Analysis of Two MZI-Based Topologies for Optical Neural Network

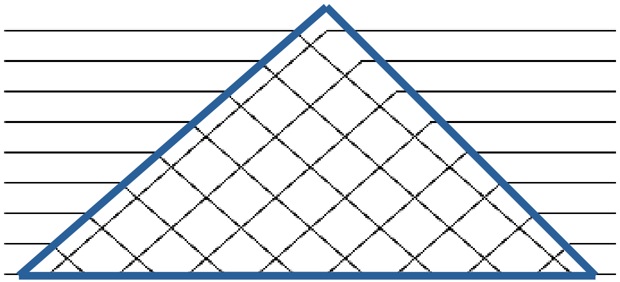
Simon Geoffroy-Gagnon
https://s-g-gagnon.research.mcgill.ca/
McGill University, Montreal
Electrical Engineering Masters

Content

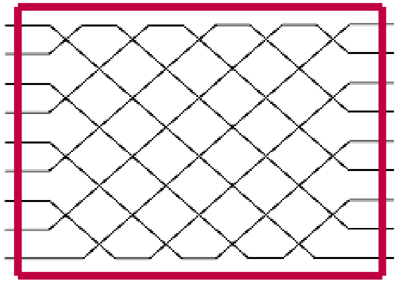
- What are Optical Processors
    - Mach-Zehnder Interferometers – the building blocks
    - The two different optical processors studied in this work
- What are Neural Networks
- How are Optics used in NNs
- Why do experimental uncertainties matter
- Results of simulations
- Future work
- Conclusion

What are MZI-Based Optical Processors?

- Use multiple Mach-Zehnder Interferometers (MZIs) connected in a mesh to compute matrix multiplications
- Can be used to create arbitrary unitary matrices
- Different mesh topologies of MZIs can create unitary matrices – Which topology is better?

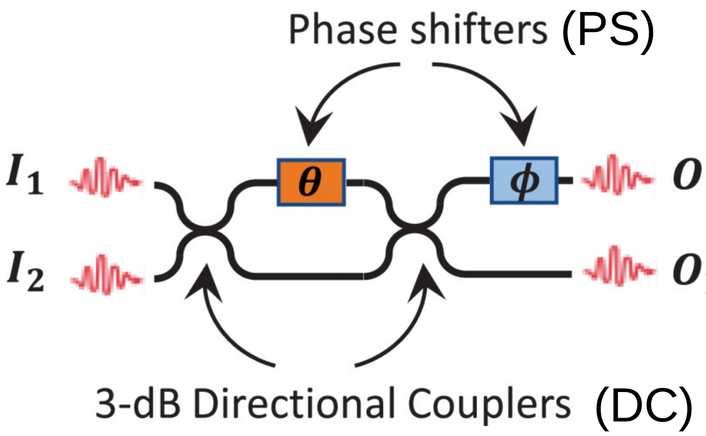


VS



Hughes, T. W. *et al.,* Phy. Rev. A., v.11 n.6 (2019).

Mach-Zehnder Interferometers
The building blocks of Optical Processors



- Reconfigurable MZI have two phase shifters after two 50:50 directional couplers
- This creates a unitary transformation matrix:

$$\mathbf{D}_{\mathrm{MZI}} = \mathbf{D}_{\mathrm{PS},\phi} \cdot \mathbf{D}_{\mathrm{DC}} \cdot \mathbf{D}_{\mathrm{PS},\theta} \cdot \mathbf{D}_{\mathrm{DC}} = je^{j\left(\frac{\theta}{2}\right)} \begin{bmatrix} e^{j\phi}\sin\left(\frac{\theta}{2}\right) & e^{j\phi}\cos\left(\frac{\theta}{2}\right) \\ \cos\left(\frac{\theta}{2}\right) & -\sin\left(\frac{\theta}{2}\right) \end{bmatrix}$$

