More with less: Fault-tolerance and information-theoretic optimality in programmable photonics

Ryan Hamerly[1,2], Jasvith Raj Basani[3], Alex Sludds[1], Sri Krishna Vadlamani[1], and Dirk Englund[1]

[1]MIT RLE, 50 Vassar St., Cambridge, MA 02138
[2]NTT Research, 940 Stewart Dr., Sunnyvale, CA 94085
[3]JQI / UMD, College Park, MD 20742

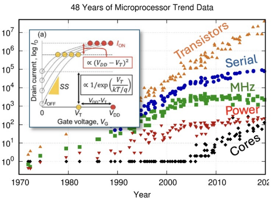
MIT AI Hardware Program, 11/13/2024

The Physical Limits of Computing

Dennard Scaling

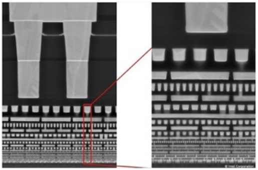
Thermodynamic voltage limit.

$$\frac{\mathrm{d}V}{\mathrm{d}\log(I)} = \frac{kT}{e} = 60\ \mathrm{mV/dec}$$



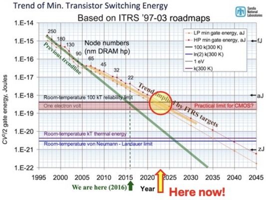
Data Movement in Cu

- Energy: $E/\mathrm{bit} = \frac{1}{4}CV^2$
- Speed: $\mathrm{bits/s} = (4 \times 10^{16})A/L^2$

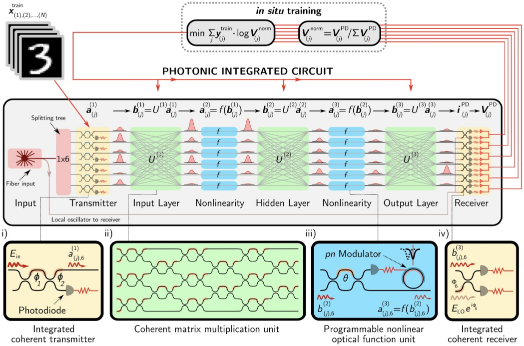


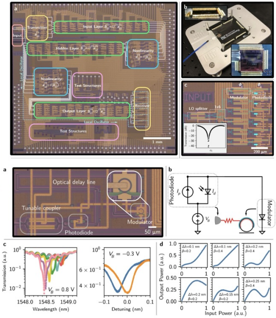
Landauer / Noise Limit

$$E_{\mathrm{bit}} \geq kT\log(2)$$



Programmable Photonics, AI, and the Future of Computing





- Leverage the bandwidth and data movement advantages of photonics in an analog compute paradigm.
- Challenge: achieving nonlinearity, **programmability**, and scalability in a single platform.

Background: Programmable Photonics

The difference between “computers” and other types of machines is that computers are programmable.

- Modern foundry processes (esp. SiPh) have made “programmable photonics” a new reality.
- But photonic components are NOT small or cheap! Need to economize.

Raises the question: what is the “optimal” photonic circuit for a given task?

- In general, problems of this nature are either:
  - NP-hard (layout problem, for gate circuits)
  - Uncomputable (Kolmogorov complexity, for Turing machines).
- Hard to generically answer the question.

Specific Example: Multiport Interferometer

Definition: any feedforward linear optical circuit with N input- and M output-ports.

- Emulates matrix multiplication $y_N = A_{M\times N} x_N$.
- Can be fixed (e.g. AWG) or programmable.

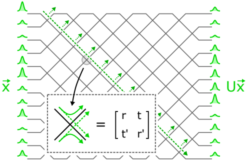
Many functions:

- Optical neural networks (of course!)
- Boson sampling, LOQC
- Optical signal processing
- Mode sorting, MIMO

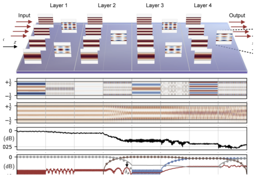
A good specific *case study* to examine the question of optimal photonic circuits.

Examples

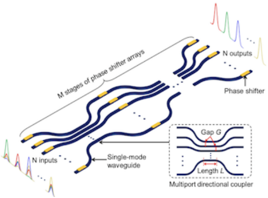
Clements Mesh



Programmable MMI



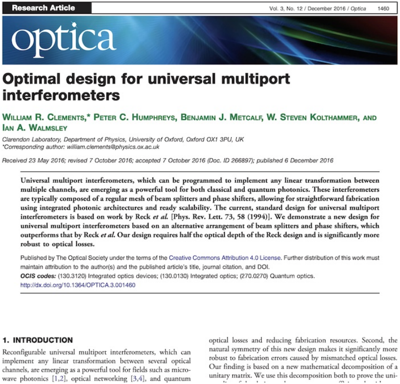
MPLC



What makes a multiport interferometer *optimal*?

Various Factors:

- Universality
- Size / Shape
- Error Robustness
- BW / length-matching
- Programming algorithm
- Efficient use of phase shift



Research Article Vol. 3, No. 12 / December 2016 / Optica 1460

optica

Optimal design for universal multiport interferometers

William R. Clements,* Peter C. Humphreys, Benjamin J. Metcalf, W. Steven Kolthammer, and Ian A. Walmsley

Clarendon Laboratory, Department of Physics, University of Oxford, Oxford OX1 3PU, UK
**Corresponding author: william.clements@physics.ox.ac.uk*

Received 23 May 2016; revised 7 October 2016; accepted 7 October 2016 (Doc. ID 266897); published 6 December 2016

Universal multiport interferometers, which can be programmed to implement any linear transformation between multiple channels, are emerging as a powerful tool for both classical and quantum photonics. These interferometers are typically composed of a regular mesh of beam splitters and phase shifters, allowing for straightforward fabrication using integrated photonic architectures and ready scalability. The current, standard design for universal multiport interferometers is based on work by Reck *et al.* [Phys. Rev. Lett. 73, 58 (1994)]. We demonstrate a new design for universal multiport interferometers based on an alternative arrangement of beam splitters and phase shifters, which outperforms that by Reck *et al.* Our design requires half the optical depth of the Reck design and is significantly more robust to optical losses.

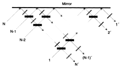
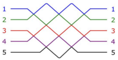
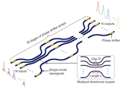
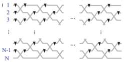
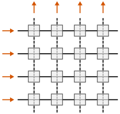
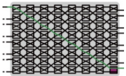
OCIS codes: (130.3120) Integrated optics devices; (130.0130) Integrated optics; (270.0270) Quantum optics.

http://dx.doi.org/10.1364/OPTICA.3.001460

1. INTRODUCTION

Reconfigurable universal multiport interferometers, which can implement any linear transformation between several optical channels, are emerging as a powerful tool for fields such as microwave photonics [1,2], optical networking [3,4], and quantum optical losses and reducing fabrication resources. Second, the natural symmetry of this new design makes it significantly more robust to fabrication errors caused by mismatched optical losses. Our finding is based on a new mathematical decomposition of a unitary matrix. We use this decomposition both to prove the uni-

What makes a multiport interferometer *optimal*?

