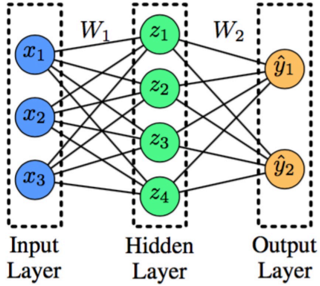
Countering Variations and Thermal Effects for Accurate Optical Neural Networks

Grace Li Zhang, Bing Li, Ulf Schlichtmann

Chair of Electronic Design Automation
Technical University of Munich

Neuromorphic Circuits

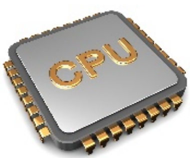
$Z_1 = A(w_{11}*x_1 + w_{21}*x_2 + w_{31}*x_3)$

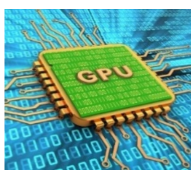


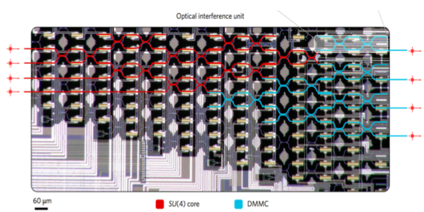
Massive multiplications and addition operations

Synapse Network

A Simple Neural Network





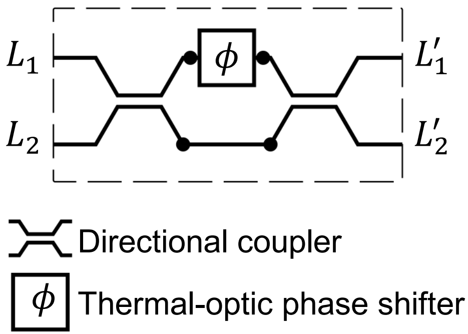


hardware accelerator for neural networks

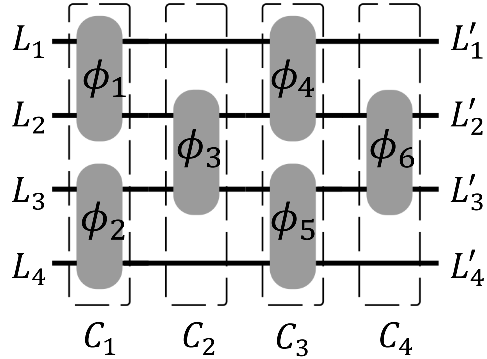
Optical computing chip [1]

[1] Y. Shen, N.C. Harris, S. Skirlo et al., “Deep learning with coherent nanophotonic circuits,” Nature Photonics, vol. 11, pp. 441–446, 2017.

MZI-based Optical Neural Networks



Mach-Zehnder Interferometer (MZI)



Grid-like architecture of MZI array

$$L_1 = A_1 e^{j\theta_1} \qquad L_2 = A_2 e^{j\theta_2}$$

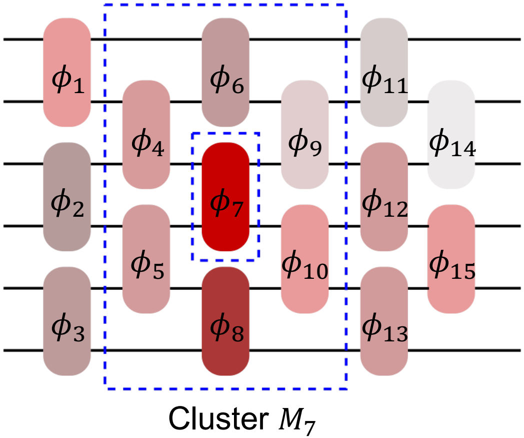
$$T = T_{C_4} T_{C_3} T_{C_2} T_{C_1}$$

$$\begin{pmatrix} L'_1 \\ L'_2 \end{pmatrix} = j e^{j\frac{\phi}{2}} \begin{pmatrix} \sin\frac{\phi}{2} & \cos\frac{\phi}{2} \\ \cos\frac{\phi}{2} & -\sin\frac{\phi}{2} \end{pmatrix} \begin{pmatrix} L_1 \\ L_2 \end{pmatrix}$$
$$= T \begin{pmatrix} L_1 \\ L_2 \end{pmatrix}$$

