
ECCV'20
ONLINE
23-28 AUGUST 2020
16TH EUROPEAN CONFERENCE ON
COMPUTER VISION
WWW.ECCV2020.EU

ECCV'20 ONLINE
23-28 AUGUST 2020

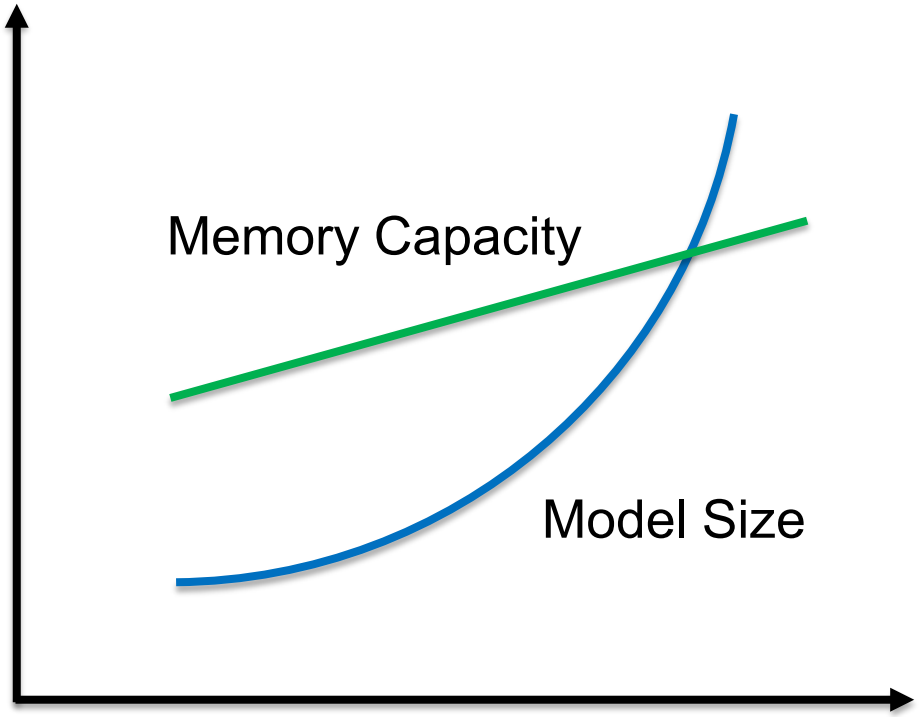
TEXAS
The University of Texas at Austin