	Unitary				**Generic**	
	Reck	Clements	MPLC	Fldzhyan	Blass	PILOSS
Size	triangle	square	??	square	diamond	square
ER*	good	good	great	great	great	good
LM†	N	Y	maybe	Y	N	Y
Alg‡	easy	moderate	hard	hard	easy	iterative
						

*ER: error robustness. †LM: length-matching. ‡Alg: algorithm. Many mesh topologies considered.[1] Previous questions on error robustness of meshes[2] were resolved by our group's work on photonic *hardware error correction*[3].

[1] M. Reck, *PRL* **73**, 58 (1994); W. Clements, *Optica* **3**(12), 1460 (2016).; R. Tanomura, *JLT* **38**(1), 60 (2020); S. Fldzhyan, *Opt. Lett.* **45**, 2632 (2020); J. Blass, in IRE Int. Conv. Record **8**(1), 48 (1960); K. Suzuki, *Opt. Exp.* **22**(4), 3887 (2014).

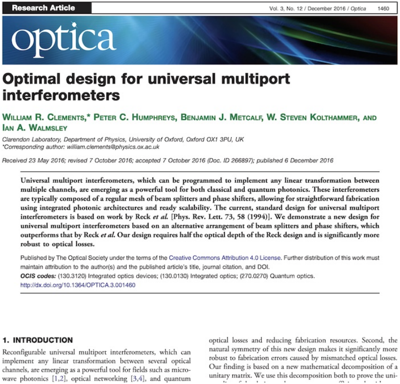
[2] M. Fang, *Opt. Exp.* **27**, 14009 (2019).

[3] S. Bandyopadhyay et al., *Optica* **8**(10), 1247 (2021); R. Hamerly et al., *PRApp* **18**, 024019 (2022); R. Hamerly et al., *Nat. Comm.* **13** (2022).

What makes a multiport interferometer *optimal*?

Various Factors:

- Universality
- Size / Shape
- Error Robustness
  $\Rightarrow$ focus of this talk
- BW / length-matching
- Programming algorithm
- Efficient use of phase shift
  $\Rightarrow$ focus of this talk


Research Article Vol. 3, No. 12 / December 2016 / Optica 1460

optica

Optimal design for universal multiport interferometers

William R. Clements,* Peter C. Humphreys, Benjamin J. Metcalf, W. Steven Kolthammer, and Ian A. Walmsley

Clarendon Laboratory, Department of Physics, University of Oxford, Oxford OX1 3PU, UK
**Corresponding author: william.clements@physics.ox.ac.uk*

Received 23 May 2016; revised 7 October 2016; accepted 7 October 2016 (Doc. ID 266897); published 6 December 2016

Universal multiport interferometers, which can be programmed to implement any linear transformation between multiple channels, are emerging as a powerful tool for both classical and quantum photonics. These interferometers are typically composed of a regular mesh of beam splitters and phase shifters, allowing for straightforward fabrication using integrated photonic architectures and ready scalability. The current, standard design for universal multiport interferometers is based on work by Reck *et al.* [Phys. Rev. Lett. 73, 58 (1994)]. We demonstrate a new design for universal multiport interferometers based on an alternative arrangement of beam splitters and phase shifters, which outperforms that by Reck *et al.* Our design requires half the optical depth of the Reck design and is significantly more robust to optical losses.

OCIS codes: (130.3120) Integrated optics devices; (130.0130) Integrated optics; (270.0270) Quantum optics.

http://dx.doi.org/10.1364/OPTICA.3.001460

1. INTRODUCTION

Reconfigurable universal multiport interferometers, which can implement any linear transformation between several optical channels, are emerging as a powerful tool for fields such as microwave photonics [1,2], optical networking [3,4], and quantum

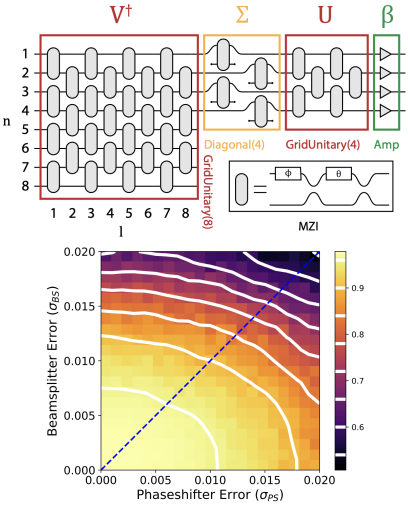
optical losses and reducing fabrication resources. Second, the natural symmetry of this new design makes it significantly more robust to fabrication errors caused by mismatched optical losses. Our finding is based on a new mathematical decomposition of a unitary matrix. We use this decomposition both to prove the uni-

(Part I) Programmable Photonics and Error Robustness

(Part I) Programmable Photonics and Error Robustness

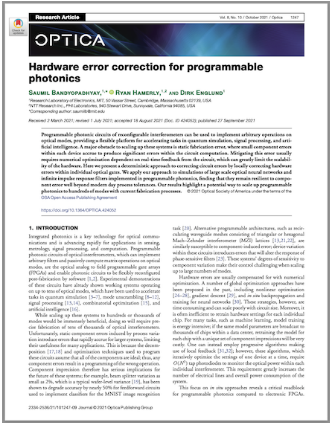
A big problem of meshes: hardware errors

- Analog systems: subject to errors
- Light passes through $O(N)$ components
- Interferometric
- Hardware errors cascade down light path
- Calibration is hard



M. Fang, *Opt. Exp.* **27**, 14009 (2019).

Hardware Error Correction

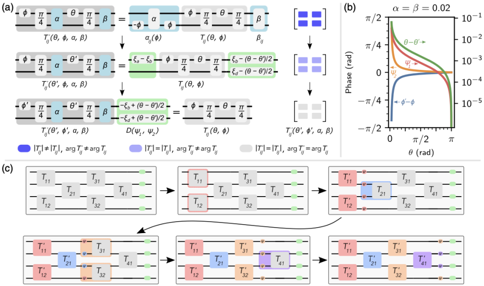

Research Article

OPTICA

Hardware error correction for programmable photonics

Saumil Bandyopadhyay, Ryan Hamerly, and Dirk Englund

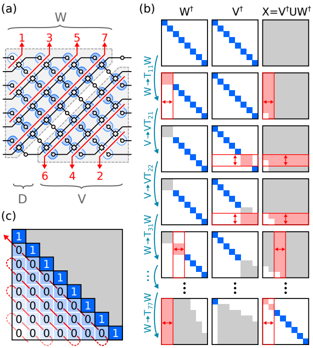
1. INTRODUCTION

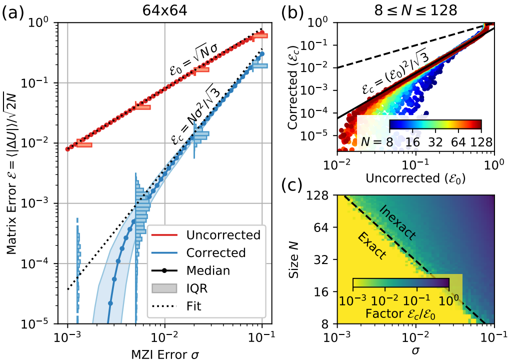


- If errors are known, then there is a fast, easy way to correct them.
- But relies on accurate pre-calibration.

