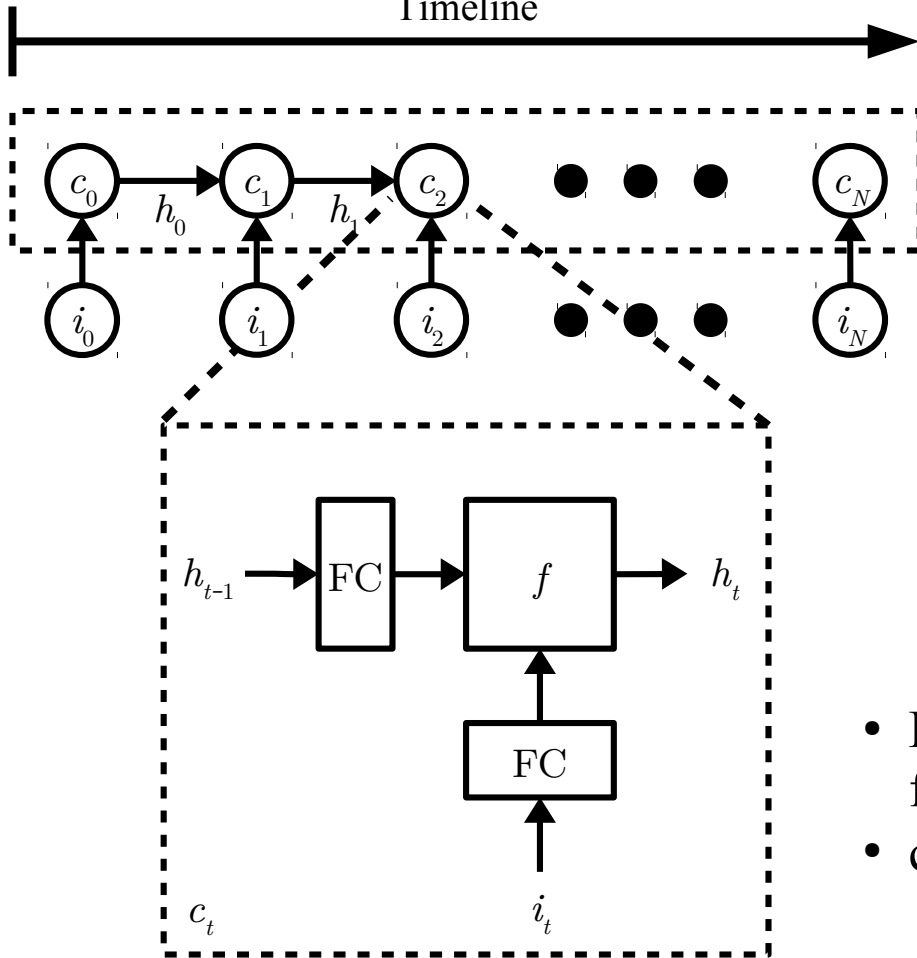


EcoRNN: Fused LSTM RNN Implementation with Data Layout Optimization

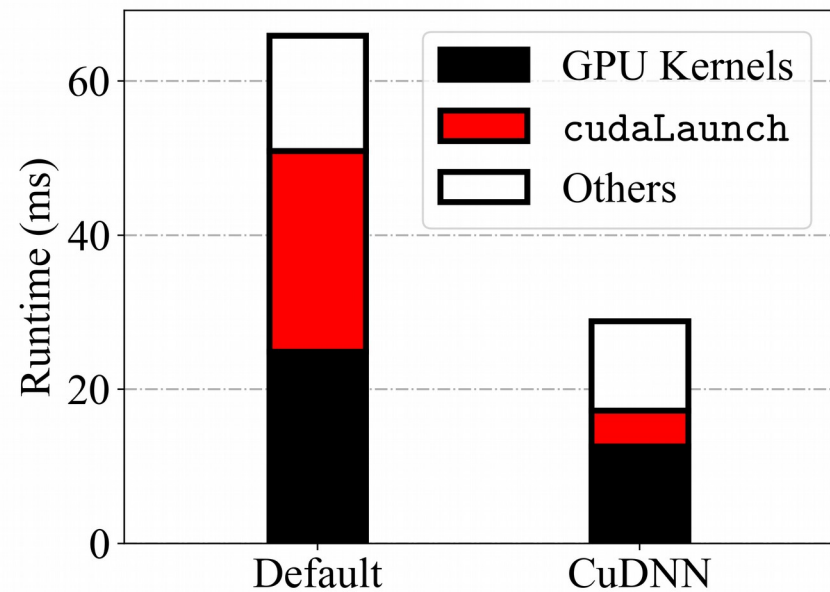


LSTM RNN and Cell

Timeline

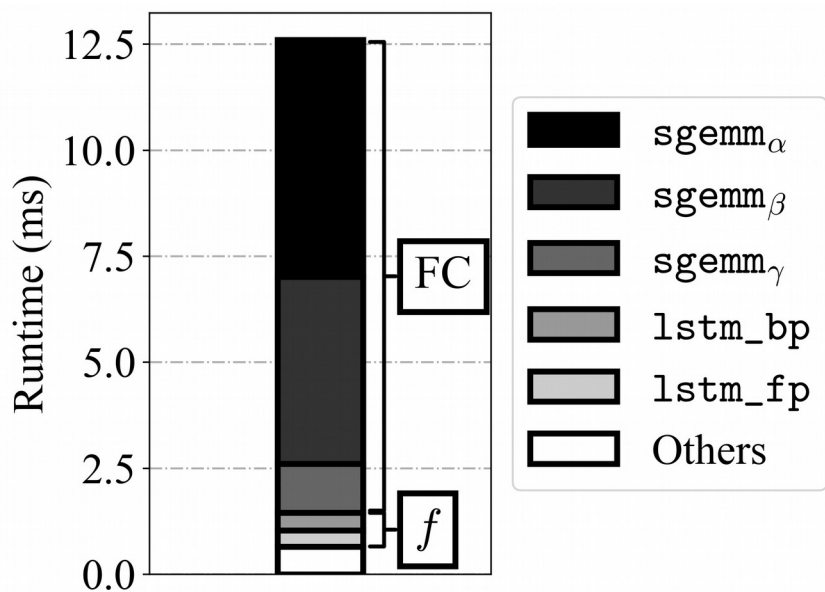


Runtime Breakdown of MXNet Default and cuDNN

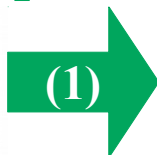


