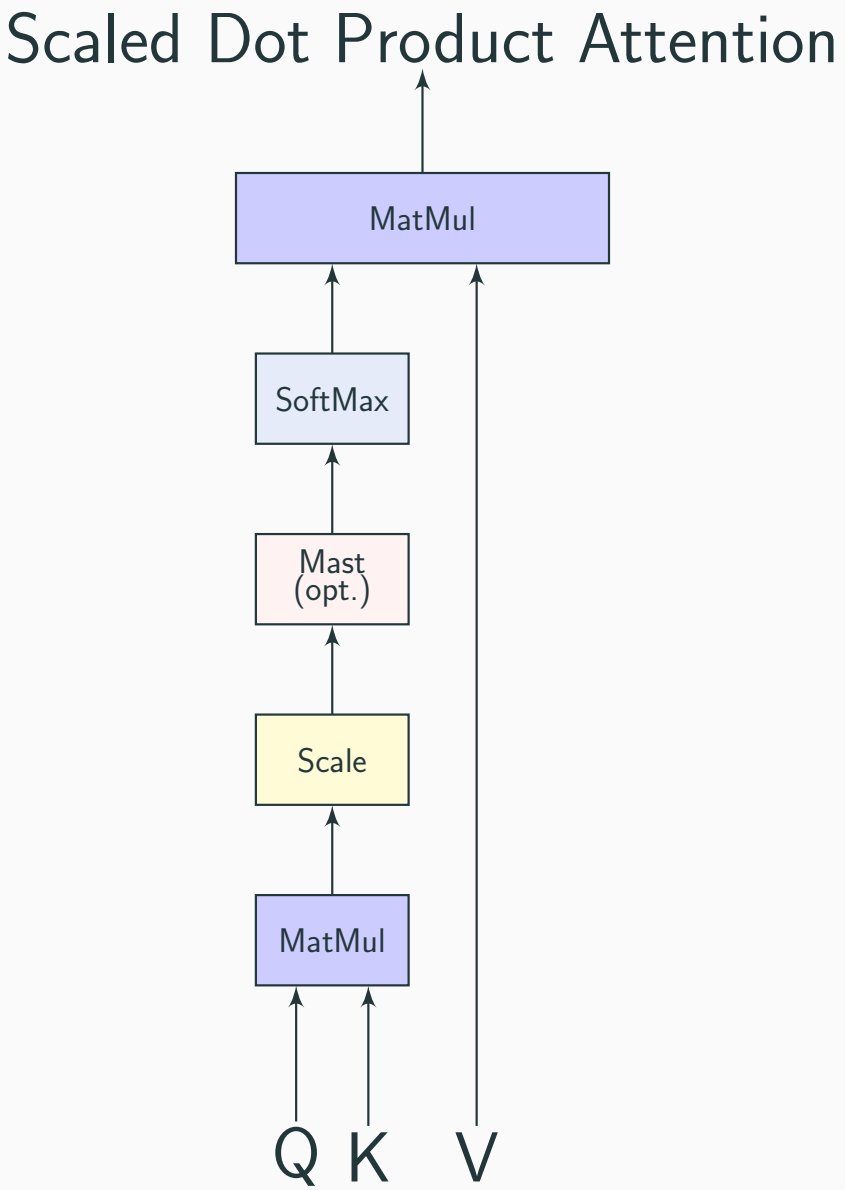


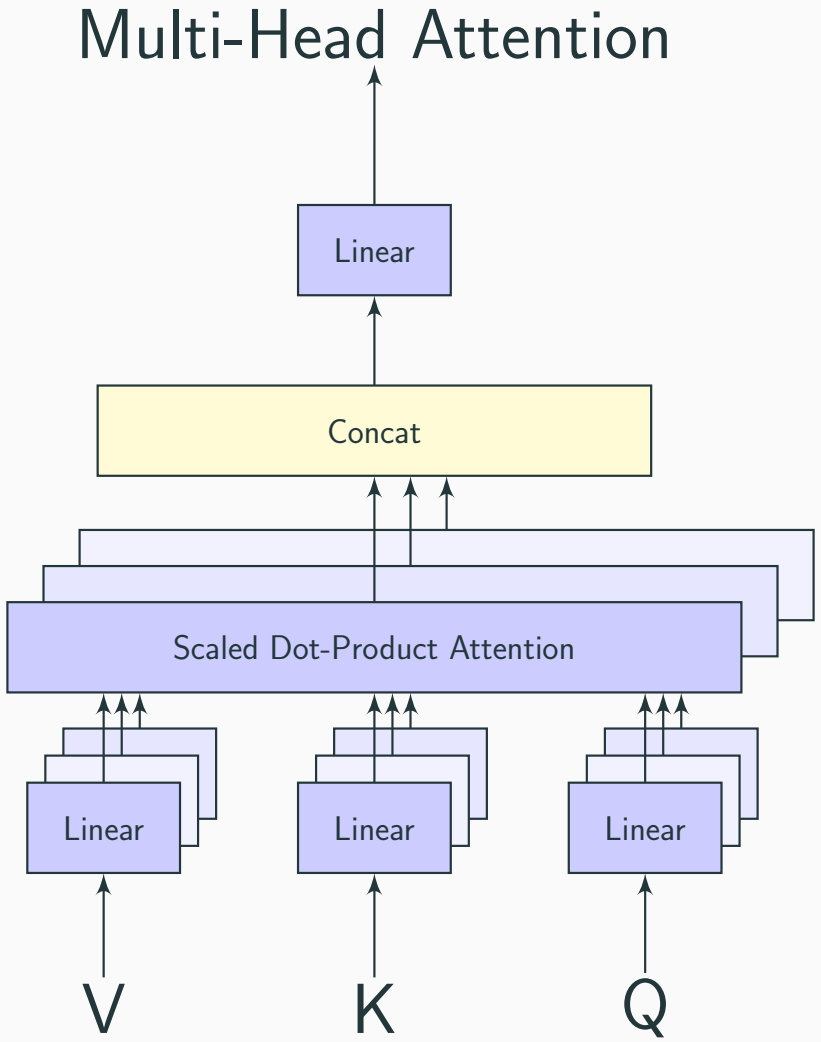
Coverage mechanism location aware Attention