S. Bandyopadhyay et al., *Optica* **8**(10), 1247 (2021).

Error Correction by Self-Configuration



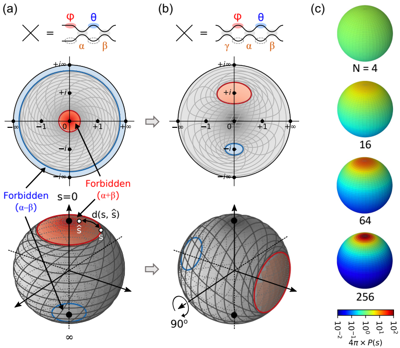


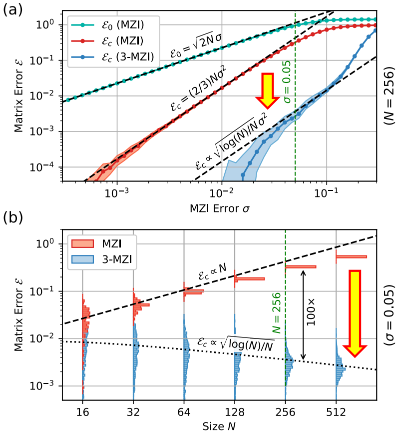
Through a feedback mechanism,

- It is possible to correct errors by self-configuring without calibration
- Never learn what the errors are! More robust.

R. Hamerly et al., *PRApp* **18**, 024018 (2022); *PRApp* **18**, 024019 (2022).

Asymptotic Fault-Tolerance of 3-MZI Mesh





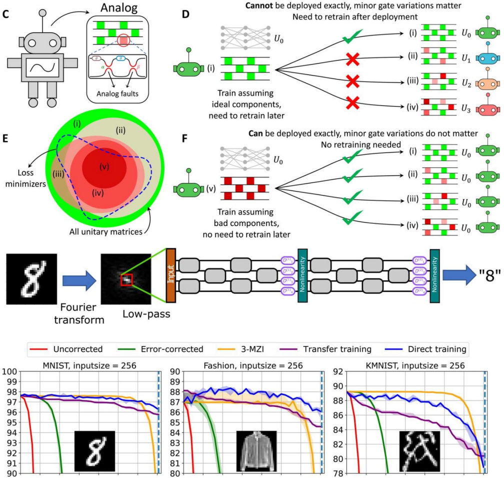
Adding third splitter to each MZI:

1. Mobius transform to locally Cartesian frame, high ER
2. Facilitates error correction, asymptotically fault-tolerant

R. Hamerly et al., *Nat. Comm.* **13** (2022).

Error-Aware “Corner Training” for Meshes

1. Train on ideal circuit, errors degrade performance (“uncorrected”)
2. Train on ideal circuit, use hardware error correction to mitigate errors on real HW (“corrected”)
3. Train on *circuit with large fixed errors*, use HEC to eliminate errors on real HW (“corner”)



S. Vadlamani et al., Sci. Adv. 9, eadh3436 (2023)

(Part II) Programmable Photonics and "Efficiency"

(Part II) Programmable Photonics and "Efficiency"

In photonics, $\Delta n \ll 1$, which means phase is expensive. In general, it is hard to make phase shifters that are simultaneously

1. low-loss,
2. compact,
3. low-power,
4. fully $[0, 2\pi)$-tunable, and
5. fast.

Cost of phase depends on the platform:

- Thermo-optic phase shifters: heater power $P = P_\pi \times \psi/\pi$, with typical $P_\pi = 20$ mW. (For a 64x64 Clements mesh, that's a total of 80 W!). Sensitive to *total phase* $\sum_i \psi_i$.
- MOSCAP & PCM phase shifters. These phase shifters induce a loss $\alpha = \alpha_\pi \times \psi/\pi$. Total loss is sensitive to total phase. Also can lead to unbalanced-loss errors, sensitive to RMS phase $\left(\sum_i \psi_i^2\right)^{1/2}$.
- Pockels phase shifters. Here, phase is limited by voltage via $V_\pi L$. Hence, the maximum phase $\psi_{\max}$ (usually 2π) is most relevant.

Summary of Argument

1. System: Multiport Interferometer $\Leftrightarrow$ a map between a set of phase shifts $\{\psi_i\}$ and a unitary matrix U.
2. Objective: Minimize the phase shift, quantified by moments $\|\psi\|_0 = |\psi|_{\max}, \quad \|\psi\|_1 = \sum_i |\psi_i|, \quad \|\psi\|_2 = \left(\sum_i \psi_i^2\right)^{1/2}$ assuming some distribution of target matrices (here Haar-random unitaries).
3. SoA: Reck and Clements MZI meshes.
4. Improvement: Use of 3-MZI instead of MZI in meshes.
5. Bound: Derive lower bounds on the moments of ψ based on information entropy. Show that the proposed 3-MZI mesh is *near-optimal* when compared to this bound.

Summary of Argument

1. **System: Multiport Interferometer $\Leftrightarrow$ a map between a set of phase shifts $\{\psi_i\}$ and a unitary matrix U.**
2. Objective: Minimize the phase shift, quantified by moments $\|\psi\|_0 = |\psi|_{\max}, \quad \|\psi\|_1 = \sum_i |\psi_i|, \quad \|\psi\|_2 = \left(\sum_i \psi_i^2\right)^{1/2}$ assuming some distribution of target matrices (here Haar-random unitaries).
3. SoA: Reck and Clements MZI meshes.
4. Improvement: Use of 3-MZI instead of MZI in meshes.
5. Bound: Derive lower bounds on the moments of ψ based on information entropy. Show that the proposed 3-MZI mesh is *near-optimal* when compared to this bound.

Summary of Argument

1. System: Multiport Interferometer $\Leftrightarrow$ a map between a set of phase shifts $\{\psi_i\}$ and a unitary matrix U.
2. **Objective: Minimize the phase shift, i.e. the moments** $\|\psi\|_0 = |\psi|_{\max}, \quad \|\psi\|_1 = \sum_i |\psi_i|, \quad \|\psi\|_2 = \left(\sum_i \psi_i^2\right)^{1/2}$ **assuming some distribution of target matrices (here Haar-random unitaries).**
3. SoA: Reck and Clements MZI meshes.
4. Improvement: Use of 3-MZI instead of MZI in meshes.
5. Bound: Derive lower bounds on the moments of ψ based on information entropy. Show that the proposed 3-MZI mesh is *near-optimal* when compared to this bound.

Summary of Argument

1. System: Multiport Interferometer $\Leftrightarrow$ a map between a set of phase shifts $\{\psi_i\}$ and a unitary matrix U.
2. Objective: Minimize the phase shift, quantified by moments $\|\psi\|_0 = |\psi|_{\max}, \quad \|\psi\|_1 = \sum_i |\psi_i|, \quad \|\psi\|_2 = \left(\sum_i \psi_i^2\right)^{1/2}$ assuming some distribution of target matrices (here Haar-random unitaries).
3. **SoA: Reck and Clements MZI meshes.**
4. Improvement: Use of 3-MZI instead of MZI in meshes.
5. Bound: Derive lower bounds on the moments of ψ based on information entropy. Show that the proposed 3-MZI mesh is *near-optimal* when compared to this bound.