An Efficient Training Framework for Reversible Neural Architectures

Zixuan Jiang, Keren Zhu, Mingjie Liu, Jiaqi Gu, David Z. Pan

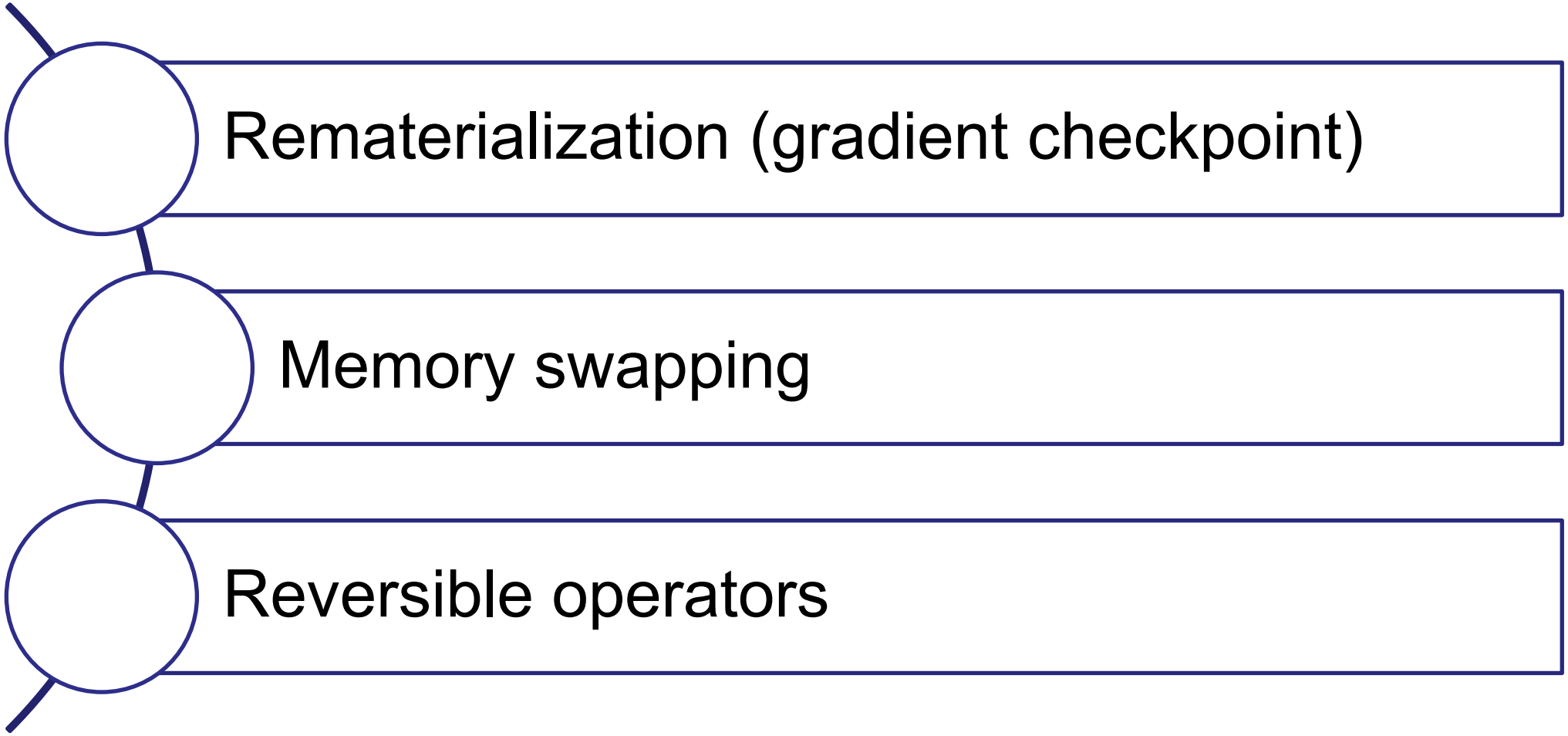
Electrical and Computer Engineering
The University of Texas at Austin

