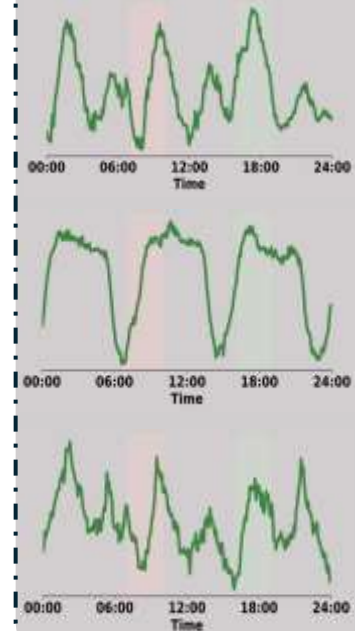
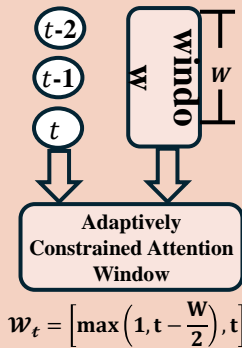


Ground Truth

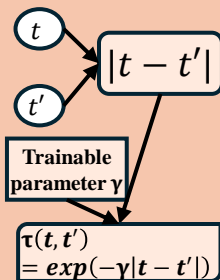
 $e_1$  $e_2$  $e_3$  $e_4$  $e_5$  $e_6$ 

## Efficient Transformer

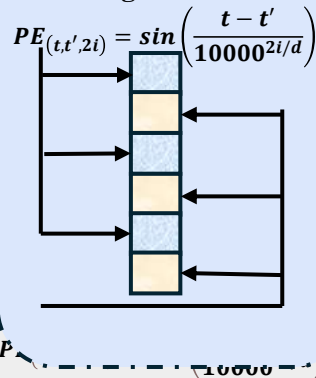
## Dynamic Windowed Masked Attention



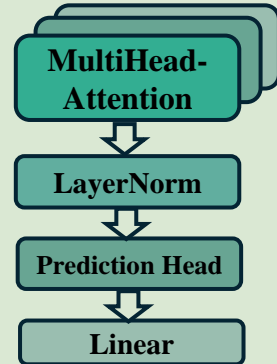
## Learnable Exponential Decay



## Precomputed Position Encodings



## Layer Normalization



Prediction