Summary of Argument

1. System: Multiport Interferometer $\Leftrightarrow$ a map between a set of phase shifts $\{\psi_i\}$ and a unitary matrix U.
2. Objective: Minimize the phase shift, quantified by moments $\|\psi\|_0 = |\psi|_{\max}, \quad \|\psi\|_1 = \sum_i |\psi_i|, \quad \|\psi\|_2 = \left(\sum_i \psi_i^2\right)^{1/2}$ assuming some distribution of target matrices (here Haar-random unitaries).
3. SoA: Reck and Clements MZI meshes.
4. **Improvement: Use of 3-MZI instead of MZI in meshes.**
5. Bound: Derive lower bounds on the moments of ψ based on information entropy. Show that the proposed 3-MZI mesh is *near-optimal* when compared to this bound.

Summary of Argument

1. System: Multiport Interferometer $\Leftrightarrow$ a map between a set of phase shifts $\{\psi_i\}$ and a unitary matrix U.
2. Objective: Minimize the phase shift, quantified by moments $\|\psi\|_0 = |\psi|_{\max}, \quad \|\psi\|_1 = \sum_i |\psi_i|, \quad \|\psi\|_2 = \left(\sum_i \psi_i^2\right)^{1/2}$ assuming some distribution of target matrices (here Haar-random unitaries).
3. SoA: Reck and Clements MZI meshes.
4. Improvement: Use of 3-MZI instead of MZI in meshes.
5. **Bound: Derive lower bounds on the moments of ψ based on information entropy. Show that the proposed 3-MZI mesh is *near-optimal* when compared to this bound.**

Summary of Argument

1. System: Multiport Interferometer $\Leftrightarrow$ a map between a set of phase shifts $\{\psi_i\}$ and a unitary matrix U.
2. **Objective: Minimize the phase shift, i.e. the moments**
   $\|\psi\|_0 = |\psi|_{\max}, \quad \|\psi\|_1 = \sum_i |\psi_i|, \quad \|\psi\|_2 = \left(\sum_i \psi_i^2\right)^{1/2}$
   assuming some distribution of target matrices (here Haar-random unitaries).
3. SoA: Reck and Clements MZI meshes.
4. Improvement: Use of 3-MZI instead of MZI in meshes.
5. Bound: Derive lower bounds on the moments of ψ based on information entropy. Show that the proposed 3-MZI mesh is *near-optimal* when compared to this bound.

FOM 1: L_∞ Norm

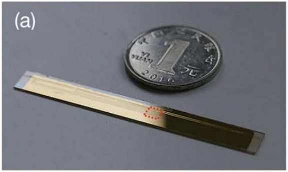
L_∞ norm is defined by the peak-to-peak variation of ψ: $\psi_{pp} \equiv 2\|\psi\|_\infty \equiv \psi_{\max} - \psi_{\min}$

- Fast lossless pure-phase modulation (Pockels, piezo-optomechanical) is very inefficient, e.g. $\Delta n \sim 10^{-4}$.
- If full 2π tunability ($\psi \in [-\pi, +\pi]$, $\psi_{pp} = \pi$), is required, the phase-shifter length is limited by:

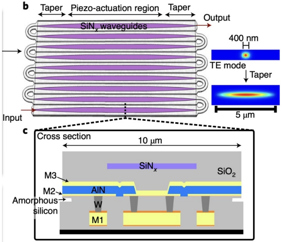
$$L \geq \frac{2 V_\pi L}{V_{pp}}$$

  which leads to a severe voltage-length tradeoff.
- If phase range can be reduced to $\psi_{pp} \ll \pi$, then $L \to L\psi_{pp}/\pi$ reduces correspondingly.

Here, we use the IQR as a stand-in for ψ_{pp}.



R. Wu, *Opt. Lett.* **44**, 4698 (2019).



M. Dong, *Nat. Phot.* **16**, 59 (2022).

FOM 2: L1 Norm

L1 norm is defined by the mean value of $|\psi|$: $\|\psi\|_1 = \langle|\psi|\rangle$. This affects:

- Average per-MZI heating power of TOPS meshes:

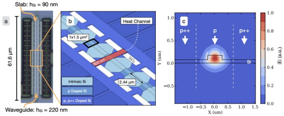
  $$P_{\mathrm{MZI}} = P_\pi \|\psi\|_1/\pi$$

  Typically $P_\pi \approx 20$ mW, although underetch can reduce this by about 10×. Total power is multiplied by the number of MZIs: $N(N-1)/2$.

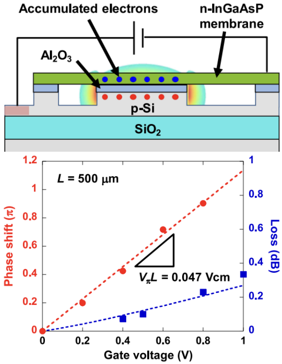
- Average excess loss due to phase shifts on lossy platforms such as MOSCAPs, where $\alpha \propto \psi$:

  $$\alpha_{\mathrm{MZI}} = \alpha_\pi \|\psi\|_1/\pi$$

  For silicon MOSCAP, $\alpha_\pi \sim 1$ dB. For III-V MOSCAP, $\alpha_\pi = 0.23$ dB. The total mesh loss is multiplied by the number of columns $\alpha_{\mathrm{tot}} = (N\alpha_\pi/\pi)\|\psi\| + \alpha_\pi$.



N. Harris, *Opt. Exp.* **22**, 10487 (2014).



M. Takenaka, *JLT* **37**(5), 1474 (2019).

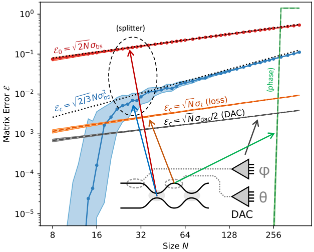
FOM 3: L2 Norm

L2 norm is defined by the RMS mean value of $|\psi|$: $\|\psi\|_2 = \langle|\psi|^2\rangle^{1/2}$. This affects:

- The average "uncorrectable" error due to lossy phase shifters. Each lossy phase shifter induces an uncorrectable non-unitary error $\|\Delta U_{\rm ps}\| = \alpha_\pi|\psi|/2\pi$.
- The errors add in quadrature, so $\mathcal{E} = \langle\|\Delta U_{\rm ps}\|\rangle/\sqrt{N}$ is given by:

$$\mathcal{E} = \alpha_\pi\sqrt{N\|\psi\|_2/2\pi}$$

This is a quadrature sum of the MZI errors and the errors due to the random phase screen.



For discussion on non-unitary errors and correction, see:
R. Hamerly et al., *PRApp* **18**, 024019 (2022).

Summary of Argument

1. System: Multiport Interferometer $\Leftrightarrow$ a map between a set of phase shifts $\{\psi_i\}$ and a unitary matrix U.
2. Objective: Minimize the phase shift, quantified by moments $\|\psi\|_0 = |\psi|_{\max}, \quad \|\psi\|_1 = \sum_i |\psi_i|, \quad \|\psi\|_2 = \left(\sum_i \psi_i^2\right)^{1/2}$ assuming some distribution of target matrices (here Haar-random unitaries).
3. **SoA: Reck and Clements MZI meshes.**
4. Improvement: Use of 3-MZI instead of MZI in meshes.
5. Bound: Derive lower bounds on the moments of ψ based on information entropy. Show that the proposed 3-MZI mesh is *near-optimal* when compared to this bound.

