

Delocalized Photonic Deep Learning On The Internet's Edge

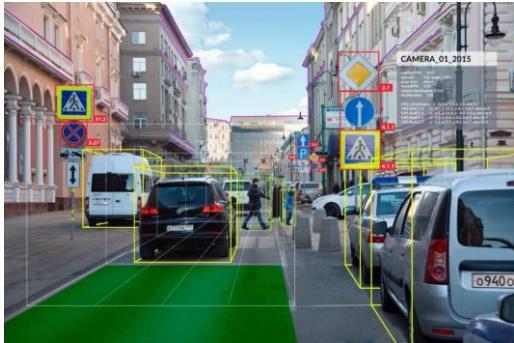
OptSys Seminar
Alexander Sludds

Outline

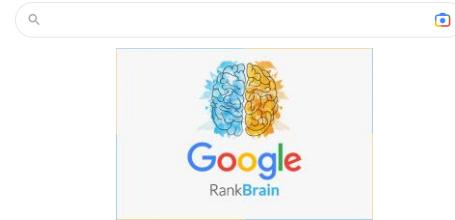
- Broad Introduction
- Delocalized Photonic Deep Learning On The Internet's Edge (Netcast)
- Discussion Of Netcast (Noise, Scaling And Crosstalk)
- Conclusion

Machine Learning Is Ubiquitous

Machine Vision



Search



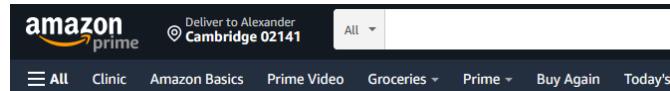
Natural Language Processing



Game Playing (Go, Starcraft, etc)



E-Commerce



ChatGPT

Write python code in qcodes for controlling a silicon photonic Mach-Zehnder Interferometer to perform machine learning

Here is a sample code in python using qcodes for controlling a silicon photonic Mach-Zehnder Interferometer (MZI) for machine learning:

```
python
import qcodes as qc
from qcodes.instrument_drivers.stanford_research.SR860 import SR860
from qcodes.instrument_drivers.rohde_schwarz.SGS100A import SGS100A
from qcodes.instrument_drivers.tektronix.AWG5814 import AWG5814
import numpy as np
from sklearn.svm import SVC
from sklearn.model_selection import train_test_split

# Connect to instruments
lockin = SR860("lockin", "GPIB::8")
signal_generator = SGS100A("signal_generator", "GPIB::13")
awg = AWG5814("awg", "TCPIP0::192.168.0.14::inst0::INSTR")
```

Create a poem about Alex Sludds' thesis defense on "Delocalized Photonic Deep Learning on the Internet's Edge"

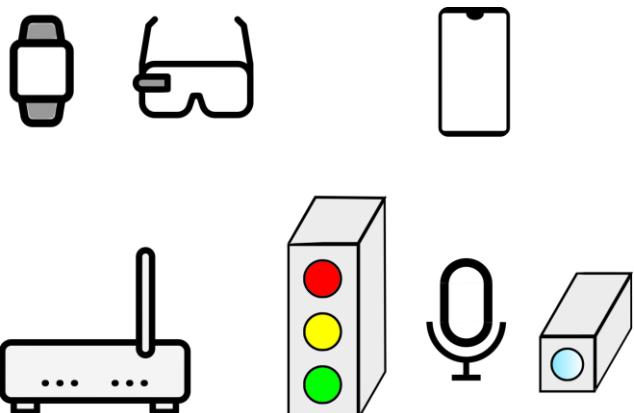
A thesis defense, a moment of pride
For Alex Sludds, the day has arrived
"Delocalized Photonic Deep Learning on the Internet's Edge"
A topic both complex and wide

With data that flows like rivers deep
And neural networks that never sleep
The edge of the internet, a place of might
Where Alex's research takes flight

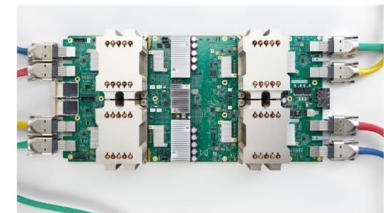
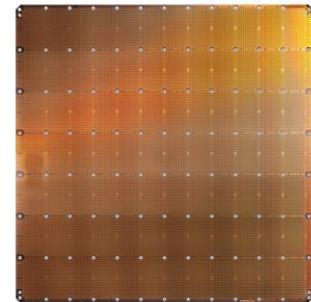
Photons dance and neurons fire
In a symphony of data desire
Delocalized, yet connected still
A network that will always thrill

How We Access ML Today: Edge Computing

Our data from milliwatt class sensor devices...



