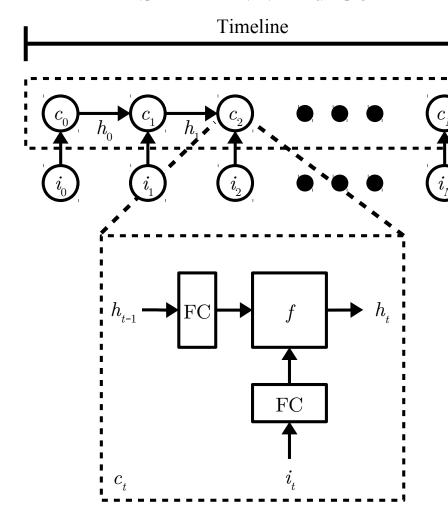
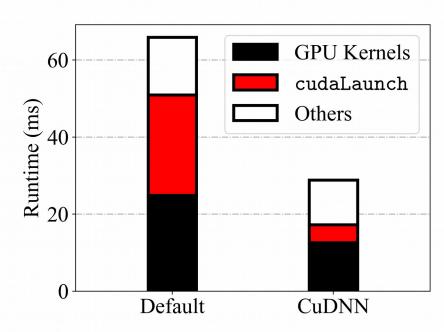
EcoRNN: Fused LSTM RNN Implementation with Data Layout Optimization



LSTM RNN and Cell



Runtime Breakdown of MXNet Default and cuDNN

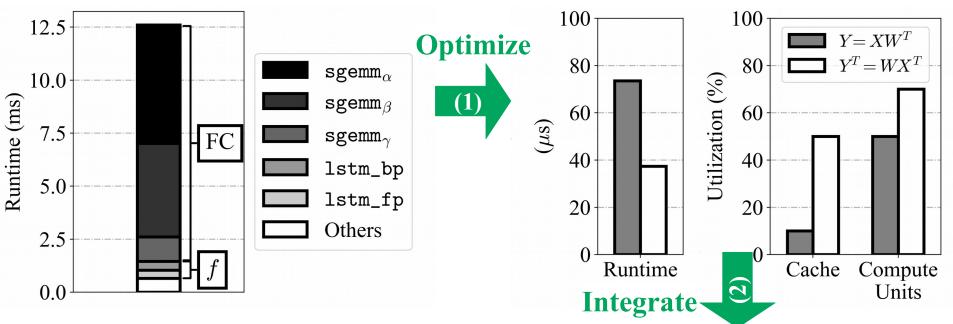


- Default implementations in machine learning frameworks have **cudaLaunch** overhead.
- cuDNN solves the issue, but is closed-sourced.

Bojian Zheng¹, Akshay Nair, Qiongsi Wu, Nandita Vijaykumar², Gennady Pekhimenko ¹University of Toronto, ²Carnegie Mellon University

cuDNN Runtime Breakdown

Data Layout Optimization (FC)



Language Modeling Benchmark

Data Layout Optimization (RNN)

