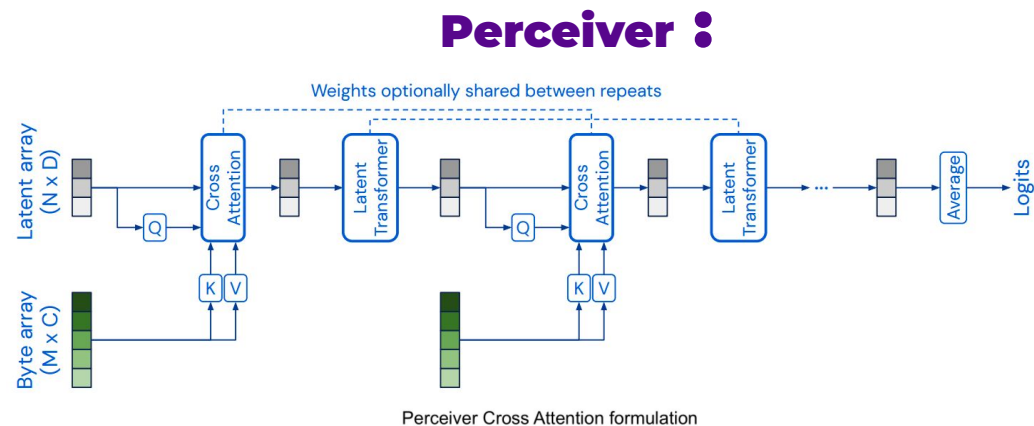


# Perceiver Is All You Need?

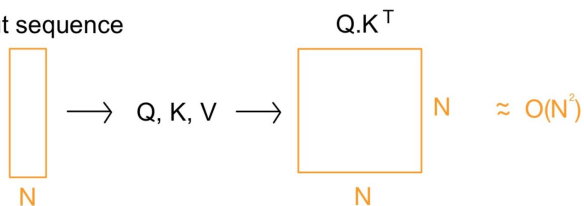
- Atharv Bhat, Saumyaa Shah, Sourabh K. Bhattacharjee, Yash Thesisia (Group 25)

- Attention is Expensive !
- Small Input Sequences
- Typically 512 / 1024

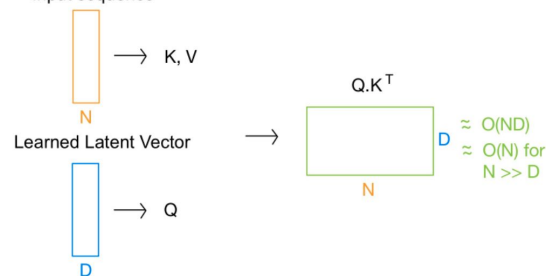


Full Attention formulation

Input sequence



Input sequence



# Long Range Arena

- Benchmark to Test Efficient transformers !

## BYTE-LEVEL TEXT CLASSIFICATION

- IMDB sentiment classification with Byte Character Encoded Inputs. Token Size : 4K

## LONG LISTOPS

- Listops (Nangia & Bowman, 2018)  
Token Size : 2K

INPUT: [MAX 4 3 [MIN 2 3 ] 1 0 [MEDIAN 1 5 8 9, 2]]

OUTPUT: 5

## BYTE-LEVEL DOCUMENT RETRIEVAL

- Given two input sequences, are they similar ? Token Size : 8K

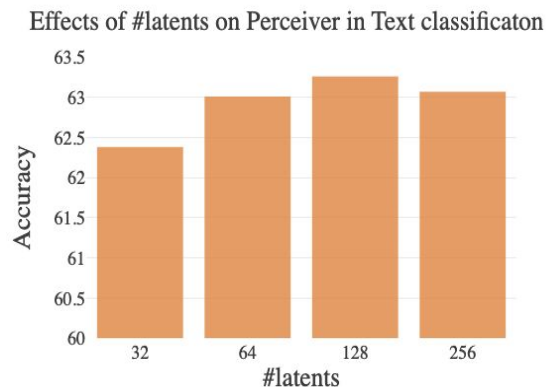
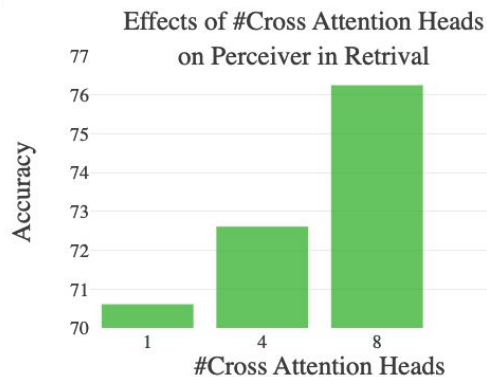
Model	ListOps	Text	Retrieval
Linear Attention	18.35	64.27	79.34
Linformer	37.25	55.91	79.37
Performer	18.35	64.10	78.62
<i>Perceiver</i>	37.15	63.76	76.25

Table 1: Comparison of baseline results of *Perceiver* to previous models

# Additional Results and Conclusion

Model	ListOps	Text
<i>Perceiver</i>	37.15	63.76
<i>Perceiver<sub>it</sub></i>	37.08	65.14

Table 2: Comparison of baseline results of *Perceiver* to *Perceiver* with iterative Attention.



- *Perceiver* model Performs well for Long Context Tasks
- We argue that, many efficient transformers Fail at long context tasks because of the inherent inductive biases that they introduce in their self attention reformulation
- Since the *Perceiver* model makes little to no assumptions about the input sequences, and it performs on PAR with SOTA efficient transformers, It can perform well on actual NLU tasks which involve working with long documents or text.