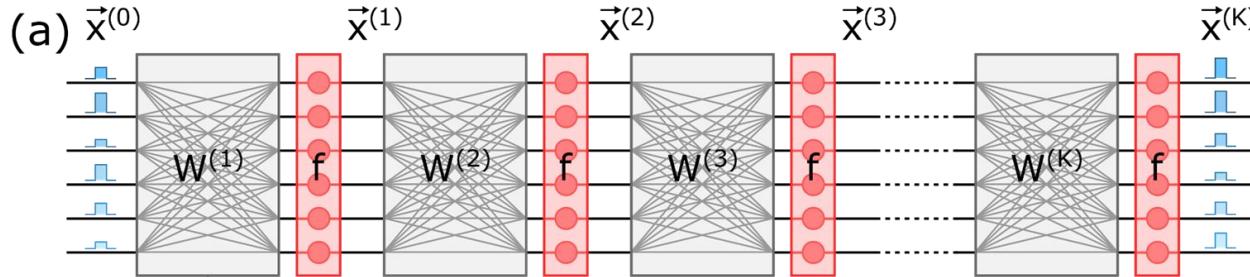
... is run on kilowatt class computers



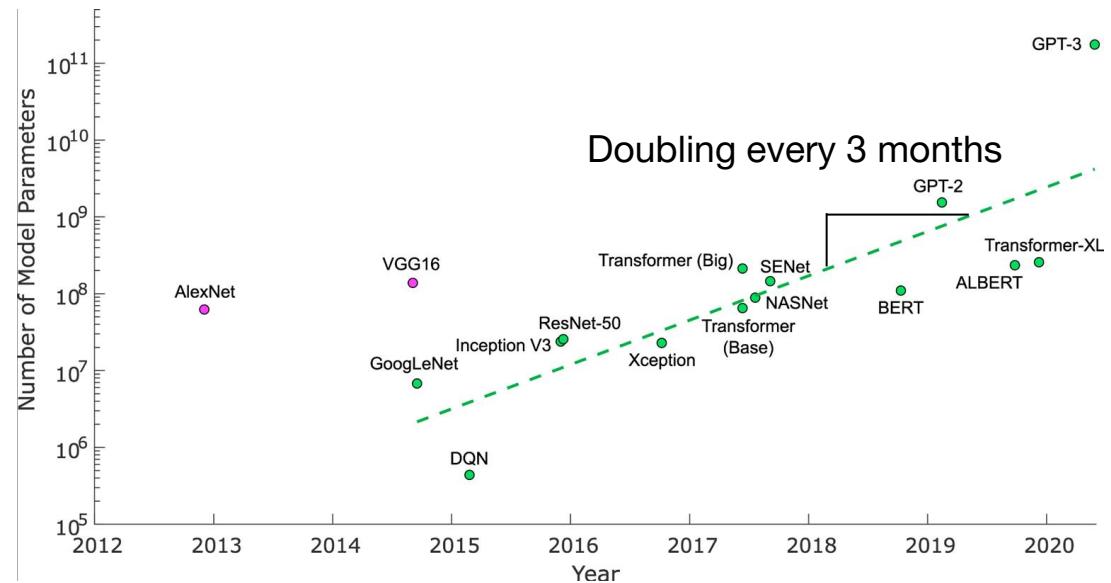
State of the art
Edge AI processor



Running A Neural Network



Models Are Growing Exponentially

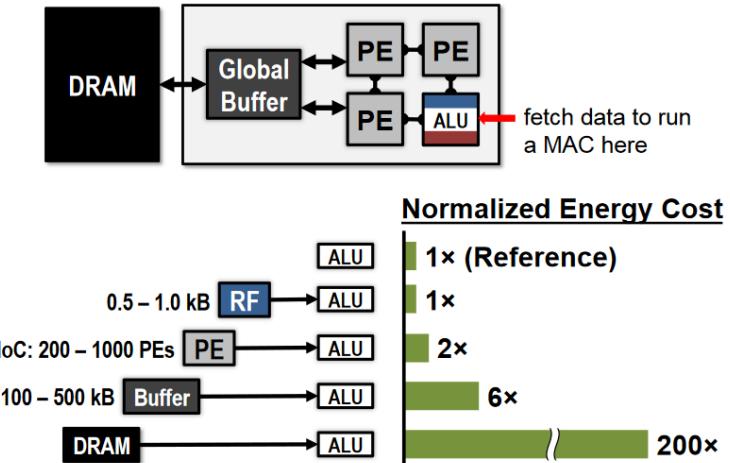
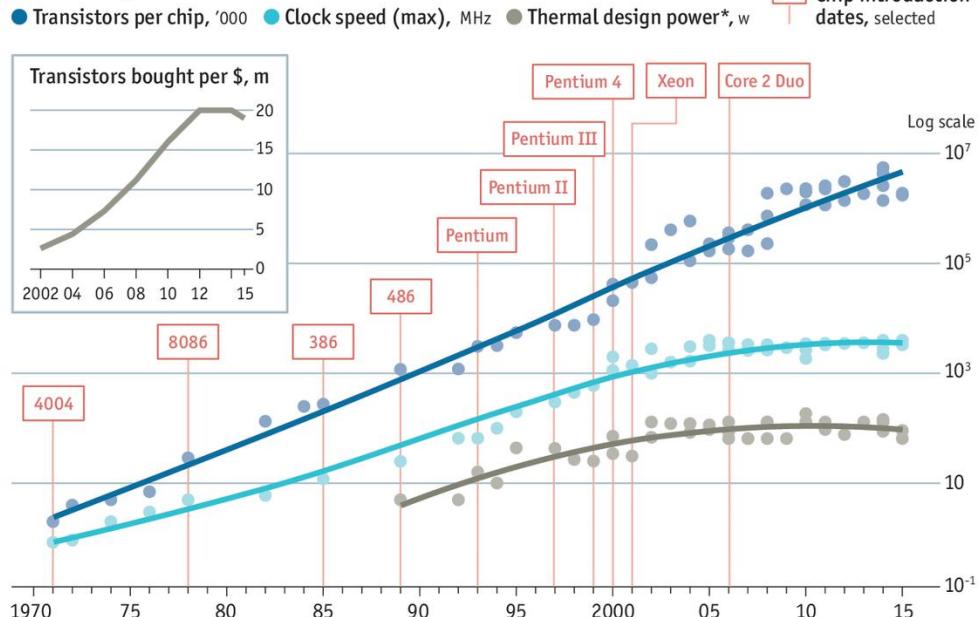