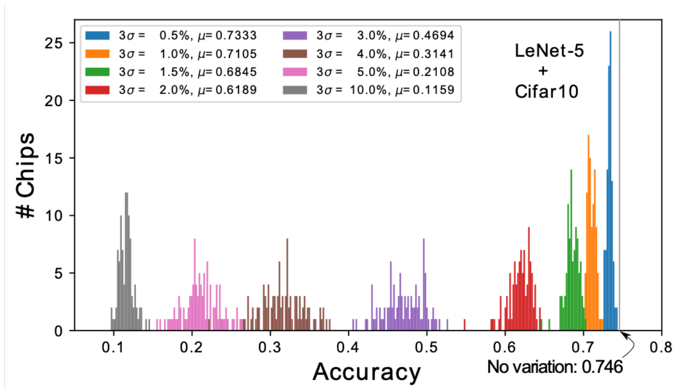
T is the transformation matrix

To apply ONN, phase shifters in MZIs need to be configured to expected values.

Challenges of MZI-based Optical Neural Networks



Thermal imbalance in an ONN.

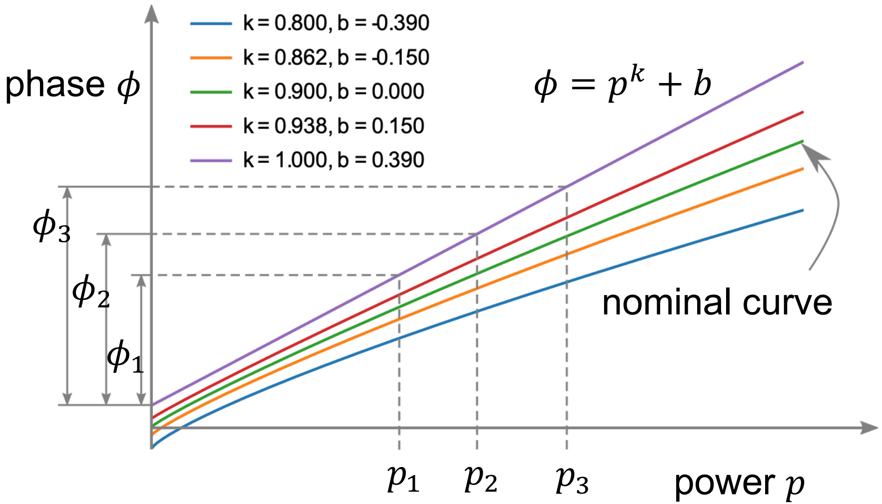


Accuracy degradation of LeNet-5 under process variations in MZI phases.

Thermal imbalance → Thermal crosstalk → $\phi \neq$ expected value

Large variations → Inference accuracy degradation → Unusable ONN

Process Variations of MZIs



Characteristic curves of five MZIs under process variations [1].

The smaller the phase of an MZI is tuned, the smaller the deviation of this phase from the expected value becomes.

Ma, J. Mower et al., “Efficient, compact and low loss thermo- optic phase shifter in silicon,” Opt. Express, vol. 22, no. 9, pp. 10 487–10 493, 2014.

