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} Training Loss for Istm Loss 10^{-1} 10^{-2} 20 10 30 40 50 60 70 80 **Epoch**