“From talking to OpenAI, GPT-4 will be about 100 trillion parameters...”

-Andrew Feldmann, CEO of Cerebras

What's Wrong With Digital Computing?

Stuttering



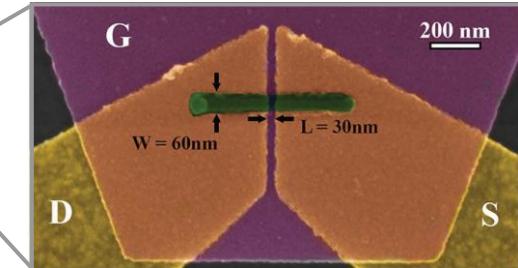
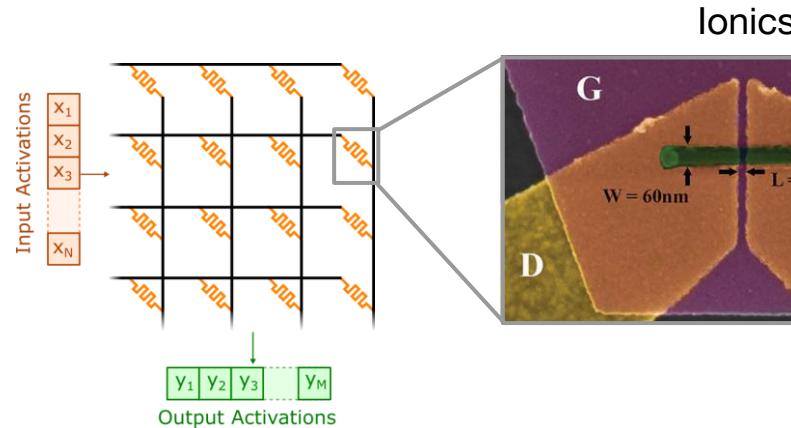
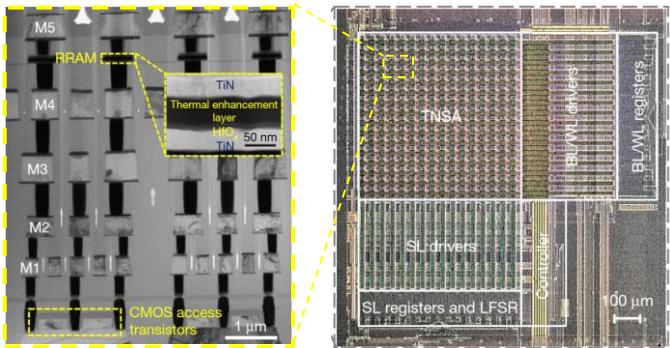
Digital CMOS is limited by data movement, not computation! "Von Neumann" Bottleneck