Large model hits memory capacity

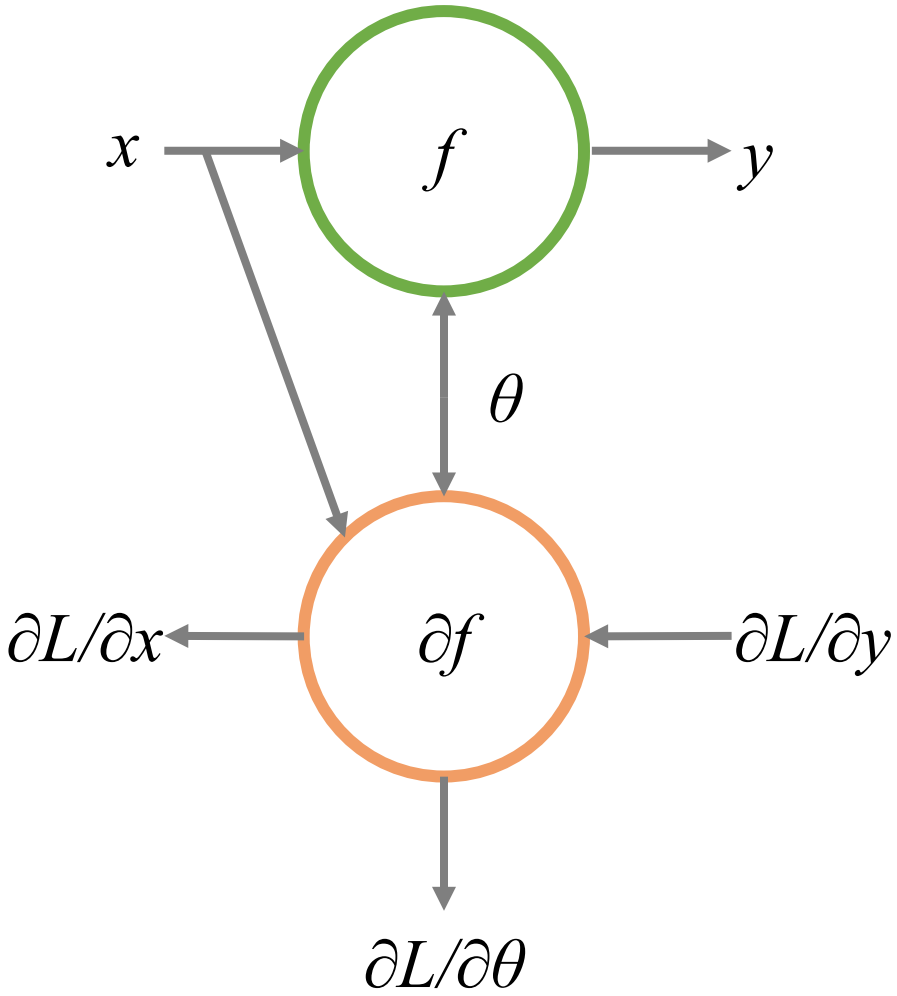


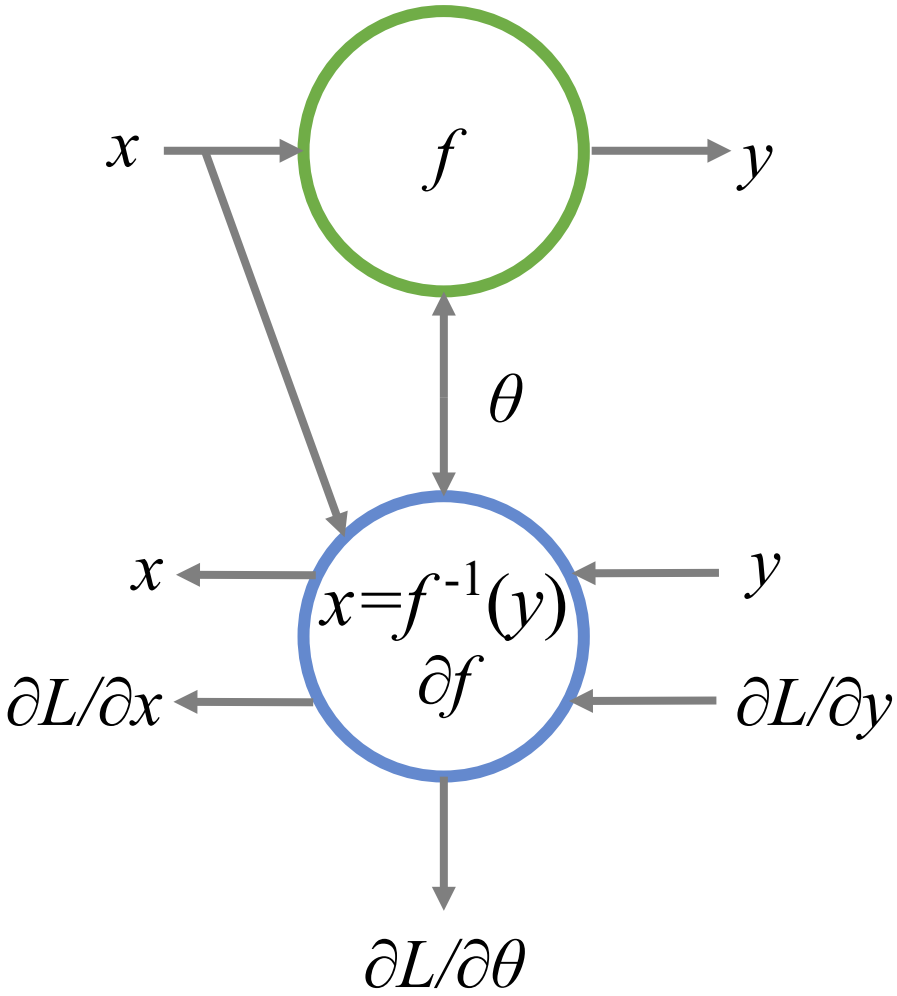
Mini-batch size of **1** in extreme cases.