- Default implementations in machine learning frameworks have **cudaLaunch** overhead.
- cuDNN solves the issue, but is **closed-sourced**.

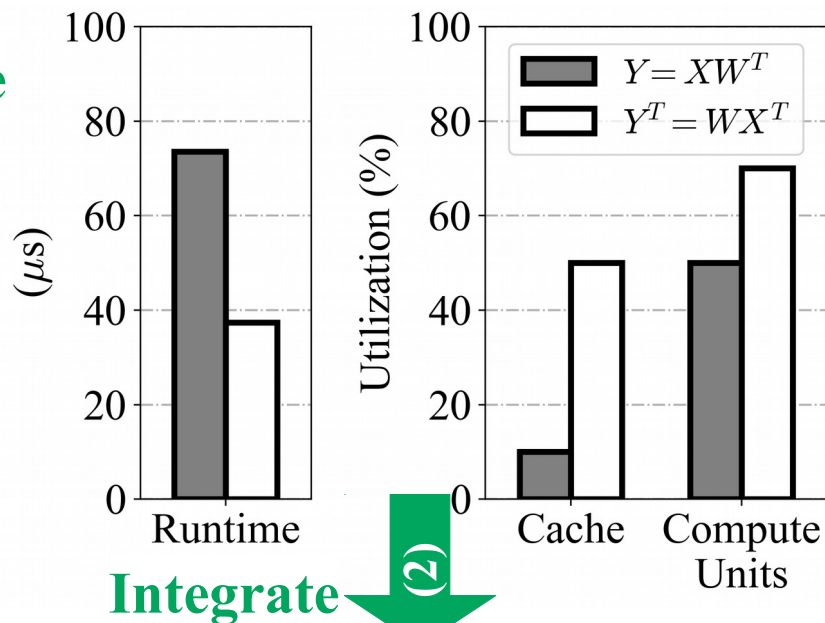
cuDNN Runtime Breakdown



Optimize



Data Layout Optimization (FC)

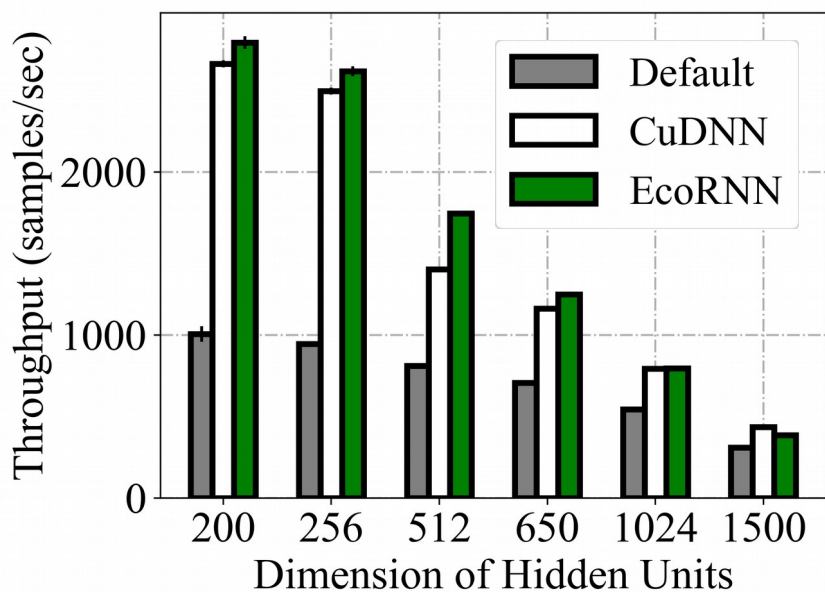


Integrate



Data Layout Optimization (RNN)

Language Modeling Benchmark



Employ