Splitting Ratios in Reck/Clements MZI meshes

A mesh is a cascade of 2×2 crossings $U = \prod_m T_m$, where

$$T = \begin{bmatrix} r & t \\ t' & r' \end{bmatrix}$$

For Haar-random U, r is the quotient of two χ^2-distributed random variables $X_k \sim \chi^2(k)$:

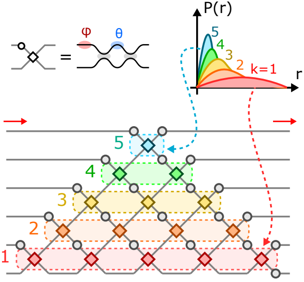
$$r = \frac{X_1}{X_1 + X_k}$$

and is given by [11]

$$P(|r|) = 2k|r|(1 - |r|2)^{k-1} \qquad (1)$$

This clusters near $r = 0$ (crossing state) for large k. A Reck / Clements mesh has $N - k - 1$ MZIs of order k

$\Rightarrow$ in large meshes, most MZIs are crosslike.

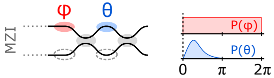


M. Reck, *PRL* **73**, 58 (1994)
W. Clements, *Optica* **3**(12), 1460 (2016)
N. Russell, *NJP* **19**(3), 033007 (2017)

Regular MZI

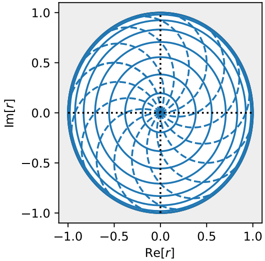
An internal phase shifter θ to control the magnitude $|r|$, and an external shifter ϕ to control its phase.

$$T_{\text{MZI}} = \frac{1}{2}\begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}\begin{bmatrix} e^{i\theta} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}\begin{bmatrix} e^{i\phi} & 0 \\ 0 & 1 \end{bmatrix}$$
$$= ie^{i\theta/2}\begin{bmatrix} e^{i\phi}\sin(\theta/2) & \cos(\theta/2) \\ e^{i\phi}\cos(\theta/2) & -\sin(\theta/2) \end{bmatrix}$$



Magnitude $|r| = \sin(\theta/2)$ only depends on θ, while ϕ only affects the angle $\arg(r) = \phi + \theta/2$.

- For small r,
  - $\theta \approx 0$
  - $\phi \in [0, 2\pi)$ uniform
    $\Rightarrow$ average phase shift of $\pi/2$
- Full control over amplitude and phase of splitting ratio: $s \equiv r/t = e^{i\phi}\tan(\theta/2) \Rightarrow$ easy to program & self-configure.



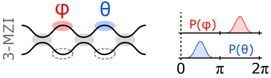
Summary of Argument

1. System: Multiport Interferometer $\Leftrightarrow$ a map between a set of phase shifts $\{\phi_i\}$ and a unitary matrix U.
2. Objective: Minimize the phase shift, quantified by moments $\|\psi\|_0 = |\psi|_{\max}, \quad \|\psi\|_1 = \sum_i |\psi_i|, \quad \|\psi\|_2 = \left(\sum_i \psi_i^2\right)^{1/2}$ assuming some distribution of target matrices (here Haar-random unitaries).
3. SoA: Reck and Clements MZI meshes.
4. **Improvement: Use of 3-MZI instead of MZI in meshes.**
5. Bound: Derive lower bounds on the moments of ψ based on information entropy. Show that the proposed 3-MZI mesh is *near-optimal* when compared to this bound.

3-MZI

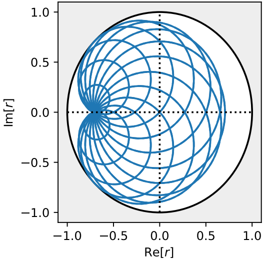
A third splitter encloses the "external" MZI:[a]

$$T_{\text{3-MZI}} = \frac{1}{2^{3/2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} e^{i\theta} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} e^{i\phi} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$$
$$= \frac{e^{i(\theta+\phi)/2}}{\sqrt{2}} \begin{bmatrix} -\cos\left(\frac{\theta-\phi}{2}\right) + i\sin\left(\frac{\theta+\phi}{2}\right) & -\sin\left(\frac{\theta-\phi}{2}\right) + i\cos\left(\frac{\theta+\phi}{2}\right) \\ \sin\left(\frac{\theta-\phi}{2}\right) + i\cos\left(\frac{\theta+\phi}{2}\right) & -\cos\left(\frac{\theta-\phi}{2}\right) - i\sin\left(\frac{\theta+\phi}{2}\right) \end{bmatrix} \quad (2)$$



Now, for $r \approx 0$, the dependence is *locally Cartesian*. The cross-state $r = 0$ does not lie at a singularity $|\partial(r, r^*)/\partial(\theta, \phi)| = 0$, as in the MZI. Implications:

- Intrinsic robustness to hardware errors.[b]
- Distributions of (θ, ϕ) are more compact
  $\Rightarrow$ can implement a mesh with less phase.



[a] K. Suzuki, *Opt. Exp.* **23**, 9086 (2015).
[b] R. Hamerly, *Nat. Comm.* **13** (2022).

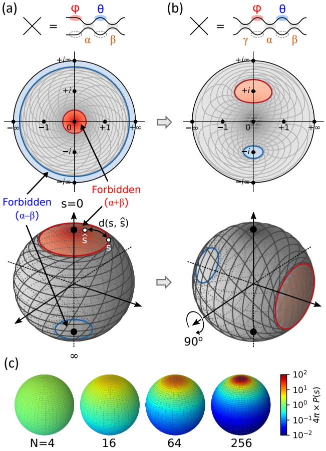
Relation to Möbius Transformations

The math is simpler when we work with splitting ratios $s = r/t$.

- Ratio denotes equivalence class w.r.t. external phase shifts, if $s = s'$, $T = T' e^{i\Psi}$.
- For standard MZI, $s = e^{i\phi}\tan(\theta/2)$, phase shifts independently control magnitude and phase of s.
- Prepending a splitter performs the Möbius transformation

$$s_{\text{3-MZI}} = \frac{s_{\text{MZI}} + i}{1 + i s_{\text{MZI}}} \approx -\frac{i}{2}(\Delta\theta + i\Delta\phi)$$

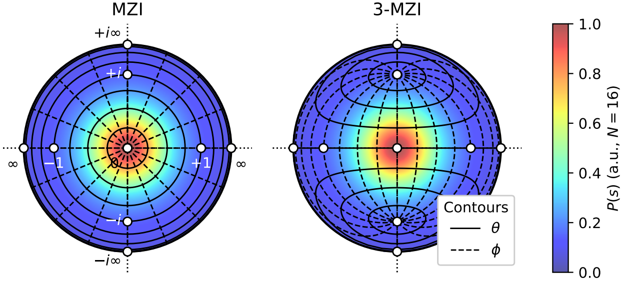
$\Rightarrow$ 90$^{\text{o}}$ rotation of Riemann sphere.