Memory saving



Reversible operators





Related work

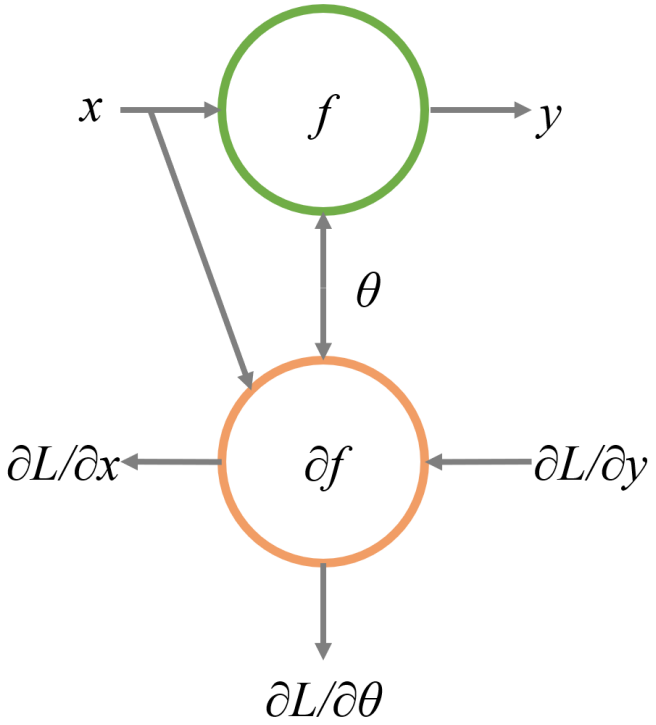
- Scheduling and graph optimization
  - Gradient checkpoint: [Chen et al. 2016], [Jain et al. 2020]
  - Memory swapping: [Zhang et al. 2019]

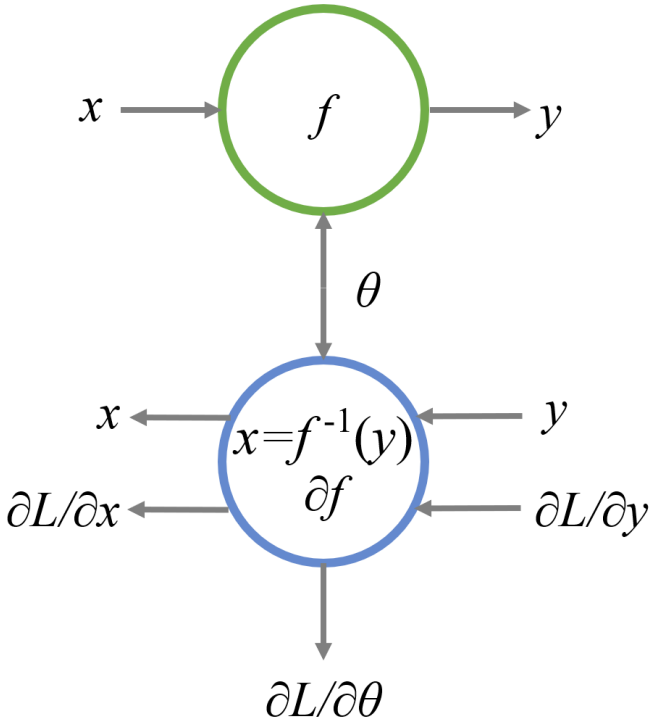
- Reversible neural networks
  - Implicitly reversible operator, e.g., convolution layers with a stride of 1
  - Inplace ABN [Bulò et al. 2018]
  - Neural ordinary differential equations [Chen et al. 2018]
  - Reversible residual architecture [Gomez et al. 2017]
  - Reformer [Kitaev et al. 2020]

- No work on scheduling for reversible neural networks.