- Connecting multiple MZIs together in a mesh creates a larger unitary matrix

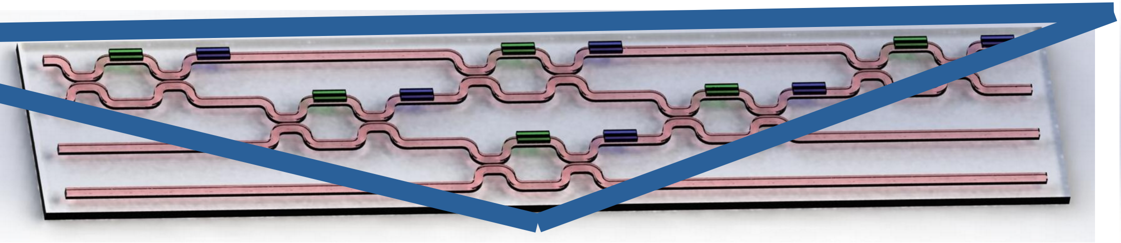
F. Shokraneh *et al,* in Asia Communications and Photonics Conference, pp. 1–3, Oct 2018.

Unitary Matrix

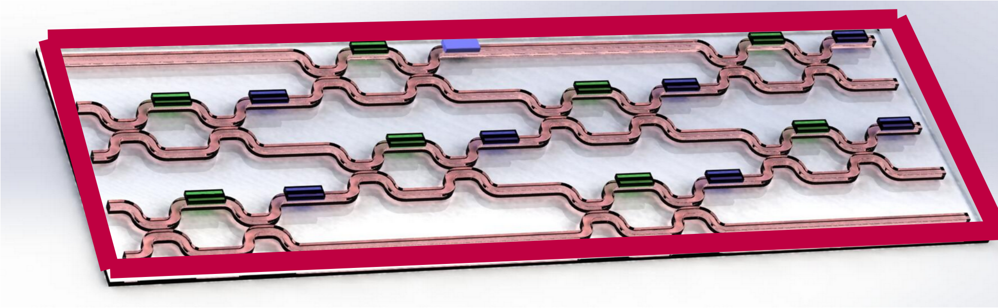
- Inverse is equal to its conjugate transpose: $\mathbf{U}^{-\mathbf{1}} = \mathbf{U}^*$
- Determinant is unitary: $|\det(\mathbf{U})| = 1$
- This causes the matrix's only possible transformation to be rotation or reflections, no stretching

$$\mathbf{D}_{\text{MZI}} = \mathbf{D}_{\text{PS},\phi} \cdot \mathbf{D}_{\text{DC}} \cdot \mathbf{D}_{\text{PS},\theta} \cdot \mathbf{D}_{\text{DC}} = je^{j\left(\frac{\theta}{2}\right)} \begin{bmatrix} e^{j\phi}\sin\left(\frac{\theta}{2}\right) & e^{j\phi}\cos\left(\frac{\theta}{2}\right) \\ \cos\left(\frac{\theta}{2}\right) & -\sin\left(\frac{\theta}{2}\right) \end{bmatrix}$$

MZI-Based Optical Processors



Triangular mesh, theorized by Reck *et al.*



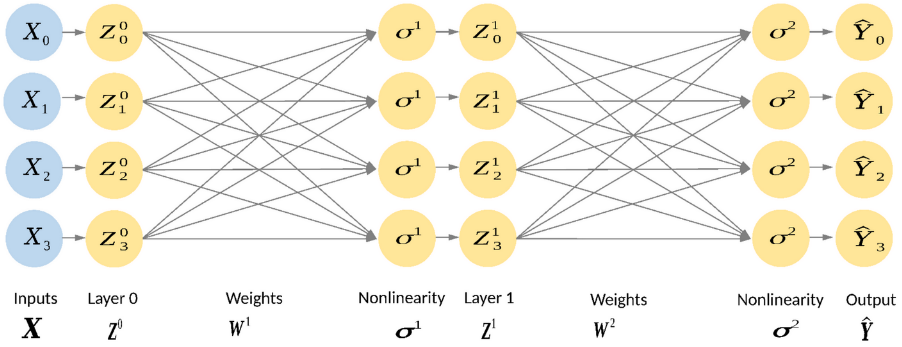
Rectangular mesh, theorized by Clements *et al.*

