L2ight: Enabling On-Chip Learning for Optical Neural Networks via Efficient in-situ Subspace Optimization

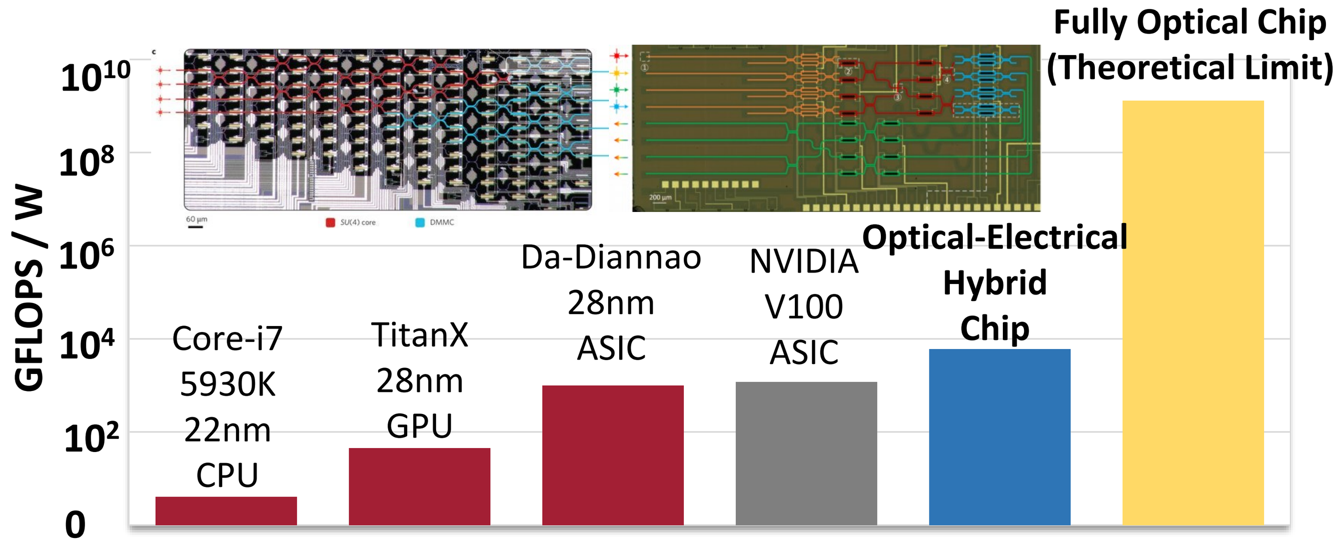
Jiaqi Gu, Hanqing Zhu, Chenghao Feng, Zixuan Jiang, Ray T. Chen, David Z. Pan