What is the optimal scheduling for reversible neural architectures?

Two modes of a reversible operator

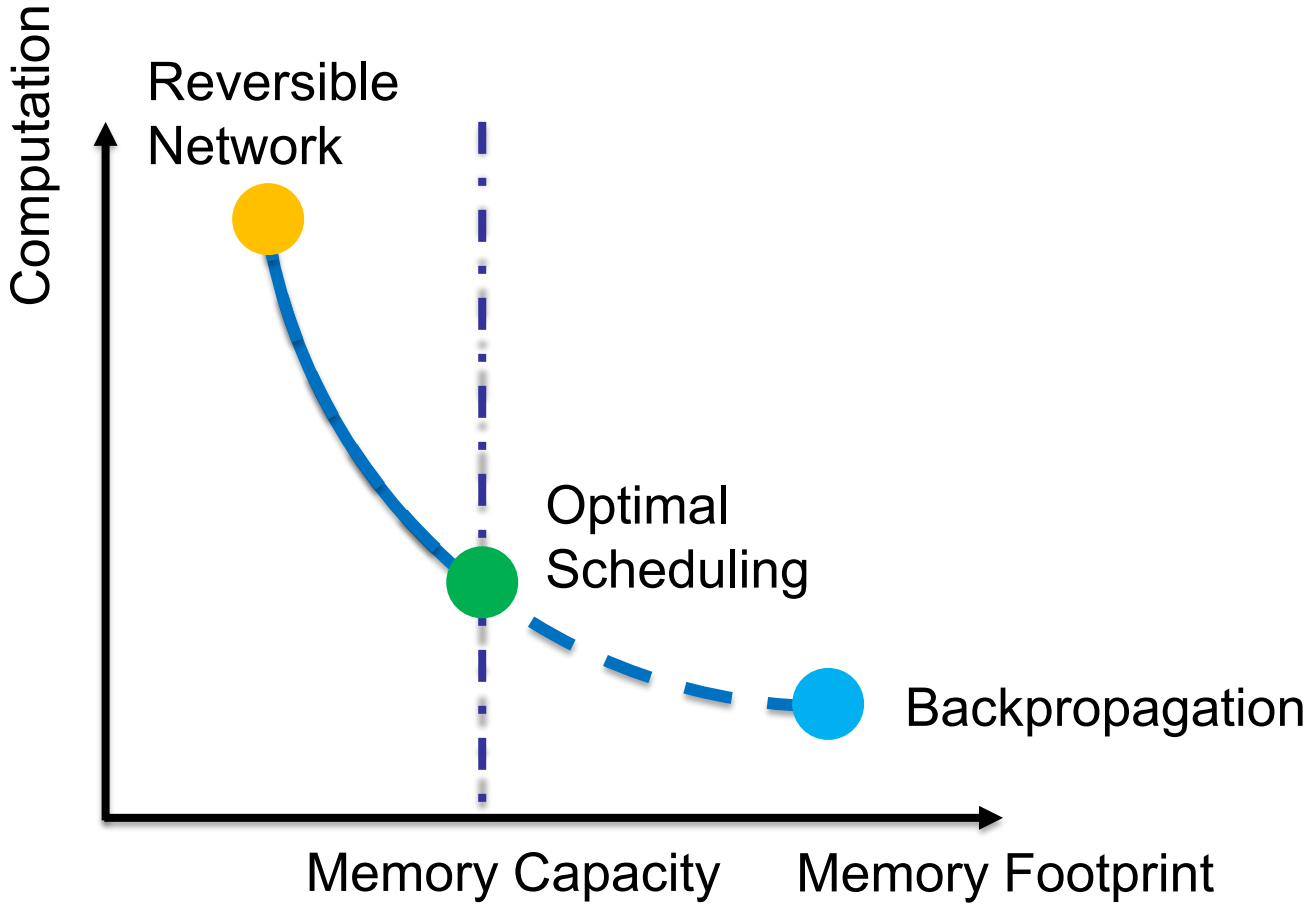




Mode	Forward	Backward	Computation cost	Memory cost
M-Mode	Discard x	Recover x from y	$x = f^{-1}(y)$	0
C-Mode	Save x	Use x directly	0	Size of x