Each use N(N-1)/2 MZI, where N is the number of ports

Reck, M. *et al.,* Phys. Rev. Lett., v.73 n.1 (2002).
Clements, W. *et al.,* Optica, v.3 n.6 (2016).

Neural Networks – what do they do

- Neural Networks: very promising Machine Learning models used for image classification, voice recognition, autonomous car control
- Uses Matrix Multiplication extensively
- Increased energy efficiency and computational speed through the use of optics



$$\mathbf{Z}^1 = \sigma(\mathbf{W}^1(\mathbf{Z}^0))$$

$$\hat{\mathbf{Y}} = \sigma(\mathbf{W}^K(\mathbf{Z}^{K-1}))$$

Fu, J. *et al,* IEEE conference on computer vision, pp. 4438-4446 (2017).
Venayagamoorthy, G. K. *et al,* COMSIG, pp. 29-32 (1998).
Pomerleau, D. A., The Handbook of Brain Theory and Neural Networks, pp. 161-181 (1996).

Computer NN vs Optical NN

Digital Matrix Multiplication Time Complexity: O(N^2.376), since computers have to go through every element one by one and multiply them

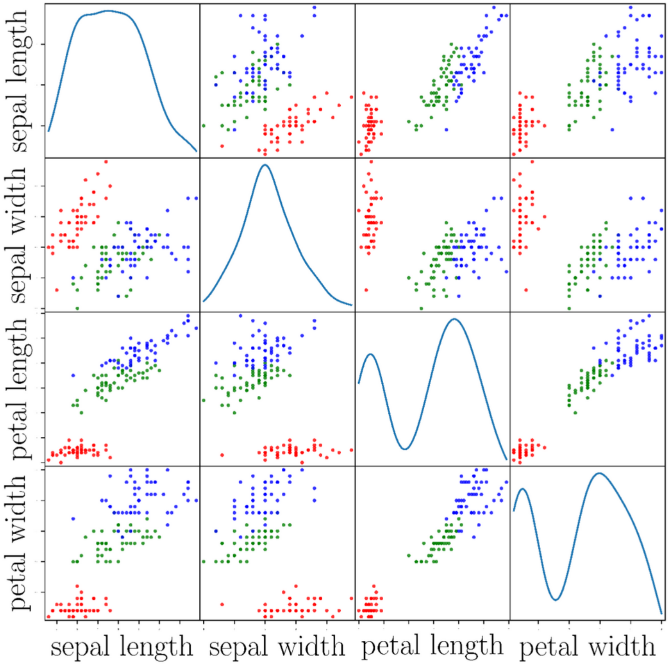
Optical Matrix Multiplication Time Complexity: O(N), because optical components have inherent parallelism

Datasets

- Neural Networks Can be used for classification
- Classifying the Iris dataset, for example

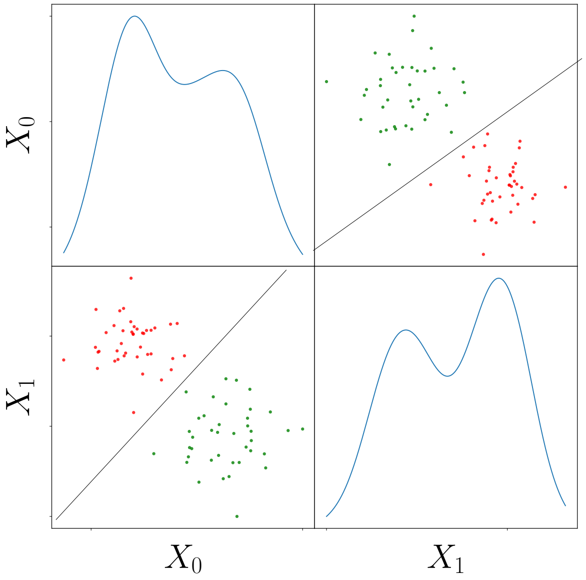
Iris Setosa (one class in the Iris dataset)

