Comparison of RNN, GRU, LSTM and S4D on WMT2014 English-French Training Loss for s4d Training Loss for rnn Training Loss for gru  $10^{0}$ Loss  $10^{-1}$  $10^{-2}$ 10 20 30 50 60 80 40 70 **Epoch**