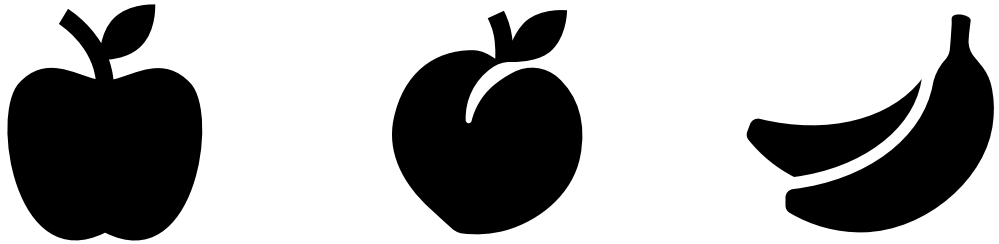
Scheduling problem

- For a neural network with n reversible operators, there are 2^n possible solutions.



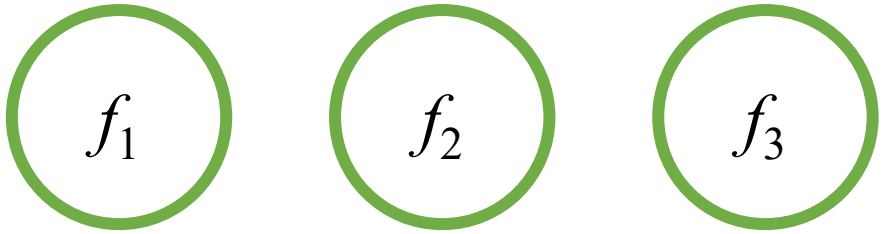
0/1 Knapsack problem

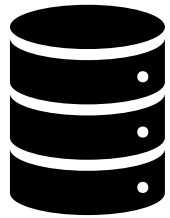
- n items
- Value, weight
- Bag capacity
- Fill the bag with the maximum value.



Scheduling problem

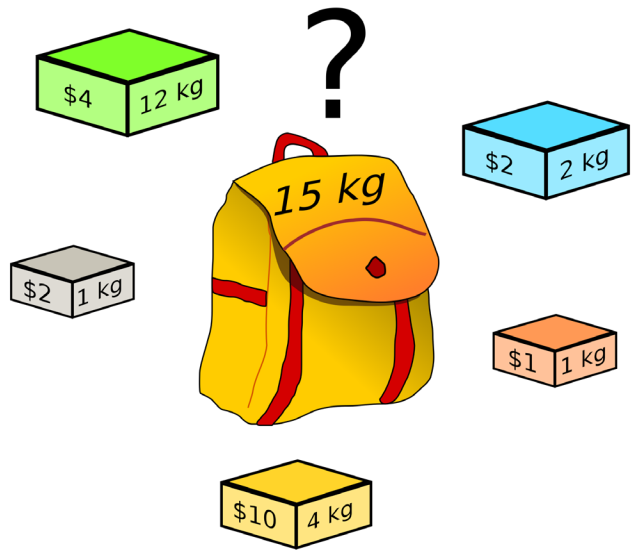
- n reversible layers
- Extra execution time, memory footprint
- Memory capacity
- Fill the memory with the maximum saved time.





