DAC YF

FLOPS: Efficient On-Chip Learning for Optical Neural Networks

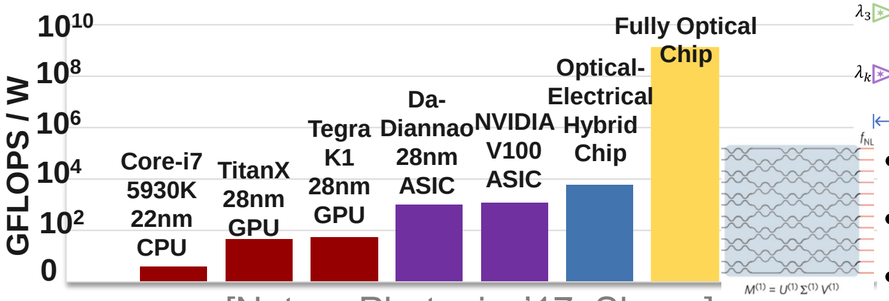
Jiaqi Gu, Ray T. Chen and David Z. Pan, The University of Texas at Austin

DESIGN AUTOMATION CONFERENCE 57
FROM CHIPS TO SYSTEMS – LEARN TODAY, CREATE TOMORROW

1.Neural Networks and AI Acceleration

ML Applications and Photonic Acceleration





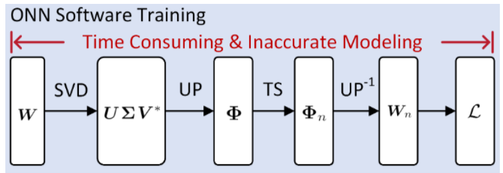
[Nature Photonics'17, Shen+]

2.Previous ONN Training Protocols

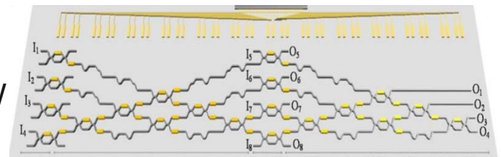
- **Software training**
  - Limited speed (>1 s)
  - Hardware-unaware

$$W = U\Sigma V^*$$

$$U(n) = D\prod_{i=n}^{2}\prod_{j=1}^{i-1} R_{ij}$$

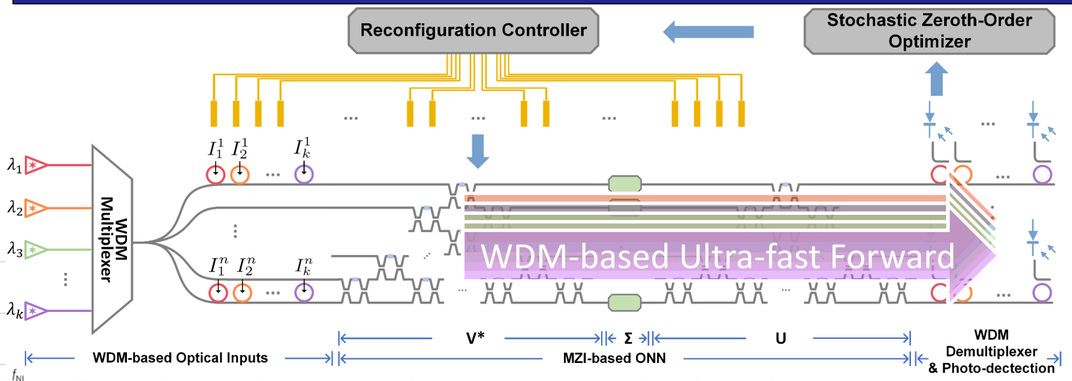


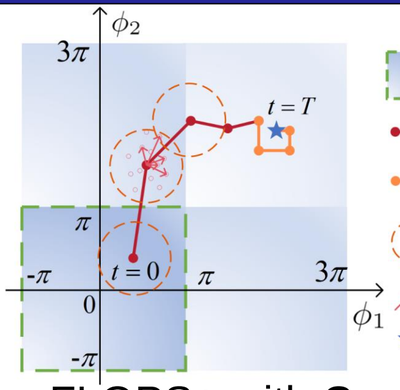
- **On-chip training**
  - Ultrafast (~1 ms)
  - 1000x faster than SW
  - Unscalable
  - Limited efficiency



[arXiv'19, Zhou+]

3.Proposed Method: FLOPS & FLOPS+



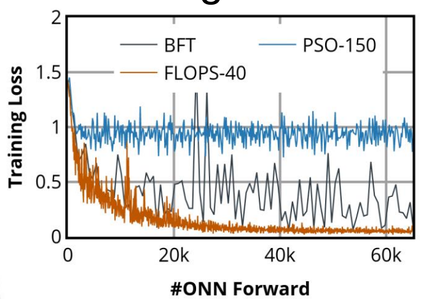


- **Stochastic zeroth-order optimization**
- **Efficiency:** WDM-based parallel gradient estimation
- **Accuracy:** Two-stage learning protocol with high accuracy
- **Robustness:** Robust learning under *in situ* device variations

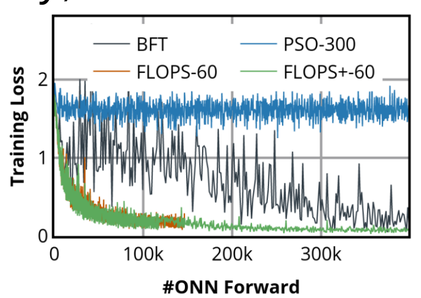
- FLOPS+ with ***SparseTune***
  - Sparse coordinate-wise fine-tuning
  - Improve Accuracy via searching
  - Sparsity guarantees efficiency

4.Experimental Results

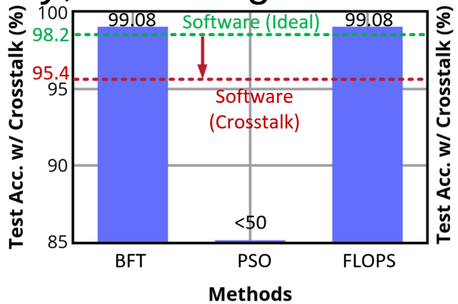
- **2-4x** higher efficiency; **10x** better scalability; **3-5%** higher robustness and accuracy



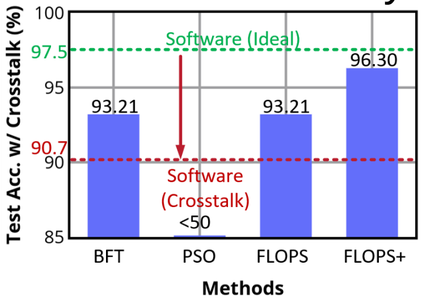
ONN config: 8-16-16-4 (448 MZIs)



ONN config: 10-24-24-6 (960 MZIs)



BFT: [Zhou+, 2019]



PSO: [Zhang+, 2019]