Iris dataset

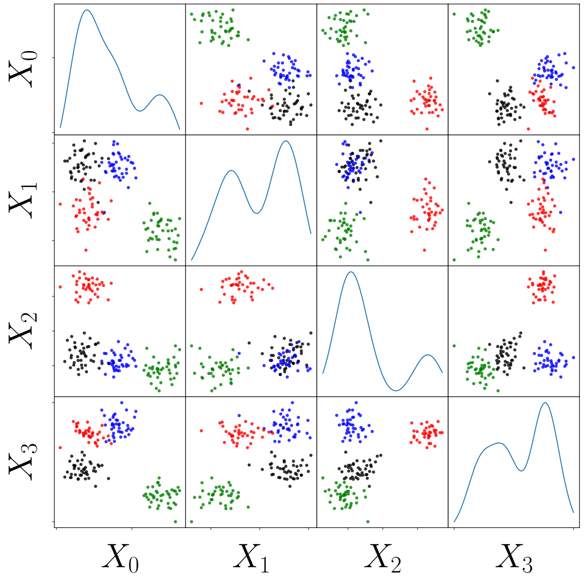


S. Sacks, https://daylily-phlox.eu/beardless-iris/other-iris-species-and-hybrids/iris-setosa-all-stripes/, Accessed: 2020-04-20.
R. A. Fisher, “The Use of Multiple Measurements in Taxonomic Problems,” Annals Eugen., vol. 7, pp. 179–188, 01 1936.

Linearly Separable Multi-variate Gaussian Datasets



Two feature dataset with two classes
Linearly separable with a straight line



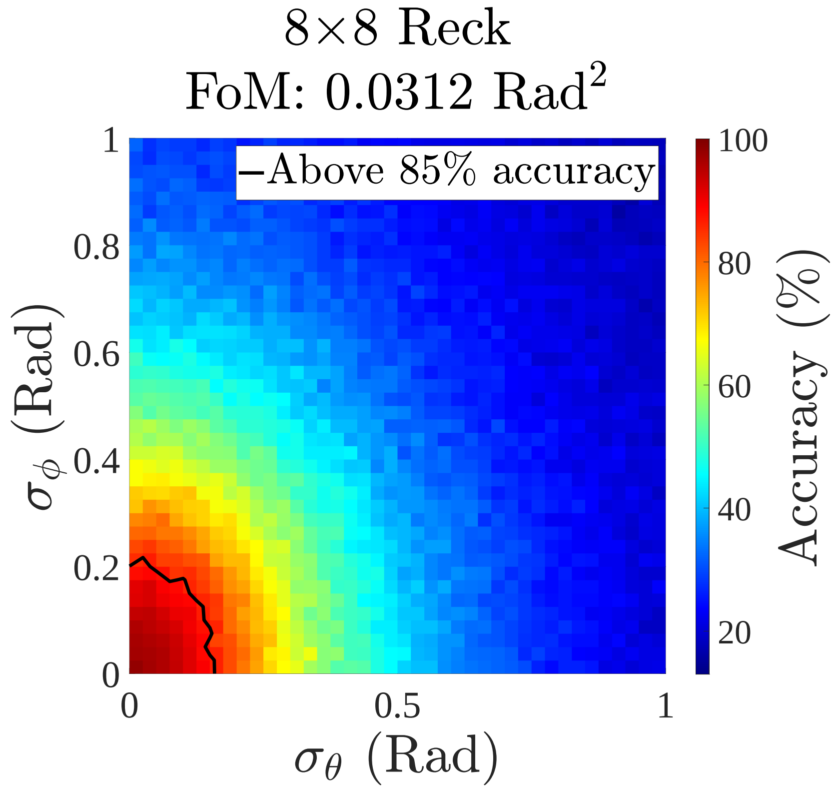
Four feature dataset with four classes
Linearly separable with a 4-dimensional hyperplane