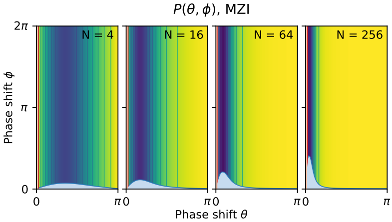
Relation to Möbius Transformations

This shows the same insight in a different basis.

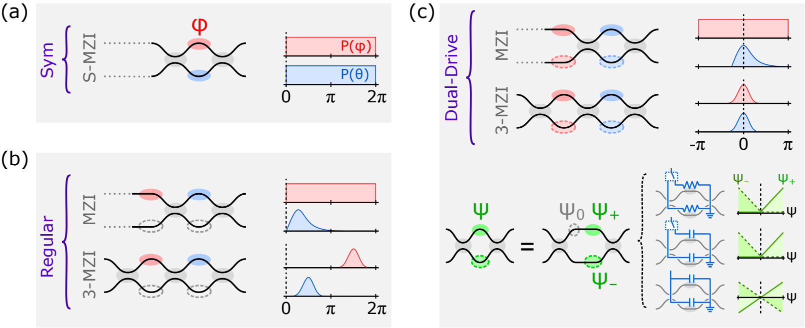
- MZI maps $(\theta, \phi) \to s$ via polar coordinates (bad).
- 3-MZI maps $(\theta, \phi) \to s$ via (locally) Cartesian coordinates (good).



Probability Density in (θ, ϕ)



Dual-Drive MZI & 3-MZI: minimizing total phase

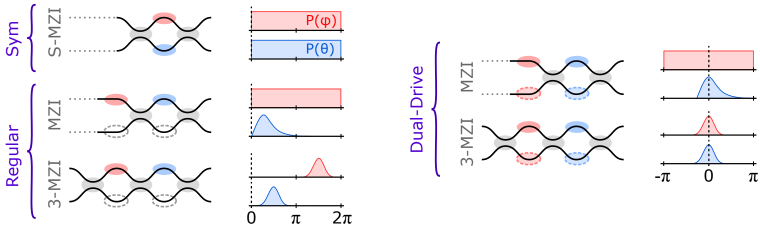


- Final problems: (i) residual phase shifts on 3MZI, i.e. phase range is small, but overall phase is still large.
- Fix with a phase offset and a dual-drive configuration:

$$\psi_\pm = \begin{cases} \pm\psi/2 & \text{Push-Pull (PP)} \\ \max(\pm\psi, 0) & \text{Single-Arm (SA)} \end{cases}$$

Comparison of MZIs on FOMs

Crossing	$\|\psi\|_1$	$\|\psi\|_2$	$\psi_{pp,\mathrm{IQR}}$
MZI	$\pi + O(N^{-1/2})$	$\sqrt{2/3}\pi + O(N^{-1})$	π
3-MZI	π	$\sqrt{5/4}\pi$	$3.8/\sqrt{N}$
MZI-DD	$\pi/2 + O(N^{-1/2})$	$\pi/\sqrt{6} + O(N^{-1})$	$\pi/2$
3-MZI-DD	$\frac{16}{3}/\sqrt{\pi N}$	$\sqrt{4\log(N/N_0'')/N}$	$1.9/\sqrt{N}$
Ratio DD/SA	$0.47 + 0.26\sqrt{N}$	$0.64\sqrt{N/\log(N)}$	$0.83\sqrt{N}$

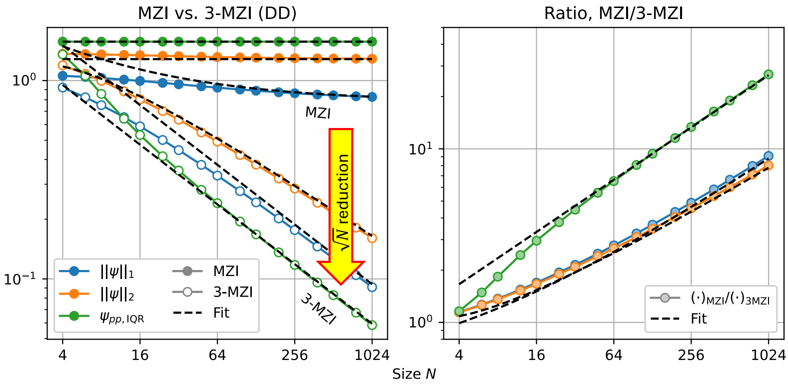


Comparison of MZIs on FOMs

Crossing	$\|\psi\|_1$	$\|\psi\|_2$	$\psi_{pp,\mathrm{IQR}}$
MZI	$\pi + O(N^{-1/2})$	$\sqrt{2/3}\pi + O(N^{-1})$	π
3-MZI	π	$\sqrt{5/4}\pi$	$3.8/\sqrt{N}$
MZI-DD	$\pi/2 + O(N^{-1/2})$	$\pi/\sqrt{6} + O(N^{-1})$	$\pi/2$
3-MZI-DD	$\frac{16}{3}/\sqrt{\pi N}$	$\sqrt{4\log(N/N_0'')/N}$	$1.9/\sqrt{N}$
Ratio DD/SA	$0.47 + 0.26\sqrt{N}$	$0.64\sqrt{N/\log(N)}$	$0.83\sqrt{N}$

- With dual-drive and 3-MZI, average phase scales as $1/\sqrt{N}$ rather than a constant.

Numerical Simulations and Scaling



Look at large mesh sizes with Haar-random target unitaries.

- $\|\psi\| \sim O(1)$ for MZI, $O(1/\sqrt{N})$ for 3-MZI.
- Roughly $10\times$ reduction of all moments $\|\psi\|_1$, $\|\psi\|_2$, $\psi_{pp,\mathrm{IQR}}$ for large meshes $N \gtrsim 256$, when using the 3-MZI.

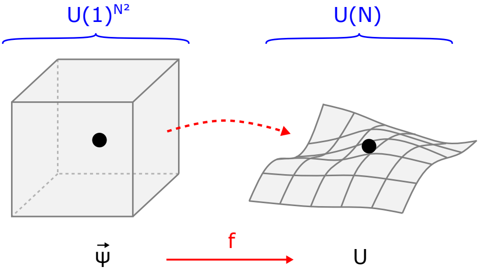
Summary of Argument

1. System: Multiport Interferometer $\Leftrightarrow$ a map between a set of phase shifts $\{\phi_i\}$ and a unitary matrix U.
2. Objective: Minimize the phase shift, quantified by moments
$\|\psi\|_0 = |\psi|_{\max}, \quad \|\psi\|_1 = \sum_i |\psi_i|, \quad \|\psi\|_2 = \left(\sum_i \psi_i^2\right)^{1/2}$
assuming some distribution of target matrices (here Haar-random unitaries).
3. SoA: Reck and Clements MZI meshes.
4. Improvement: Use of 3-MZI instead of MZI in meshes.
5. **Bound: Derive lower bounds on the moments of ψ based on information entropy. Show that the proposed 3-MZI mesh is *near-optimal* when compared to this bound.**

Method for Deriving a Lower Bound

We find a bound for the moments of ψ (i.e. the total phase shift) by examining the map $f : \mathbb{R}^{N^2} \to U(N)$ that maps phase shifts $\{\psi_k\}$ to unitaries U. The procedure involves two steps:

1. Show that the map f is *contractive*, and calculate the minimum loss information entropy ΔH.
2. Use this entropy to find a lower bound for the moments of ψ.



