Hardware-level simulations of nanophotonic neural networks

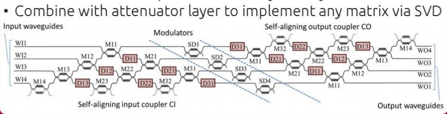
Ben Bartlett (benbartlett@stanford.edu)
Department of Applied Physics, Stanford University

Abstract

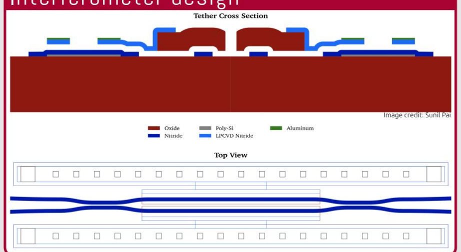
Modern computing hardware is inefficient at implementing neural networks, primarily because digital matrix multiplication is an $\Omega(N^2)$ operation. We present a fully-optical architecture consisting of meshes of self-configuring nanophotonic interferometers which is capable of performing $O(1)$ matrix multiplication on an input vector of light intensities. Using detailed physical simulations of our interferometer design, we develop a theoretical control system for our architecture which generates the on-chip layout and applied voltages necessary to implement and train an arbitrarily specified feed-forward neural network.

Introduction

- Neural networks are computationally expensive to run, even once fully trained - $O(N^{2.37})$ best complexity
- Photonic devices can use interferometric effects for constant time matrix multiplication on input vector of light intensities
- Data throughput of ~100GHz and no theoretical energy cost!
- Phase-shifted Mach Zehnder Interferometer (MZI) can perform arbitrary $SU(2)$ transformation
- Blocks of MZIs can implement arbitrary unitary transformation
- Combine with attenuator layer to implement any matrix via SVD



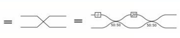
Interferometer design



Interferometric matrix multiplication

$$T_{m,n}(\theta,\phi) = \begin{bmatrix} 1 & 0 & & & & 0 \\ 0 & 1 & & & & \\ & & e^{i\phi}\cos\theta & -\sin\theta & & \\ & & e^{i\phi}\sin\theta & \cos\theta & & \\ & & & & 1 & 0 \\ 0 & & & & 0 & 1 \end{bmatrix}$$

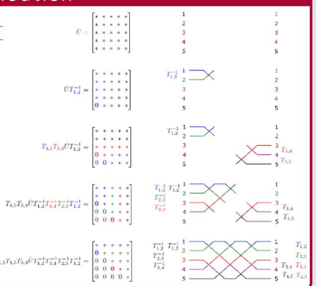
$T_{m,n}(\theta,\phi)$ =



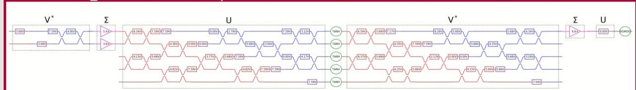
Algorithm 1 $\mathrm{U}(N) \to \mathrm{U}(2)$ matrix decomposition

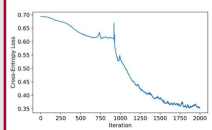
let $U \in \mathrm{U}(N)$ be an arbitrary unitary matrix and let $\hat{U} = U$
for $i = 1$ to $N-1$ **do**
  if i is odd **then**
    for $j = 0$ to $i-1$ **do**
      find θ, ϕ such that $T^{-1}_{i-j,i-j+1}(\theta,\phi)$ nullifies element $N-j, i-j$ of $\hat{U}$
      update $\hat{U} \Leftarrow \hat{U} T^{-1}_{i-j,i-j+1}(\theta,\phi)$
    end for
  else
    for $j = 1$ to i **do**
      find θ, ϕ such that $T_{N+j-i-1,N+j-i}(\theta,\phi)$ nullifies element $N+j-i, j$ of $\hat{U}$
      update $\hat{U} \Leftarrow T_{N+j-i-1,N+j-i}(\theta,\phi)\hat{U}$
    end for
  end if
end for
return $L = \{T^{-1}_{m,n}\}, \hat{U}, R = \{T_{m,n}\}$

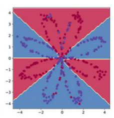
$$W = U\Sigma V^* = \left(\prod_{T_{U,L}} T^{-1}_{m,n} \cdot \hat{U} \cdot \prod_{T_{U,R}} T_{m,n}\right) \Sigma \left(\prod_{T_{V^*,L}} T^{-1}_{m,n} \cdot \hat{V}^* \cdot \prod_{T_{V^*,R}} T_{m,n}\right)$$



neuroptical: a photonic neural network simulator







```
neuroptical.components.PhotonicNeuralNetwork
m __init__(self, layerSizes, activations)
m forward_propagate(self, data)
m back_propagate(self, yhat, y)
m update_params(self, dW_list, db_list, learning_rate)
m compute_cost(self, yhat, y)
m learn(self, data, y, learning_rate, num_iterations, plot = True, showProgress = True)
m classify(self, data)
m make_drawing(self, d = None, outputs = None)
f layers
```

```
neuroptical.components.MZI
m __init__(self, N, m, n, inverted = False)
m set_params(self, theta, phi)
m unitary(self)
m operator(self)
m voltage_to_phase(self, V)
m phase_to_voltage(self, phi)
f phi
f V_phi
f V_theta
f inverted
f phase_uncert
f m
f theta
f N
f n
```