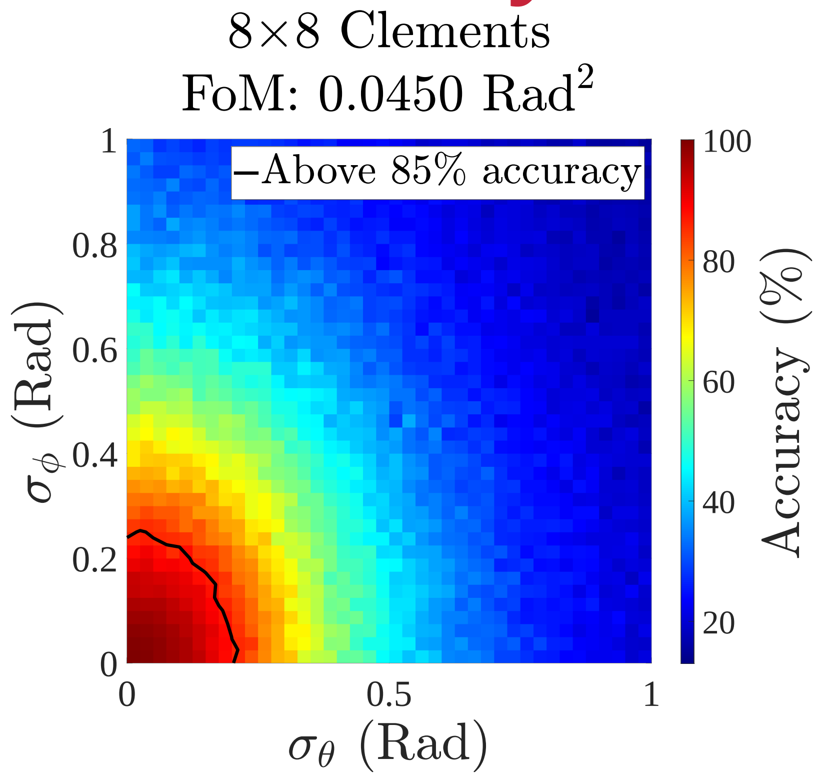
100% Classifiable with a single layer conventional neural network to give Optical Neural Network an achievable baseline

Experimental uncertainties

- Insertion Loss per MZI: Inherent loss of power
- Phase Uncertainty: Imperfect phase shifters
- Decreases the final classification accuracy of the optical NN.
- Need a way to find optimal mesh topology for tolerance to experimental uncertainties