The Economist "After Moore's Law" 2019

V. Sze: Efficient Processing of Deep Neural Networks
Proceedings of IEEE, 2017

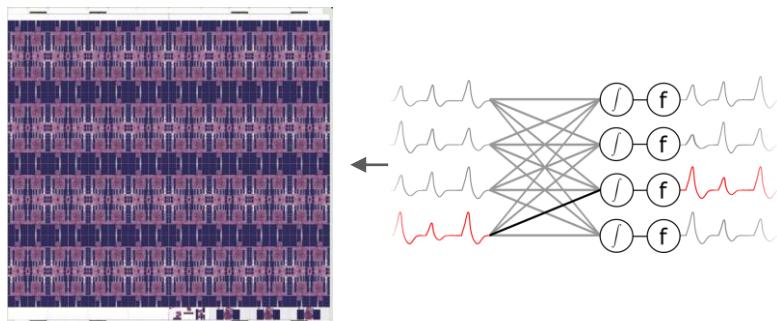
New Computing Approaches

Memristors

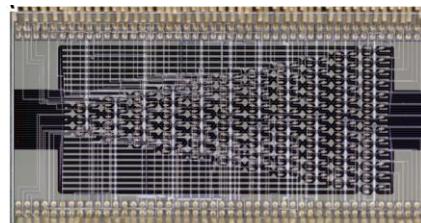


M. Onen et al, Science 2022

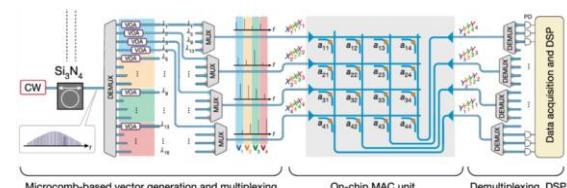
Spiking/Neuromorphic



Photonics



Y. Shen, N. Harris et al, Nature Photonics 2017



J. Feldmann et al, Nature 2021

M Davies, Proceedings of the IEEE, 2021

Why Use Photonics?

	Electronics	Photonics	
Nonlinearity	Easy (transistors)	Hard [with caveats]	Optical computing is hard Reason: electrical NL strong, optical NL weak
Memory	Easy (DRAM, flip-flop)	Hard	
Gain	Easy	OK	
Communication	$(1/2)CV^2$ energy cost	Easy (waveguides)	Data movement is easy in photonics
Fan-out	$(1/2)CV^2$ energy cost	Easy: beamsplitters, multiplexers	
Linear algebra	Hard (BLAS, gemv/gemm)	Easy (at low precision of ~10 bits)	

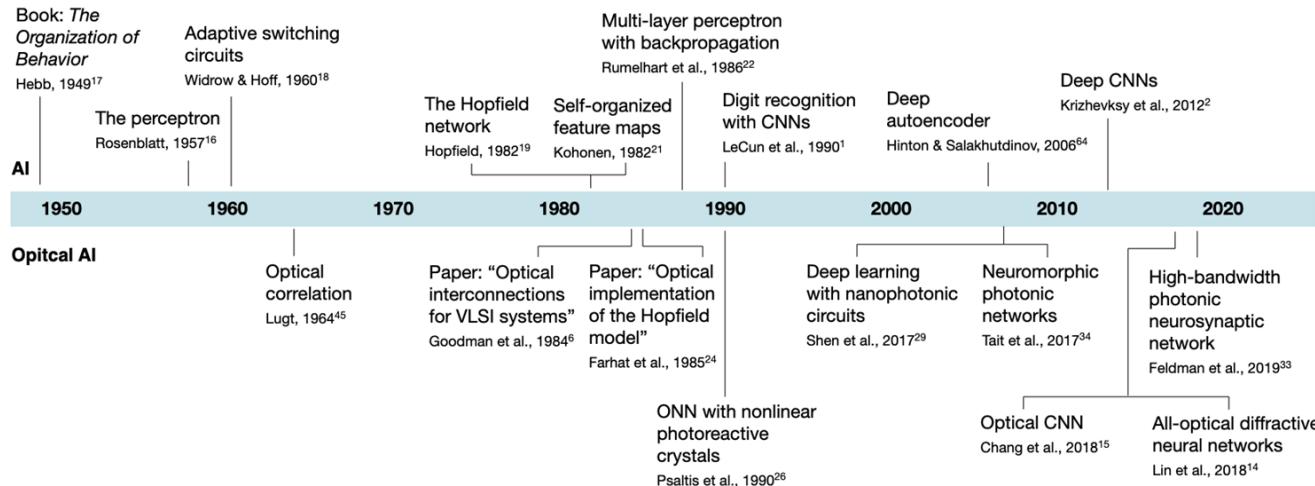
Optical Computing Is Not New, Why Now?