TEXAS — The University of Texas at Austin

NEURAL INFORMATION PROCESSING SYSTEMS

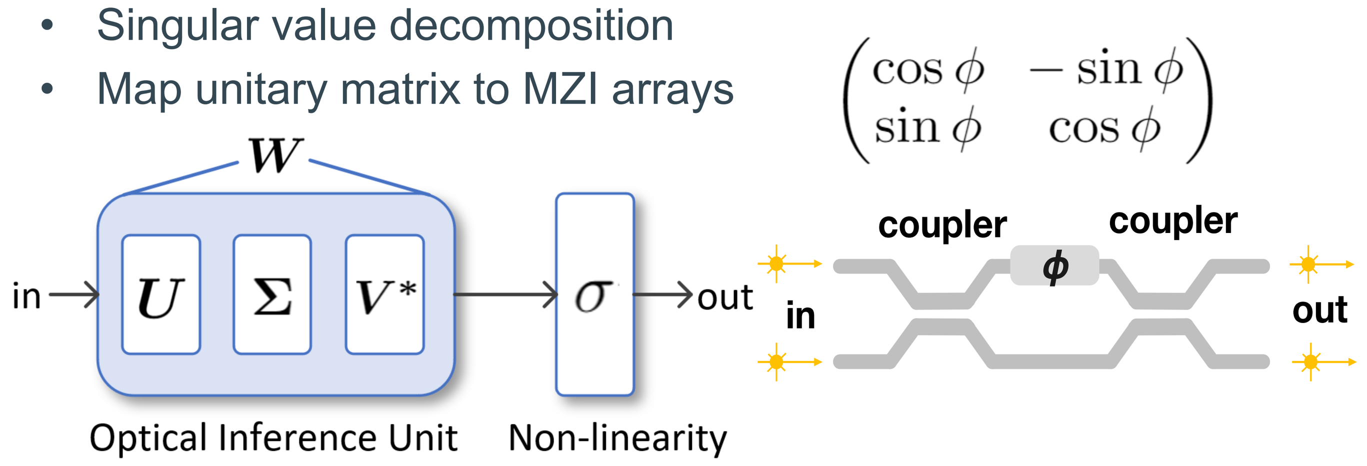
Introduction: Optical Neurocomputing

Photonics for AI Acceleration: **Ultra-Fast** & **Ultra-Efficient**



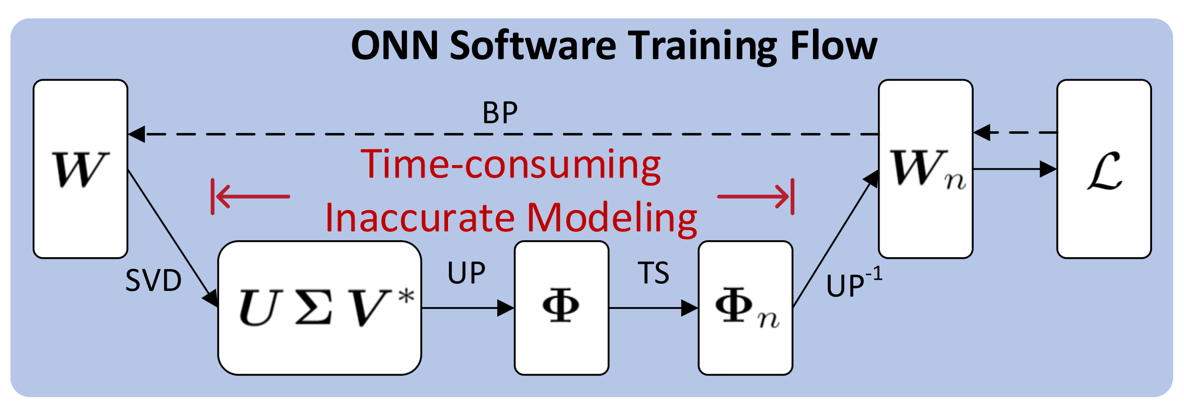
Optical Neural Network Basics

- Singular value decomposition
- Map unitary matrix to MZI arrays



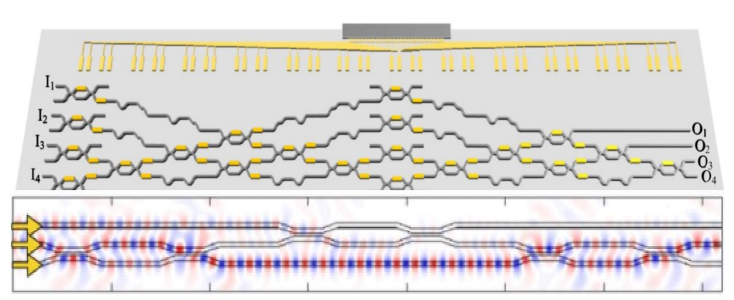
ONN On-Chip Training

- ***Robust*** Deployment & On-Chip ***Learnability***
- Ultra-fast & effective in-situ noise handling