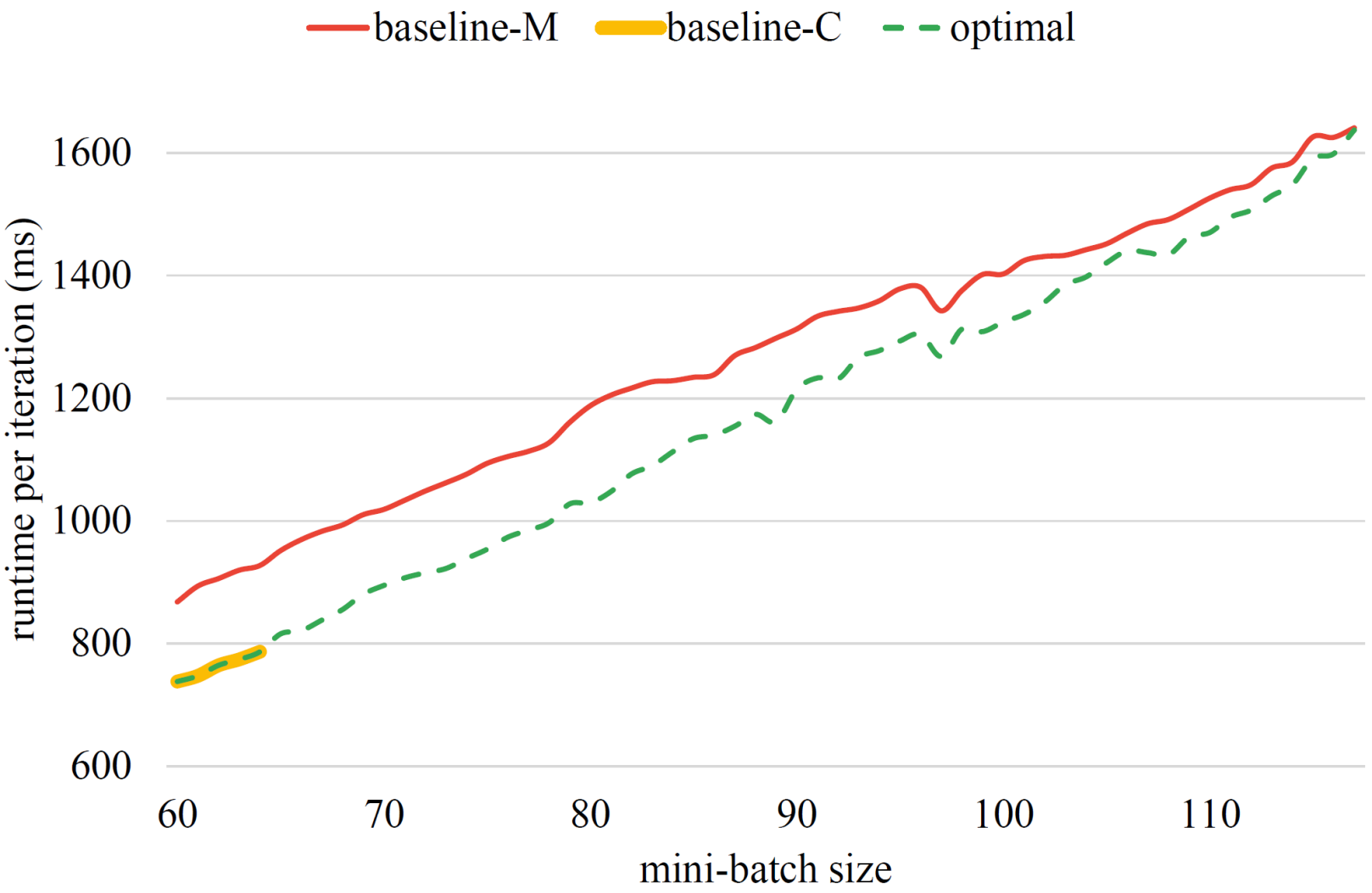
Algorithm and framework

- Use dynamic programming algorithm to solve the scheduling problem (knapsack problem).