Inference in artificial intelligence with deep optics and photonics

Gordon Wetzstein , Aydogan Ozcan, Sylvain Gigan, Shanhui Fan, Dirk Englund, Marin Soljačić,

Cornelia Denz, David A. B. Miller & Demetri Psaltis

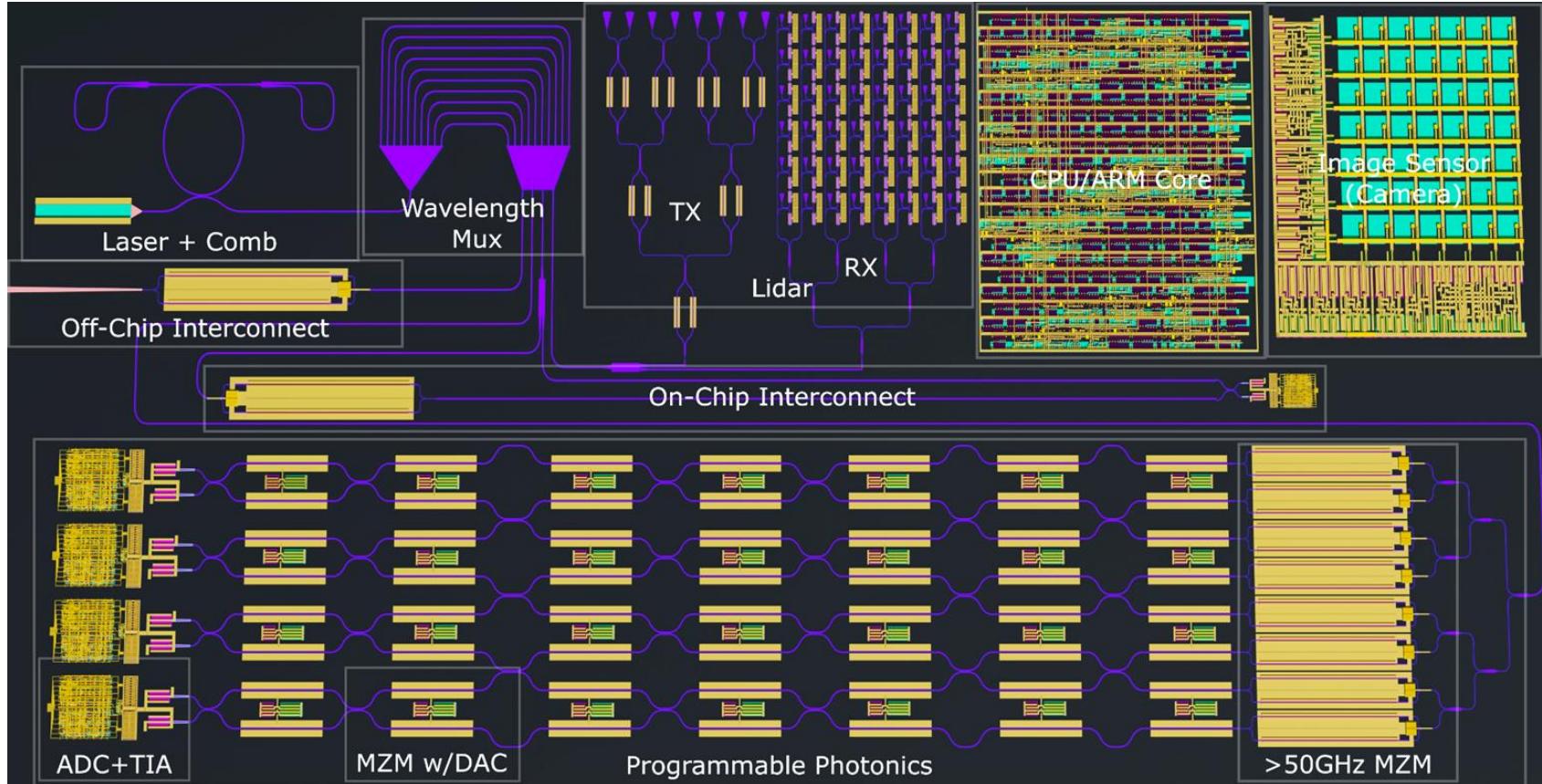
Nature 588, 39–47 (2020) | [Cite this article](#)



Why Now?