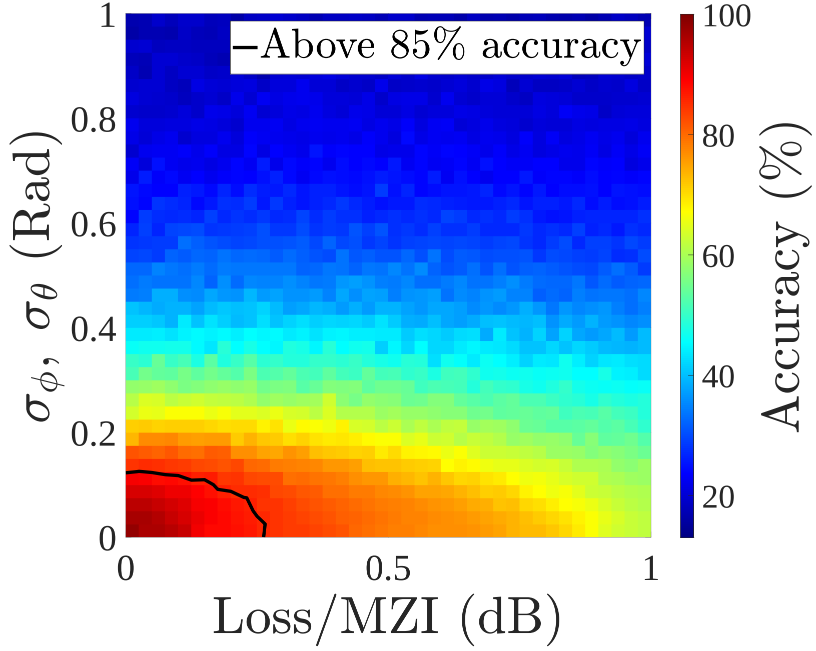
Tolerance to Phase Uncertainty



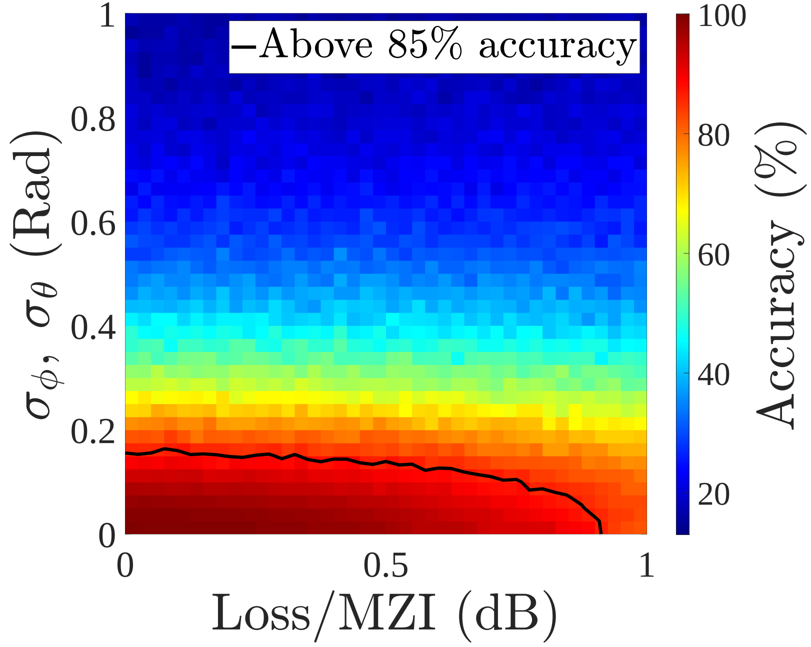


Tolerance to Insertion Loss

8×8 Reck
FoM: 0.0312 Rad · dB



8×8 Clements
FoM: 0.1319 Rad · dB



Future Work

- Currently working on a mesh that is as good as the Rectangular mesh but with the same programming method as the Triangular mesh
- Will be described in an upcoming paper

Conclusion

- Clements outperforms Reck
- Both in Phase Uncertainty Tolerance and Loss/MZI Tolerance
- This can be seen by the higher FoMs present for the Rectangular mesh
- But Rectangular mesh has more complex programming and calibration: No channels with single MZI

Bibliography

[1] F. Shokraneh, M. S. Nezami, and O. Liboiron-Ladouceur, “A 4×4 Reconfigurable Optical Processor,” in Asia Communications and Photonics Conference, pp. 1–3, Oct 2018.

[2] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, “Experimental Realization of Any Discrete Unitary Operator,” Physics Review Letters, vol. 73, pp. 58–61, Jul 1994

[3] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, “Optimal Design for Universal Multiport Interferometers,” Optica, vol. 3, pp. 1460–1465, Dec 2016.

Thank you

Questions?