Ans to Q3:

RNN has no cell state. It has feedback loops in recurrent layers but faces vanishing gradient problem due to the absence of cell state. When dealing with a time series, it tends to forget old information. When there is a distant relationship of unknown length, the absence of memory generates wrong predictions.

On the other hand, Peephole - LSTM allows peeping into memory. It has forget gate controller, input gate controller, and output gate controller. It increases memory's glance. The forget gate controller controls which information is to be stored or forgotten. Moreover, peephole connections allow the gates to utilize the previous internal state as well as the previous hidden state. This is an improvement over the memoryless RNN.

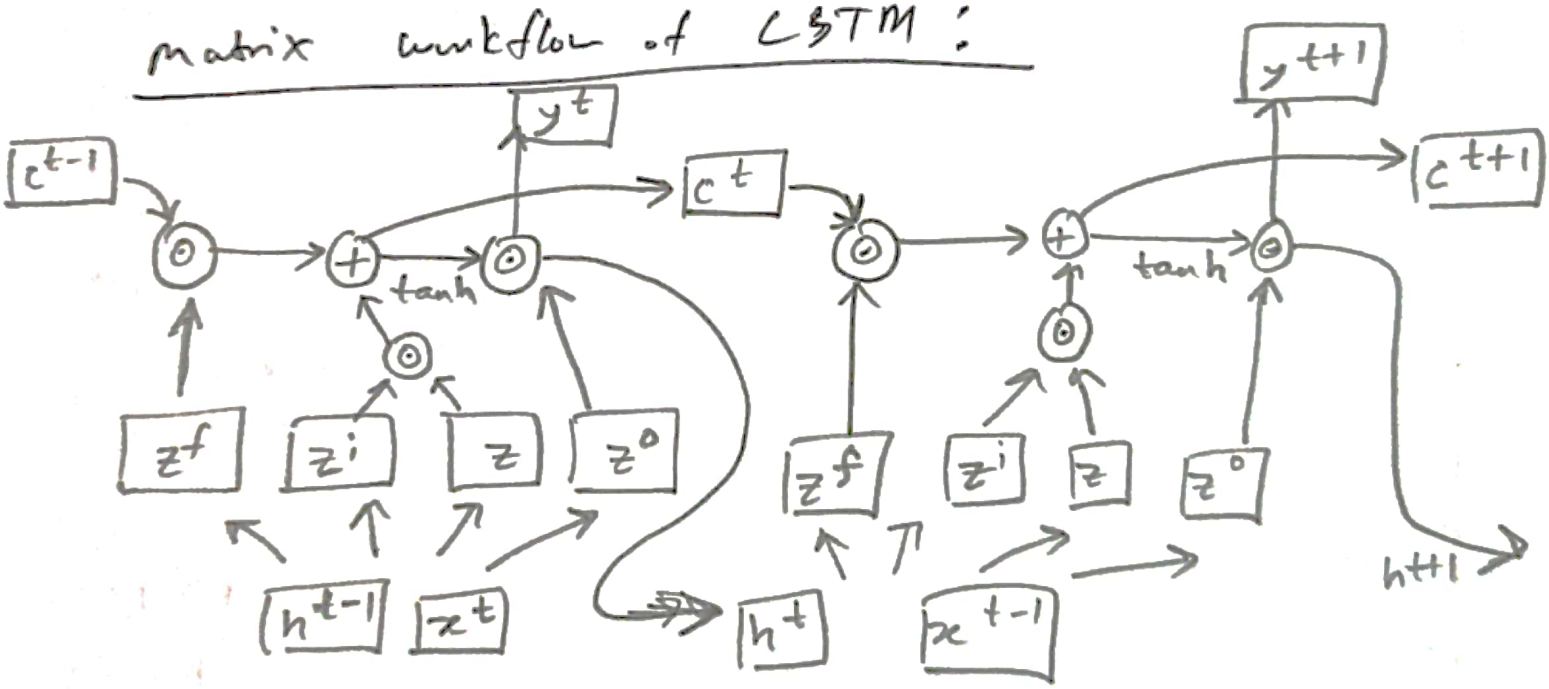
GRU is the newer generation of RNN and it is better than Peephole-LSTM. It does not read memory units so it is faster to train compared to Peephole-LSTM, which is very resource intensive. Moreover GRU uses 2 gates where the update gate keeps information of how much past information needs to be maintained. The update gate is a combination of forget gate and input gate of LSTM. The second gate is called reset gate. This tracks how much past information is to be neglected. GRU does not have cell state, like RNN.

The difference between LSTM forget gate and GRU reset gate is given below:

i) LSTM forget gate decides which information needs to be remembered or forgotten. On the other hand, GRU reset gate decides how much of past information is needed to be ignored.

ii) LSTM forget gate takes information from current input and hidden state through sigmoid function. The sigmoid function has 0, 1 as gating switch.

matrix workflow of LSTM:



The 4 'z' matrix computation should be done concurrently.