Challenges and Prior Work

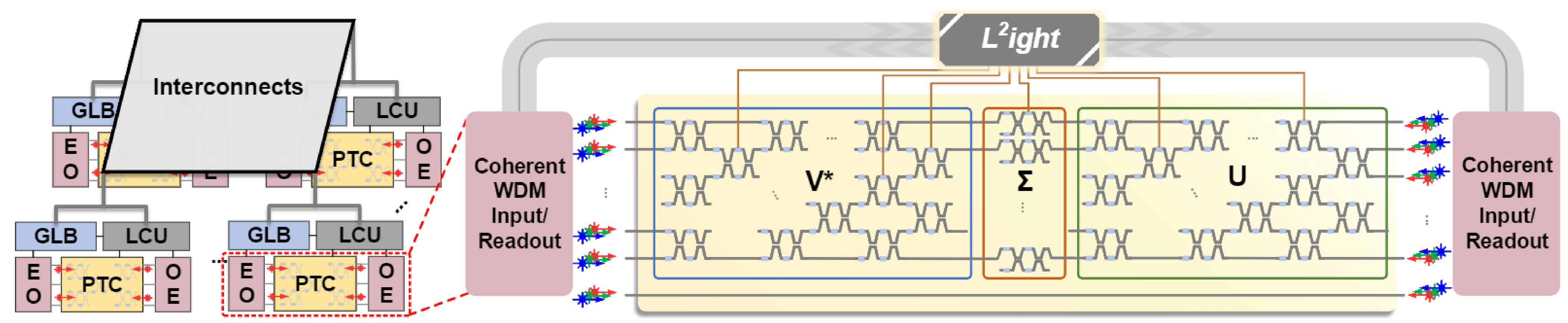
- Unscalable: 100 ~ 1000 MZIs
- Inefficient & unstable



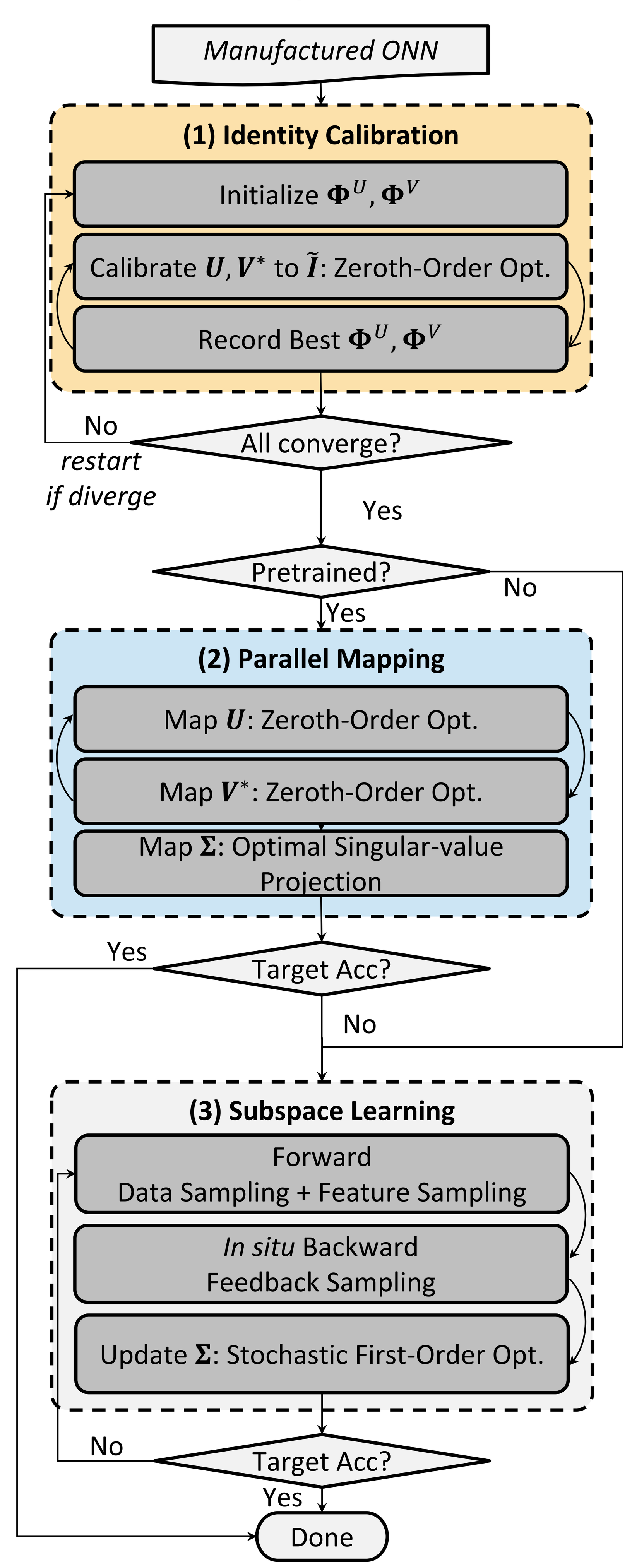
	PSO [OE'19]	AVM [Optica'18]	FLOPS [DAC'20]	MixedTrain [AAAI'21]	*Our L²ight*
#Params	~100	~100	~1,000	~2,500	**~10 M**
Algorithm	Evolution (ZO)	Adjoint Method (FO)	ZO SGD	SZO-SCD	**ZO + FO**
Resolution Req.	High	Medium	High	Medium	**Medium**
Observability Req.	Coh. I/O	Coh. I/O+ Per device monitor	Coh. I/O	Coh. I/O	**Coh. I/O**