Software Training for ONN to Reduce the Effects of Process Variations and Thermal Imbalance

Conventional cost function for ONNs :

$$\mathrm{C} = -\sum_{i=1}^{n} \hat{y}_i \log(y_i) = g(\boldsymbol{\phi})$$

- Reduce phase deviations and power consumption

$$\mathrm{C} = g(\boldsymbol{\phi}) + \alpha \cdot \sum_{i=1}^{N} \phi_i$$

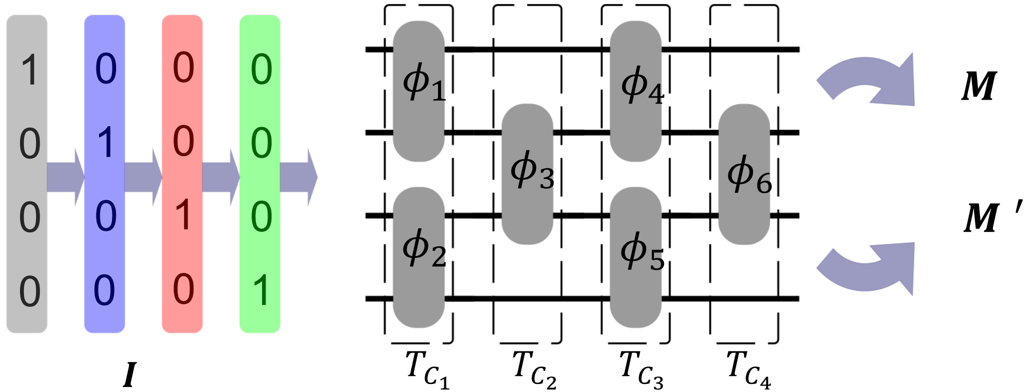
- Reduce thermal imbalance

$$\mathrm{C} = g(\boldsymbol{\phi}) + \alpha \cdot \sum_{i=1}^{N} \phi_i + \beta \cdot \sum_{i=1}^{N} \left| \phi_i - \frac{1}{|M_i|} \sum_{\phi_j \in M_i} \phi_j \right|$$

Phase update in ONN training:

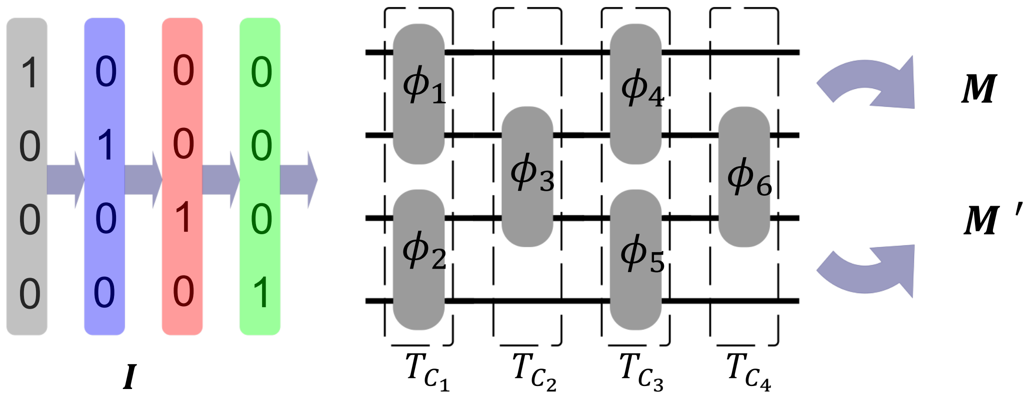
$$\phi_i = \phi_i - LR \frac{\partial \mathrm{C}}{\partial \phi_i}$$

Extraction of Process Variations with Differential Test



- First, four input patterns 1000, 0100, 0010 and 0001 are applied to the inputs of ONN. $M = T_{C_4} T_{C_3} T_{C_2} T_{C_1} I$.
- Second, the phases of MZIs in column C_4 are changed by applying a power.
- With the same input identity matrix, the ONN produces $M' = T'_{C_4} T_{C_3} T_{C_2} T_{C_1} I$.