The Contractive Map f

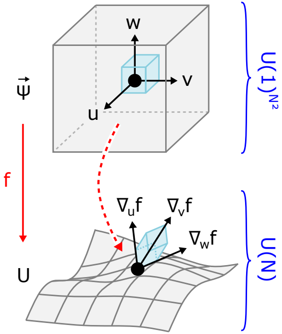
We care about the contraction of f, i.e. how it maps volumes on $\mathbb{R}^{N^2}$ to $U(N)$. These two are related by:

$$\mathrm{d}V_U = |\det(J)|\mathrm{d}V_\psi$$

$J = \partial(U)/\partial(\psi)$ is the Jacobian.
Note that, for dual-drive modulation, the dependence of U on any ψ_i takes the form:

$$U = U_{\mathrm{post}} \begin{bmatrix} e^{i\psi_i/2} & 0 \\ 0 & e^{-i\psi_i/2} \end{bmatrix} U_{\mathrm{pre}} \tag{3}$$

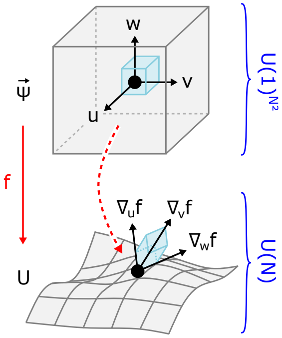
Since U_{pre} & U_{post} are unitary, $\|\Delta U\| = 2^{-1/2}\Delta\psi_i$, and therefore $|\nabla_{\psi_i} f| = 2^{-1/2}$ along all directions ψ_i, so the map is contractive.
(A more generic formula, which does not rely on dual-drive modulation, leads to a contraction by 1 along all directions, and bounds that are looser by a factor of $\sqrt{2}$.)



The Contractive Map f (cont.)

Minimum Contraction

- To keep the phase shifts small, we want a mapping (i.e. an architecture) that minimizes the contraction.
- This happens when the $\nabla_{\psi_i} f$ are all orthogonal, so the compression factor is $|\det(J)| = 2^{-N^2/2}$ ($2^{-1/2}$ per phase shifter).



Information Entropy

- Defined by $H = -\int P \log(P) \mathrm{d}V$
- Can related the entropy of distribution $P(U)$ to the pullback $P(\psi)$ using the relation $P(\psi)\mathrm{d}\psi = P(U)\mathrm{d}U$:

$$H_\psi = -\int P(\psi)\log P(\psi)\mathrm{d}V_\psi = -\int \left[P(U)\log\left(\frac{P(\psi)}{P(U)}\right) + P(U)\log P(U)\right]\mathrm{d}V_U$$

$$= H_U + \underbrace{\left\langle \log \frac{1}{|\det(J)|}\right\rangle}_{\Delta H} \geq \underbrace{H_U + (N^2/2)\log(2)}_{H_{\psi,\mathrm{min}}}$$

The Contractive Map f (cont.)

Start with the entropy bound

$$H_\psi = H_U + \left\langle \log \frac{1}{|\det(J)|} \right\rangle$$

First we get H_U, using the volume[a] of $U(N)$ and Stirling's approximation:

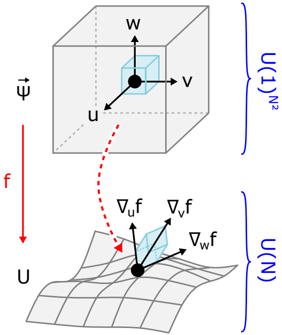
$$\mathrm{vol}(U(N)) = (2\pi)^{N(N+1)/2} \prod_{k=1}^{N-1} \frac{1}{k!}$$

$$\Rightarrow H_U = -\log[\mathrm{vol}(U(N))] = \frac{N^2}{2} \log\left(\frac{2\pi e^{3/2}}{N}\right)$$

Given $|\det(J)| \leq 2^{-N^2/2}$, we find:

$$H_\psi > H_{\psi,\min} = \frac{N^2}{2} \log\left(\frac{4\pi e^{3/2}}{N}\right)$$

for all pullbacks ψ of the Haar measure.



[a] L. Boya, Rpt. Math. Phys. **52**(3), 401 (2003).

Bound on Moments

Getting the minimum moment (or other positive-definite function $\int p(x)f(x)\mathrm{d}x$) at a fixed entropy $H = -\int p(x)\log p(x)\mathrm{d}x$ is a constrained optimization problem. Use Lagrange multipliers:

$$\min_p \int p(x)f(x)\mathrm{d}x \quad \text{s.t.} \begin{cases} \int p(x)\mathrm{d}x = 1 \\ \int p(x)\log p(x)\mathrm{d}x = -H \end{cases}$$

$$\Rightarrow \nabla_{\lambda,\mu,p} \int p(x)\big[f(x) + \lambda + \mu \log p(x)\big]\mathrm{d}x = 0$$

$$\Rightarrow \nabla_{\lambda,\mu}\left(\min_p \int p(x)\big[f(x) + \lambda + \mu \log p(x)\big]\mathrm{d}x\right) = 0$$

Analytic solution takes the form

$$p(x) = \exp\big(-(f(x) + \lambda + \mu)/\mu\big)$$

where (λ, μ) are set to satisfy the constraints.

Bound on Moments

Procedure:

1. Find analytic form $p(x) = \exp\big(-(f(x) + \lambda + \mu)/\mu\big)$ ($f(x) = |x|$, x^2, etc.)
2. Set λ, μ to satisfy constraints (normalization, entropy)
3. Plug in $H_{\psi,\min} = \frac{1}{2}\log(4\pi e^{3/2}/N)$.

(L_1 norm)

- Functional form: exponential $p = (2a)^{-1}e^{-|x|/a}$
- Entropy is $\log(2ea)$
- Moment is given by

$$\langle |x| \rangle = a = e^{H-1/2}$$

$$\|\psi\|_1 > \sqrt{\frac{\pi}{e^{1/2}N}} \approx \frac{1.38}{\sqrt{N}}$$

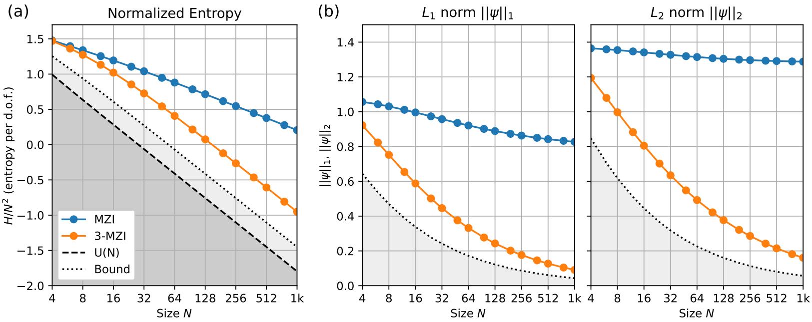
(L_2 norm)

- Functional form: Gaussian $p = (\sqrt{2\pi}\sigma)^{-1}e^{-x^2/2\sigma^2}$.
- Entropy is $\frac{1}{2}\log(2\pi e\sigma^2)$
- Moment is given by

$$\langle |x| \rangle = \sigma = e^{H-1/2}/\sqrt{2\pi}$$

$$\|\psi\|_1 > \sqrt{\frac{2e^{1/2}}{N}} \approx \frac{1.82}{\sqrt{N}}$$

Phase Moments vs. Information Bound



3-MZI comes within a factor of $2\times$ of the bound. MZI is way off.

$$\frac{\|\psi\|_1}{\|\psi\|_{1,\min}} \xrightarrow{N\to\infty} \frac{16e^{1/4}}{3\pi} \approx 2.20, \qquad \frac{\|\psi\|_2}{\|\psi\|_{2,\min}} \xrightarrow{N\to\infty} \sqrt{2e^{-1/2}\log(N/N_0'')}$$

Conclusions

The new contributions of this work are:

- Systematically studied the problem of phase-shifter economy in photonic architectures.
- Proposed a new MZI mesh architecture (3-MZI) that is more efficient than standard meshes, both in absolute terms and scaling.
- Derived lower bounds for the phase-shifter moments, answering the question "How much phase does programmable photonics need?", using information theory.
- Showed that the 3-MZI nearly saturates these theoretical bounds.