- Enough data
- Enough CPU
- Photonic ICs
- We know NNs work
- .. von Neumann architectures probably not optimal for ANNs

One Approach: Integrated Photonics

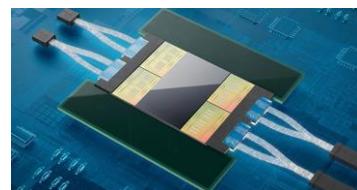
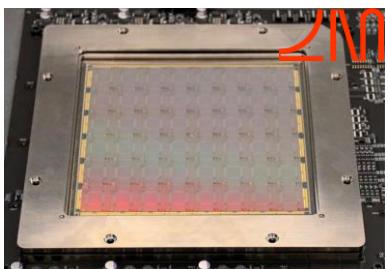
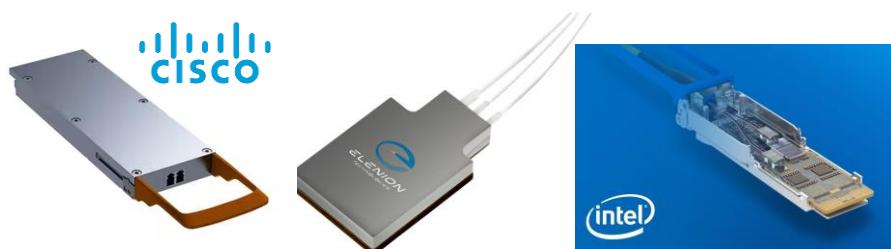


Silicon Photonics Revolution

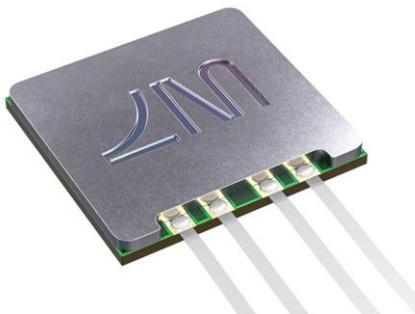
Communications

Cisco Agrees to New Deal to Buy Acacia for \$115 a Share

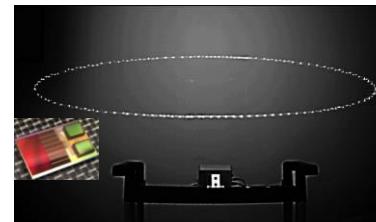
Networking-equipment giant boosts purchase price by 64% in deal that values Acacia at about \$4.5 billion