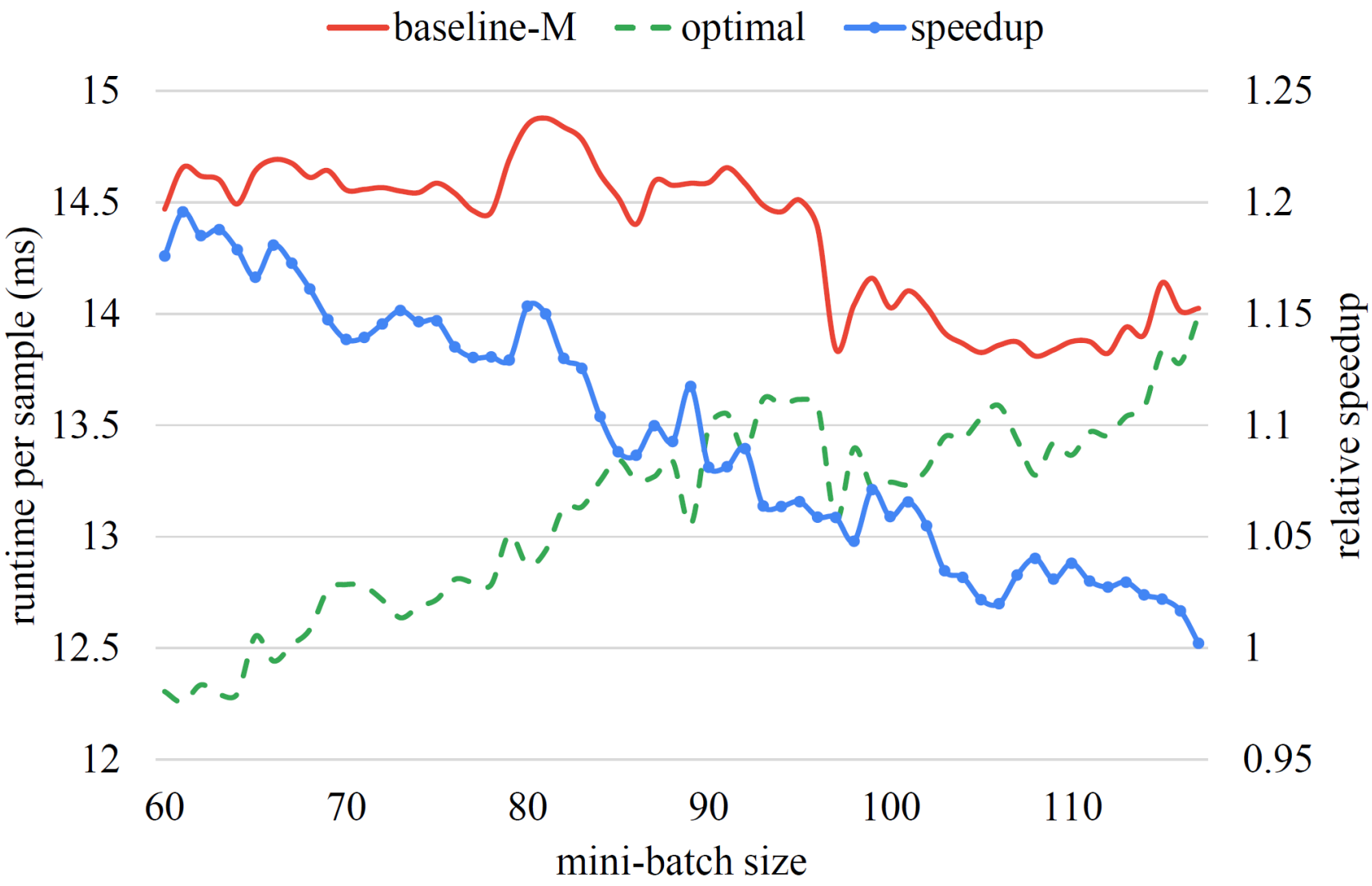


Reversibility analysis → Profiling → Decision via DP → Efficient training

Results on RevNet-104



Results on RevNet-104



Conclusions

- **New Perspective**. Scheduling for reversible architectures.
- **Optimality**. The problem can be solved use DP.
- **Automation**. Our framework provides a fully automated solution.

More in our paper

- Details regarding problem formulation
- What is the optimal batch size?
- More experiments on various reversible neural networks

Thank you!