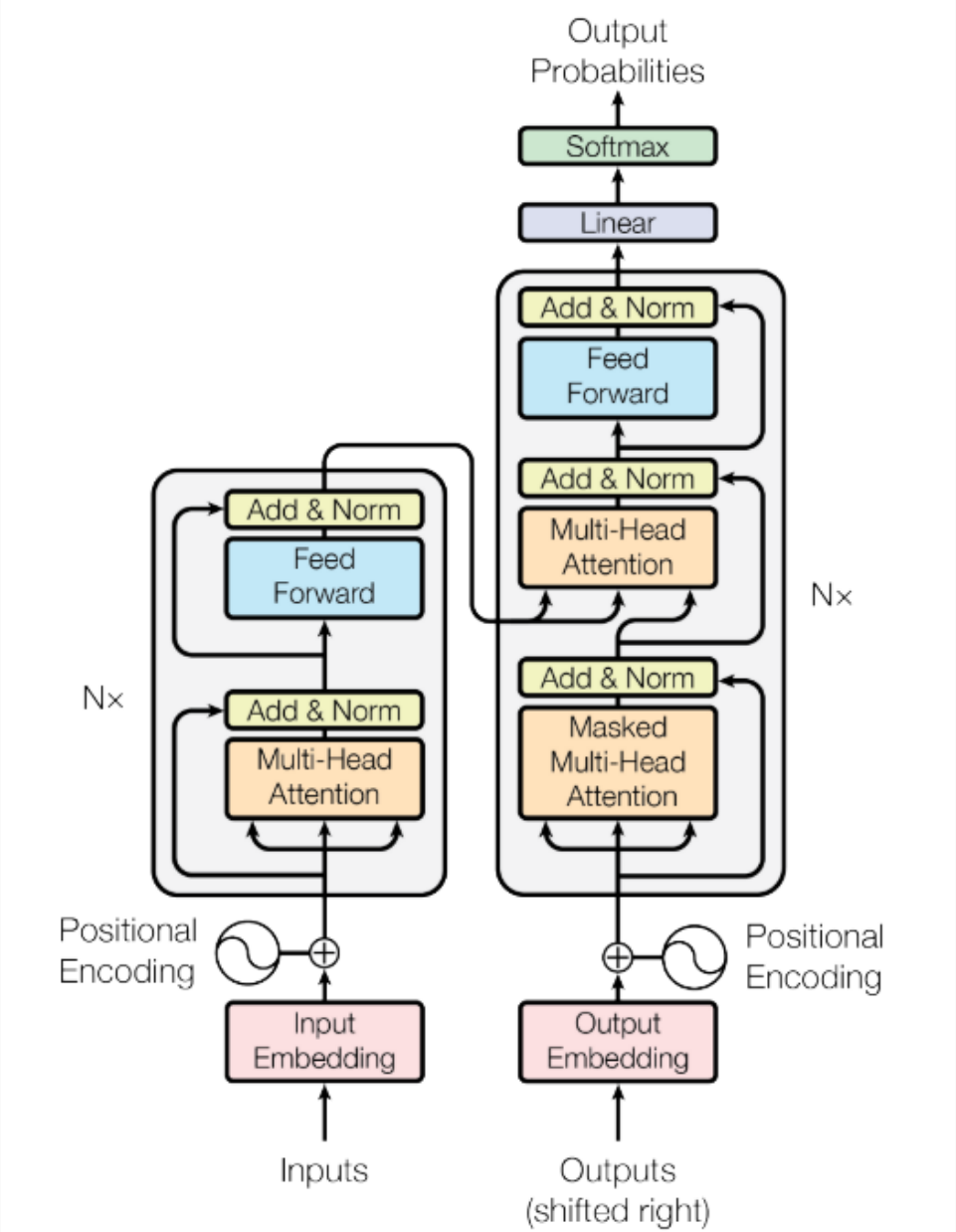
Multi-Head Attention

Scaled Dot product




Multi-Head Attention
Linear
Concat
Scaled Dot-Product Attention
Linear
Linear
Linear
V
K
Q

The Transformer - model architecture



Positional Encoding - Example

$$PE_{(pos,2_i)} = sin(pos/1000^{2i/d_{model}})$$

$$PE_{(pos,2_{i+1})} = cos(pos/1000^{2i/d_{model}})$$

Pending Discussion

- Pointer Generator Attention network
- Multi Head location based Attention
- Multi Head multi resolution location based Attention

Questions?

Multi Head Dot Product Attention

Multi Head location based Attention

Multi Head multi resolution location based Attention