Computing



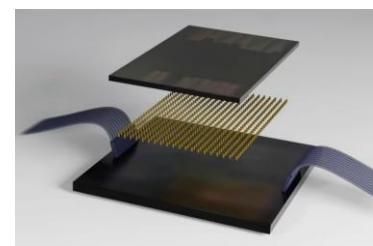
LiDAR



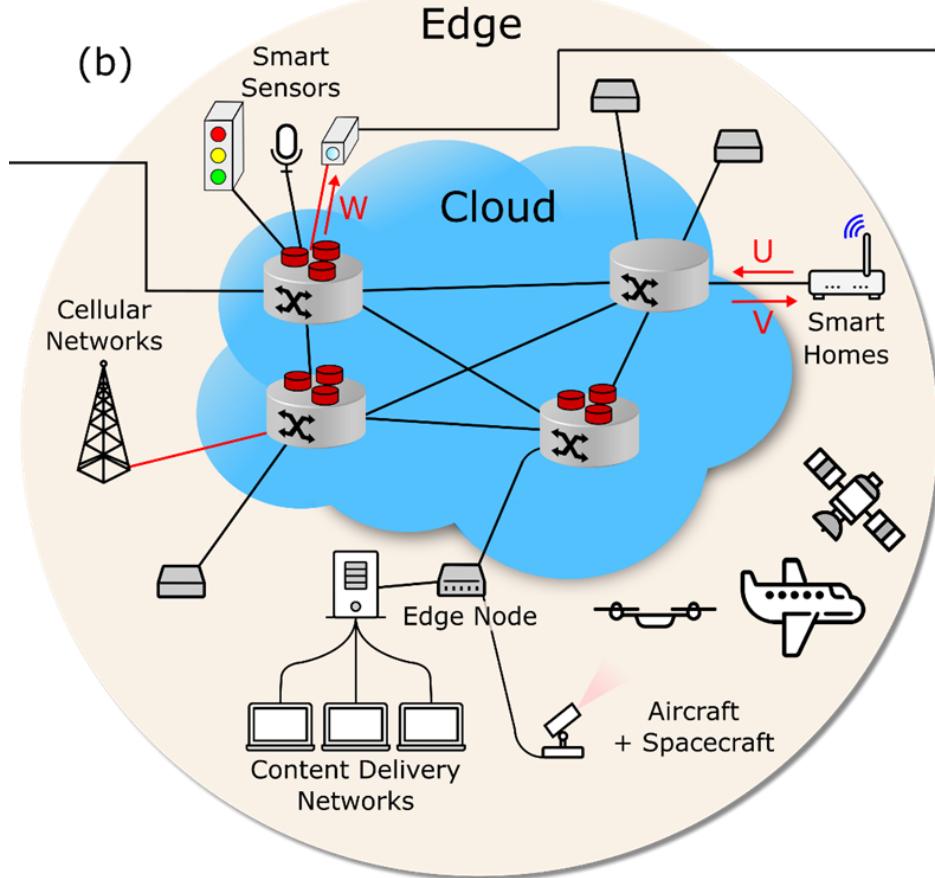
Sensing



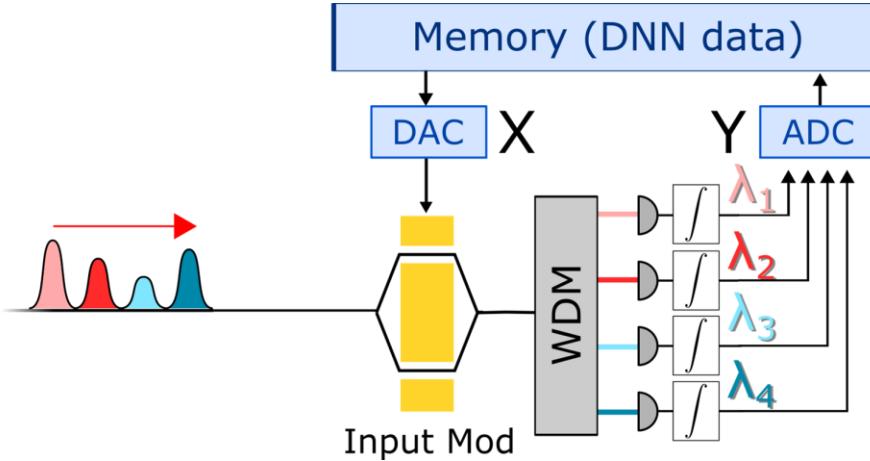
Quantum Computing



Optically Accelerated Edge Computing



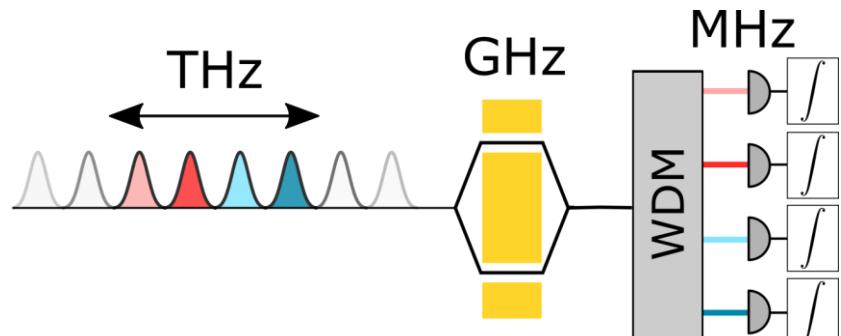
Client Energy and Bandwidth



Netcast Client Energy Consumption

Device	Number of Devices	Fan-out	Energy per Device	Energy per MAC
Modulator	1	N	$\sim 1 \text{ pJ}$	$\sim (1/N) \text{ pJ}$
DAC	1	N	$\sim 1 \text{ pJ}$	$\sim (1/N) \text{ pJ}$
ADC	1	M	$\sim 1 \text{ pJ}$	$\sim (1/M) \text{ pJ}$
Integrator	N	M	$\sim 1 \text{ fJ}$	$\sim (1/M) \text{ fJ}$
Total	—	—	—	$\sim (1/N) \text{ pJ}$

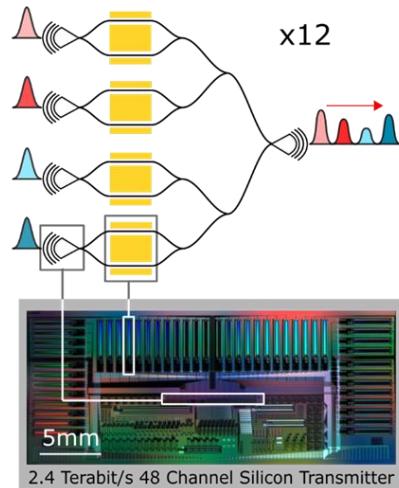
10fJ/MAC possible in near-term deployed systems



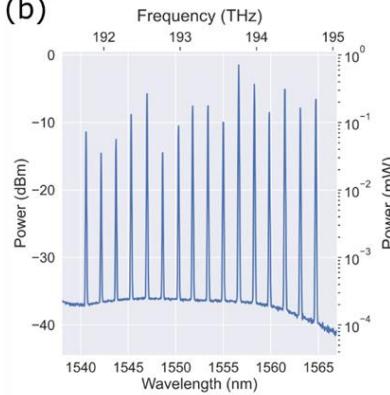
TeraMAC/s possible on client with single modulator