Extraction of Process Variations with Differential Test



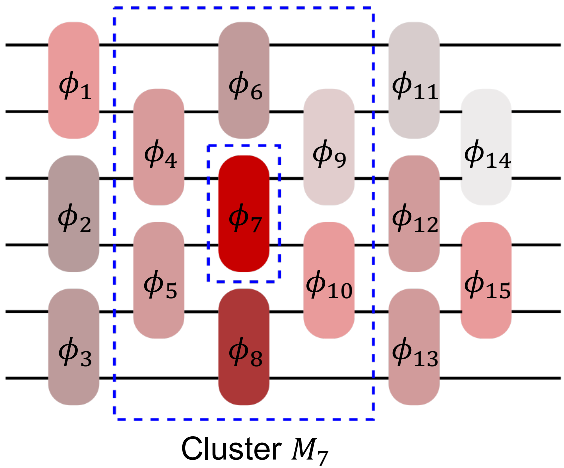
$$M'M^{-1} = \left(T'_{C_4}T_{C_3}T_{C_2}T_{C_1}\right)\left(T_{C_4}T_{C_3}T_{C_2}T_{C_1}\right)^{-1} = T'_{C_4}T_{C_4}^{-1}$$

$$= \begin{bmatrix} 1 & \mathbf{0} & 0 \\ \mathbf{0} & T'_6 & \mathbf{0} \\ 0 & \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} & 0 \\ \mathbf{0} & T^*_6 & \mathbf{0} \\ 0 & \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} & 0 \\ \mathbf{0} & T'_6T^*_6 & \mathbf{0} \\ 0 & \mathbf{0} & 1 \end{bmatrix}$$

$$T'_6T^*_6 = e^{j(\phi'_6-\phi_6)/2} \begin{bmatrix} \cos\frac{\phi'_6-\phi_6}{2} & \sin\frac{\phi'_6-\phi_6}{2} \\ -\sin\frac{\phi'_6-\phi_6}{2} & \cos\frac{\phi'_6-\phi_6}{2} \end{bmatrix} \Rightarrow \Delta\phi_6 = \phi'_6 - \phi_6$$

$\Delta\phi_6$ is used to match the characteristic curves of MZI variations.

Power Adjustment to Counter Thermal Effects



Thermal imbalance in an ONN.

- MZIs with different temperatures affect each other and cause deviations of their phases.
- To counter this deviation, instead of the target power to the k^{th} MZI, the real power applied to this MZI is modified in advance according to [1] to guarantee the expected MZI value.

[1] M. Jacques, A. Samani, E. El-Fiky, D. Patel, Z. Xing, and D. V. Plant, “Optimization of thermo-optic phase-shifter design and mitigation of thermal crosstalk on the SOI platform,” Opt. Express, vol. 27, pp. 10 456–10 471, 2019.

Accuracy of ONN without the Proposed Framework

NN	Dataset	w/o var. ext.		w/o pow. eval	
		μ	σ	μ	σ
FCNN	MNIST	11.63%	1.47%	12.07%	1.79%
LeNet-5	Cifar10	11.22%	0.86%	11.05%	0.82%
Aug.LeNet-5	Cifar10	13.76	0.82%	14.35%	0.93%

sampled ONNs: 100
3σ of the phases at 2π is 20%.

The inference accuracy of ONNs drops drastically down to an unusable level if either the process variation extraction or the power adjustment is not applied.

Ying Zhu, Grace Li Zhang, Bing Li, Xunzhao Yin, Cheng Zhuo, Huaxi Gu, Tsung-Yi Ho, Ulf Schlichtmann, "Countering Variations and Thermal Effects for Accurate Optical Neural Networks", IEEE/ACM International Conference On Computer Aided Design (ICCAD), 2020

Accuracy of ONN with the Proposed Framework

NN	Dataset	Acc. Software Training			Acc. MZI ONN		Power Reduction
		NN	ONN	ONN_p	μ	σ	
FCNN	MNIST	97.40%	97.13%	94.41%	92.80%	0.6%	25.84%
LeNet-5	Cifar10	75.56%	75.91%	74.60%	74.11%	0.26%	20.24%
Aug.LeNet-5	Cifar10	82.81%	82.08%	81.23%	80.80%	0.22%	13.27%

NN: neural network with traditional training
ONN: conventional optical neural network training
ONN_p: optical neural networks training considering process variations and thermal effects

ONNs *after testing and tuning* can maintain a similar inference accuracy in realizing neural networks while providing the advantage of high speed and low power consumption.

Thank you for your attention!