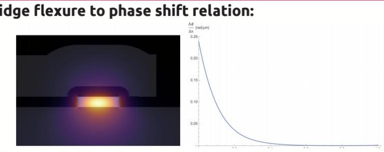
```
neuroptical.components.MZIBlock
m __init__(self, N)
m get_matrix(self)
m set_matrix(self, U)
m make_drawing(self, d = None, outputs = None)
f Tinv_block
f D
f T_block
f N
```

```
neuroptical.components.OIU
m __init__(self, inputSize, outputSize)
m get_matrix(self)
m set_matrix(self, newMatrix)
m make_drawing(self, d = None, outputs = None)
f D
f Ublock
f Vblock
```

```
neuroptical.components.NetworkLayer
m __init__(self, inputSize, outputSize, activation = "relu")
m initialize(self)
m get_W(self)
m set_W(self, newW)
m forward_pass(self, A_prev)
m activation(self, x)
m backward_pass(self, dA)
m linear_backward(self, dZ)
m activation_backward(self, dA)
m make_drawing(self, d = None, outputs = None)
f outputSize
f activationType
f cache
f inputSize
f bias
f oiu
```

Voltage to phase shift response

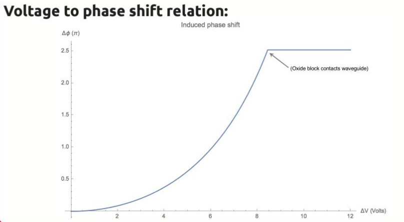
Bridge flexure to phase shift relation:



Voltage to bridge flexure relation:

- Model as Euler-Bernoulli beam: $\frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 w(x)}{\partial x^2}\right) = q(x) = \frac{\epsilon_0 \Delta V^2 \delta y}{2\,(d + w(x))}$
- Impose rigid boundary conditions: $w(-\frac{\delta x}{2}) = w(\frac{\delta x}{2}) = w'(-\frac{\delta x}{2}) = w'(\frac{\delta x}{2}) = 0$

Voltage to phase shift relation:



Future work

- Full on-chip backpropagation
- Extend fault tolerance of decomposition routine
- Efficient convolutional and recurrent architectures
- Include thermal tension from fabrication in bridge model
- Extend to low photon number regine for quantum information processing experiments

References

[1]D. A. Miller, "Self-configuring universal linear optical component," Photonics Research, vol. 1, no. 1, p. 1, 2013, ISSN: 2327-9125. DOI:10.1364/PRJ.1.000001.

[2]L. Vivien, A. Polzer, D. Marris-Morini, J. Osmond, J. M. Hartmann, P. Crozat, E. Cassan,C. Kopp, H. Zimmermann, and J. M. Fédéli, "Zero-bias 40Gbit/s germanium waveguide photodetector on silicon," Optics Express, vol. 20, no. 2, p. 1096, 2012, ISSN: 1094-4087.DOI:10.1364/OE.20.001096.

[3]D. A. Miller, "Attojoule Optoelectronics for Low-Energy Information Processing and Communications,"Journal of Lightwave Technology, vol. 35, no. 3, pp. 346–396, 2017, ISSN: 07338724.DOI:10.1109/JLT.2017.2647779.

[4]Y. Shen, N. C. Harris, S. Skirlo, D. Englund, and M. Soljacic, "Deep learning with coherent nanophotonic circuits," inSummer Topicals Meeting Series, SUM 2017, 2017, ISBN:9781509065707.DOI:10.1109/PHOSST.2017.8012714.

[5]A. Annoni, E. Guglielmi, M. Carminati, G. Ferrari, M. Sampietro, D. A. Miller, A. Melloni,and F. Morichetti, "Unscrambling light - Automatically undoing strong mixing between modes,"Light: Science and Applications, vol. 6, no. 12, 2017, ISSN: 20477538.DOI:10.1038

[6]M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental realization of any discreteunitary operator,"Physical Review Letters, vol. 73, no. 1, pp. 58–61, 1994, ISSN: 00319007.DOI:10.1103/PhysRevLett.73.58.

[7]W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, "AnOptimal Design for Universal Multiport Interferometers," no. 2, pp. 1–8, 2016, ISSN: 2334-2536.DOI:10.1364/OPTICA.3.001460.