L²ight: Our Synergistic On-Chip Learning Framework

- **Scalability**: First to handle ***million-parameter*** ONNs
- **Efficiency**: Multi-level sparsity for ***30×*** higher efficiency

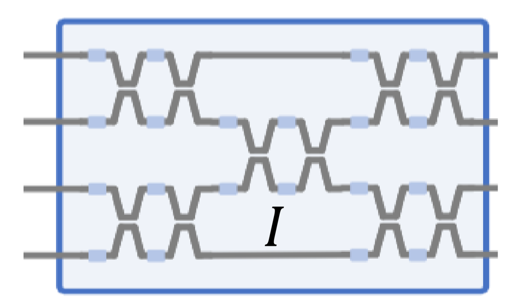


Three-stage Framework



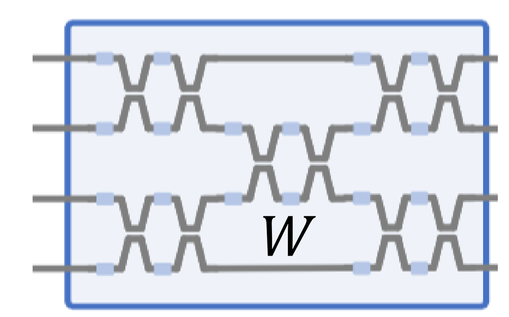
Identity Calibration (IC)

- Variation-Agnostic Circuit State Preparation
- Large-scale ZOO → Batched subtasks



