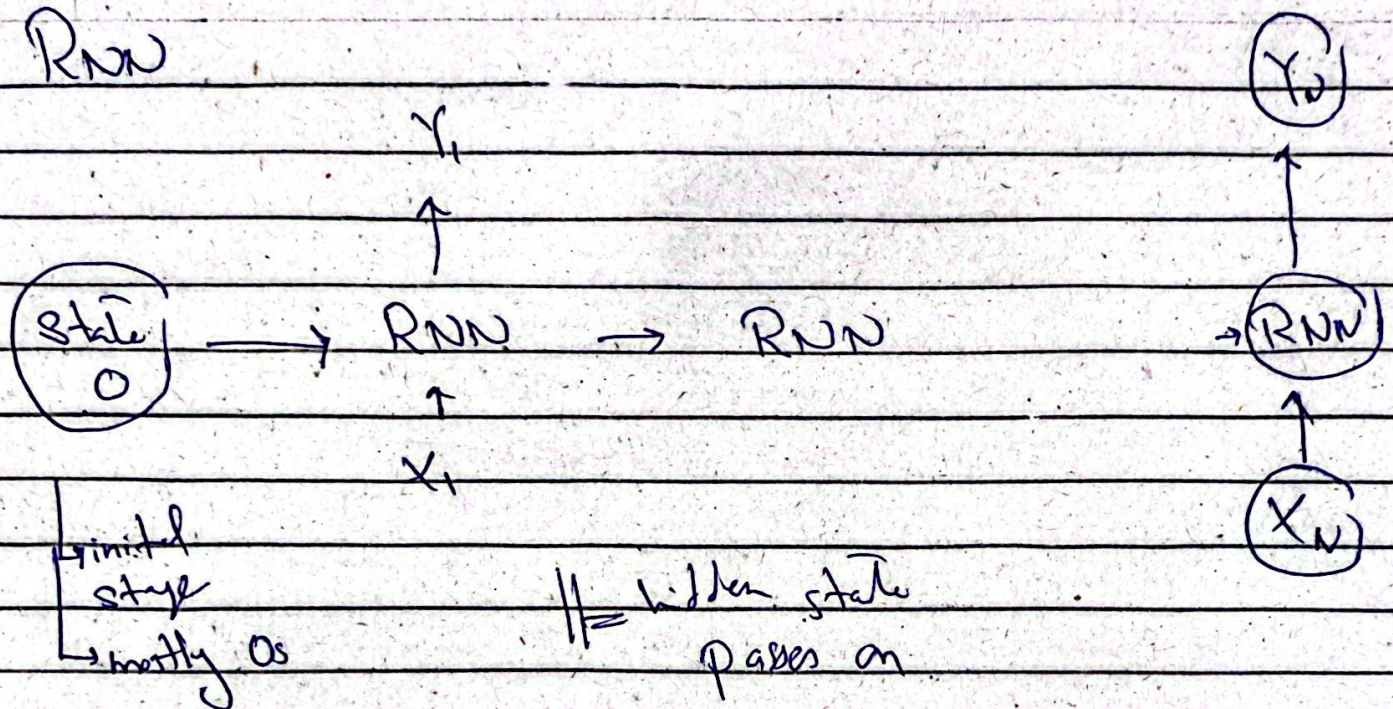


TRANSFORMERS:

RNN



RNN Limitation

- ↳ bad at remembering long sequences
- ↳ vulnerable to vanishing gradient

- ↳ They learn the sequence with time steps

In order to get result for Time stamp, you'll have to wait for till network finish for $t-1$, that's why its computationally inefficient.
So no parallelization