Future Steps

Hmmm? Definitely need to look at:

- Other mesh geometries. Preliminary results show 3-MZI nonunitary meshes can saturate the information bound with certain target matrix distributions.
- How close do non-mesh schemes (programmable MMI, diffractive network, MPLC) come to the bound?
- Generalization to nonlinearity?
- Neural network training subject to L_∞ bounds?

Acknowledgements:

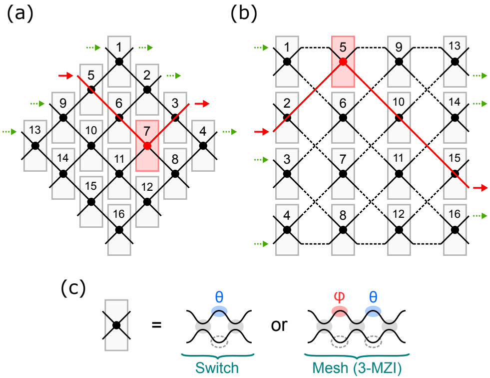
- Alex Sludds, Jasvith Basani, Saumil Bandyopadhyay, Sri Vadlamani, Dirk Englund

Funding: NTT Research, TSMC, NSF EAGER, DARPA NaPSAC

But wait, there's more!

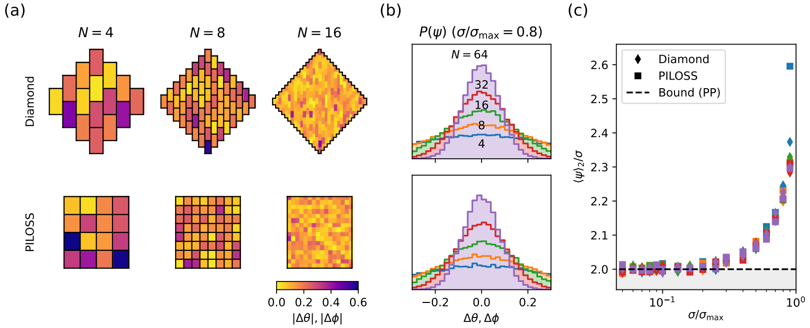
Phase-Efficient Non-Unitary Meshes

- Most optics literature focuses on unitaries.
- But applications are usually non-unitary
- Crossbar (diamond, PILOSS) can accommodate non-unitary meshes
- Can also be phase-efficient!



(a) Diamond and (b) PILOSS mesh. (c) Crossing type.

Phase-Efficient Non-Unitary Meshes

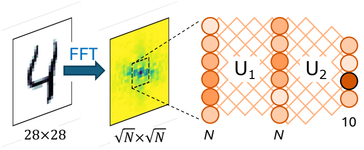


Non-unitary random matrices M (M_{ij} std. dev. σ) realized with 3-MZI diamond / PILOSS meshes:

- Marchenko-Pastur theory: $\sigma_{\max} \sim 1/\sqrt{N}$ for $|M| < 1$.
- Average phase shift decreases as $\langle\psi\rangle \sim 1/\sqrt{N}$
- Information-entropy bound: $\langle\psi\rangle = 2\sigma$. *3-MZI saturates the bound.*

Our architecture is not just near-optimal, but actually optimal.

L_∞-constrained Neural Network Training

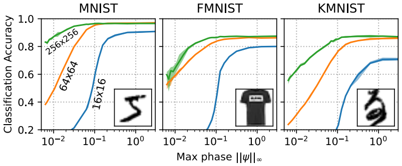


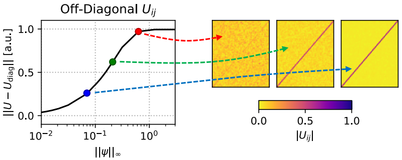
"Hard" phase bound $\psi < \|\psi\|_\infty$ is much more difficult than reducing average phase

- Mesh is no longer universal

But L_∞-constrained training works!

- 2-layer $N \times N$ Clements network
- MNIST, FMNIST, KMNIST
- High accuracy down to 0.1 rad
- Required phase decreases with N





Closing References

Hardware Error Correction

- S. Bandyopadhyay et al., *Optica* **8**(10), 1247 (2021)
- R. Hamerly et al., *PRApp* **18**, 024018 (2022)
- R. Hamerly et al., *PRApp* **18**, 024019 (2022)

New Mesh Architectures

- R. Hamerly et al., *Nat. Comm.* **13** (2022)
- J. Basani et al., *Nanoph.* **12**, 975 (2023)

Error-Aware Training

- S. Vadlamani et al., *Sci. Adv.* **9**, eadh3436 (2023)

Phase-Efficient Meshes

- R. Hamerly et al., arXiv:2408.09673.

Bibliography I

[1] M. Reck, *PRL* **73**, 58 (1994).

[2] W. Clements, *Optica* **3**(12), 1460 (2016).

[3] R. Tanomura, *JLT* **38**(1), 60 (2020).

[4] S. Fldzhyan, *Opt. Lett.* **45**, 2632 (2020).

[5] J. Blass, in IRE Int. Conv. Record **8**(1), 48 (1960).

[6] K. Suzuki, *Opt. Exp.* **22**(4), 3887 (2014).

[7] M. Fang, *Opt. Exp.* **27**, 14009 (2019).

[8] S. Bandyopadhyay et al., *Optica* **8**(10), 1247 (2021).

[9] R. Hamerly et al., *PRApp* **18**, 024019 (2022).

[10] R. Hamerly et al., *Nat. Comm.* **13** (2022).

[11] N. Russell, *NJP* **19**(3), 033007 (2017).

[12] B. Bell, *APL Phot.* **6**, 070804 (2021).

[13] K. Suzuki, *Opt. Exp.* **23**, 9086 (2015).

[14] R. Wu, *Opt. Lett.* **44**, 4698 (2019).

[15] M. Dong, *Nat. Phot.* **16**, 59 (2022).

[16] N. Harris, *Opt. Exp.* **22**, 10487 (2014).

[17] M. Takenaka, *JLT* **37**(5), 1474 (2019).

[18] K. Suzuki, *JLT* **37**, 116 (2018).

[19] L. Boya, Rpt. Math. Phys. **52**(3), 401 (2003).