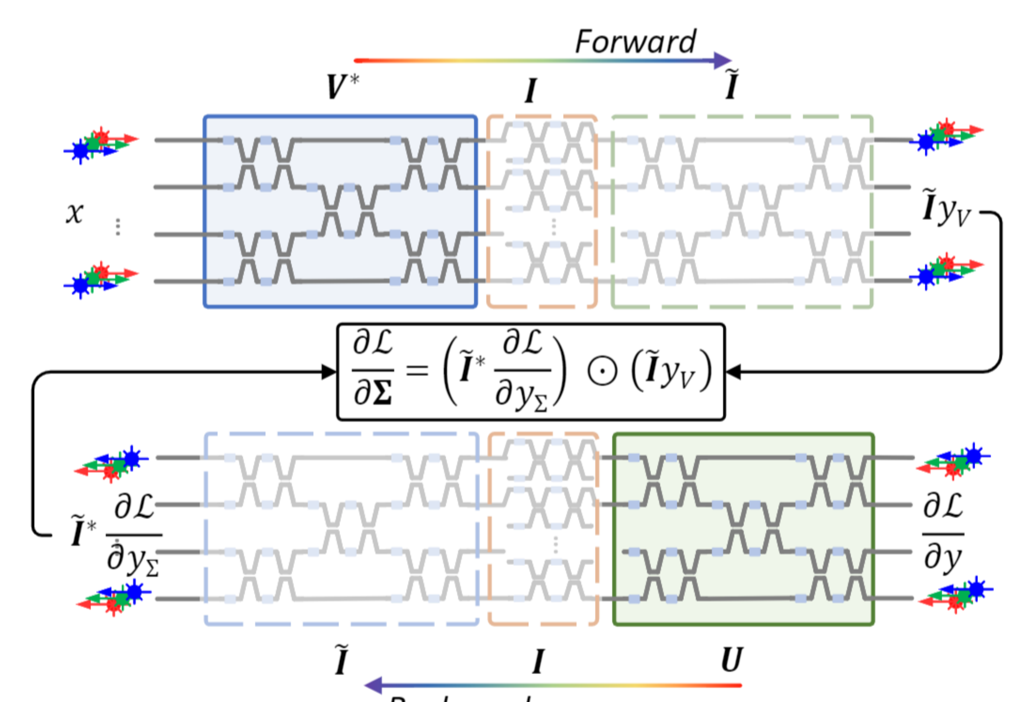
Parallel Mapping (PM)

- Decouple ZOO from stochasticity → Efficient mapping



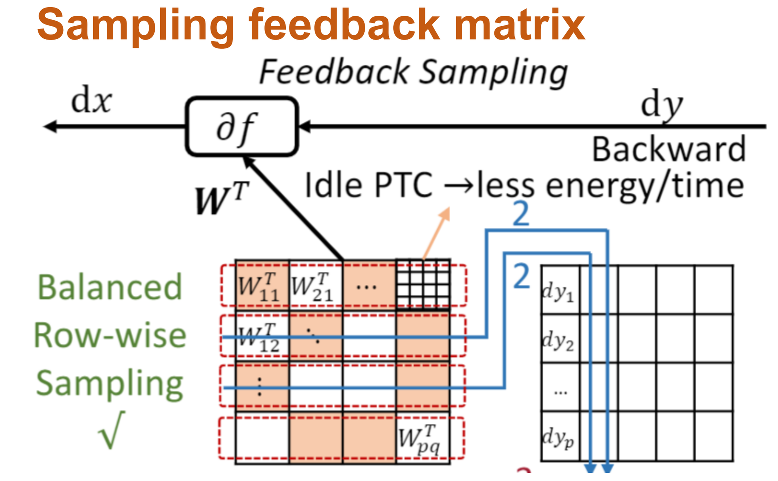
Subspace Learning (SL): Only optimize singular values

- ***First-order*** subspace gradient via ***Reciprocity***
- Trade full-space trainability → Efficient first-order gradients

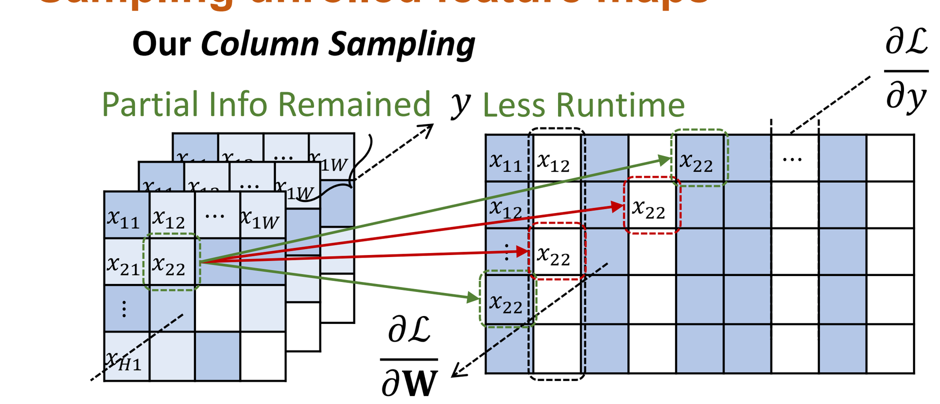


Efficient Multi-level Sparse Subspace Learning

Sampling feedback matrix

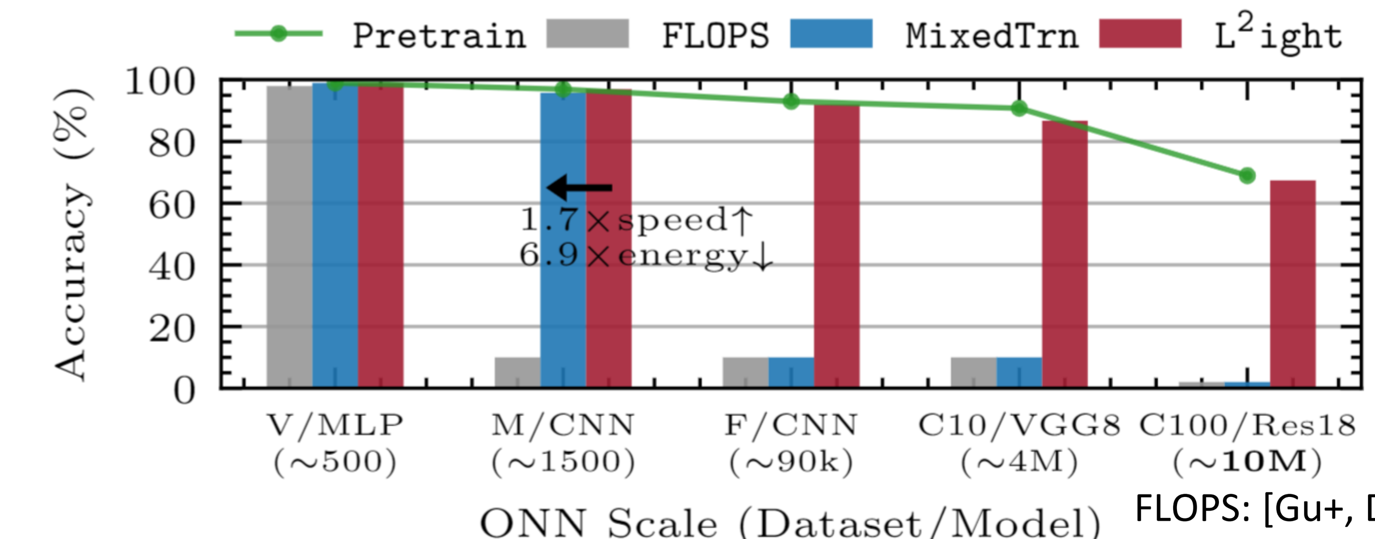


Sampling unrolled feature maps



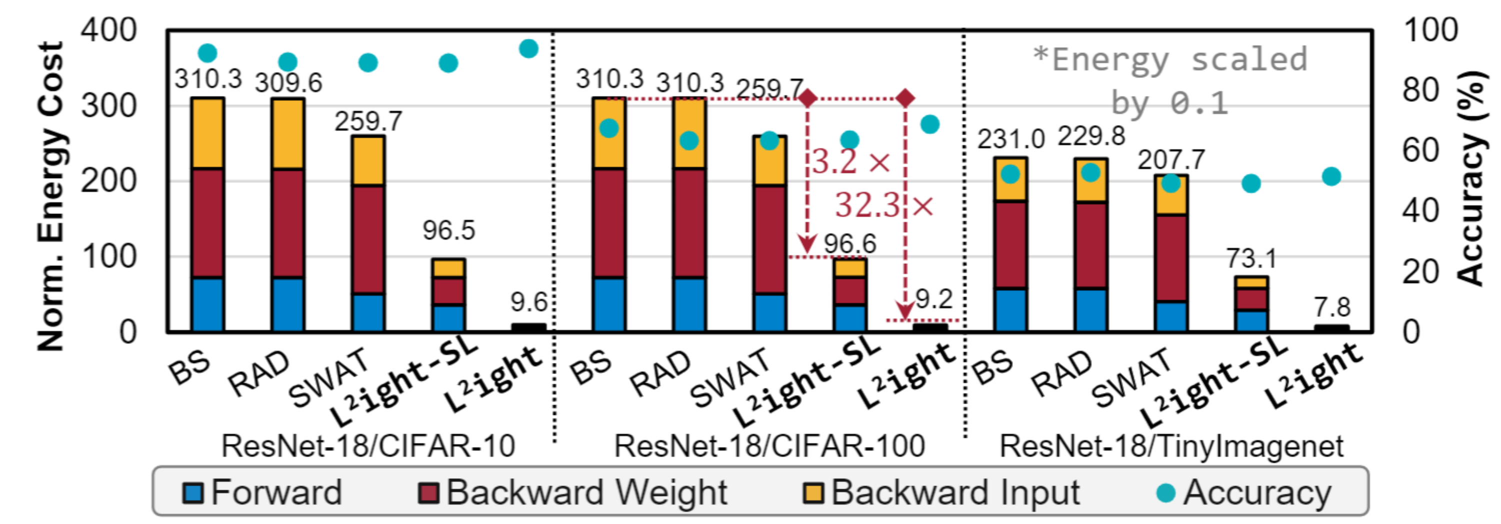
Experimental Results: Scalability & Efficiency

1000× more scalable: million-parameter ONNs



FLOPS: [Gu+, DAC'20]
MixedTrn: [Gu+, AAAI'21]

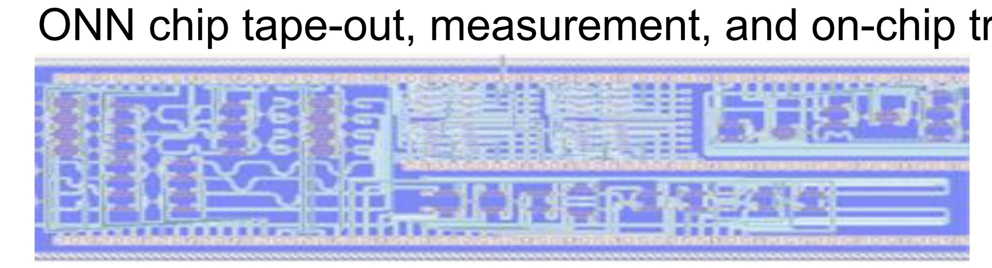
30× more energy-efficient + self-learnability and transferability



RAD: [Oktay+, ICLR'21]
SWAT: [Raihan+, NeurIPS'20]

Future Work

ONN chip tape-out, measurement, and on-chip training deployment



PyTorch, lumerical, SYNOPSYS, amf

References

Jiaqi Gu, Zheng Zhao, *et al.* (2020). "FLOPS: Efficient On-Chip Learning for Optical Neural Networks Through Stochastic Zeroth-Order Optimization." In: DAC.

Jiaqi Gu, Chenghao Feng, *et al.* (2021). "Efficient on-chip learning for optical neural networks through power-aware sparse zeroth-order optimization.". In: AAAI.

Yichen Shen, Nicholas C. Harris, *et al.* (2017). "Deep learning with coherent nanophotonic circuits.". In: Nature Photonics.

Website: https://jeremiemelo.github.io
Open-source: https://github.com/JeremieMelo/L2ight
PyTorch-ONN Library: https://github.com/JeremieMelo/pytorch-onn