Experiment

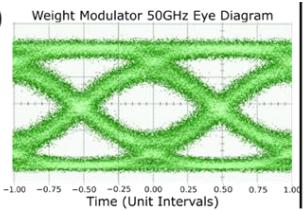
(a) Smart Transceiver



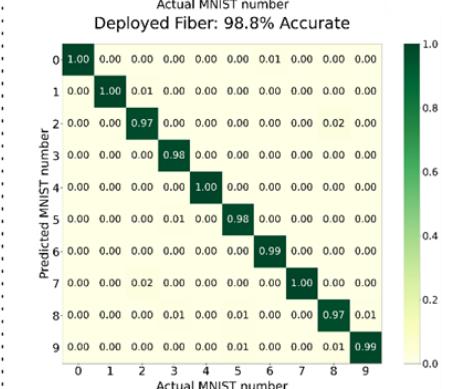
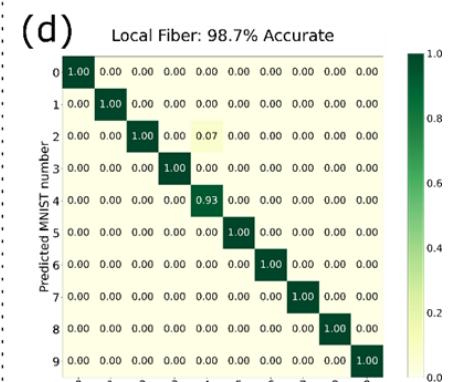
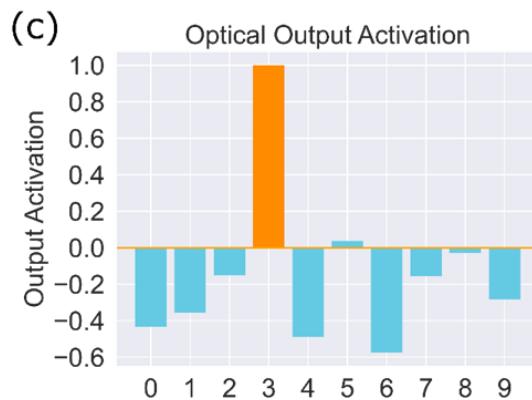
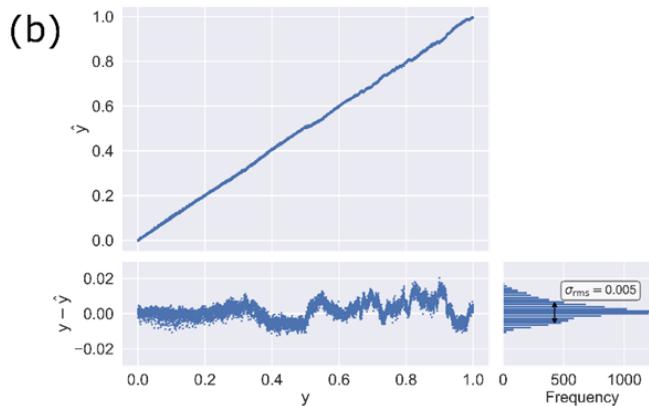
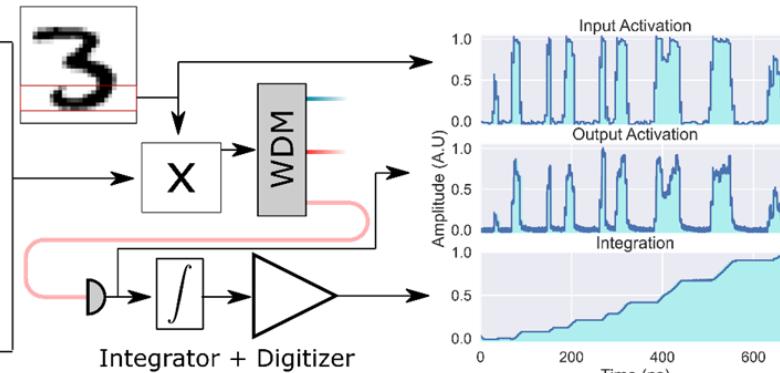
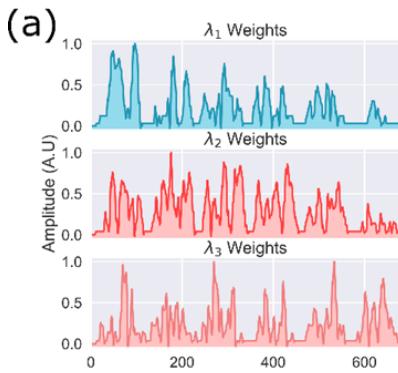
(b)



(c)

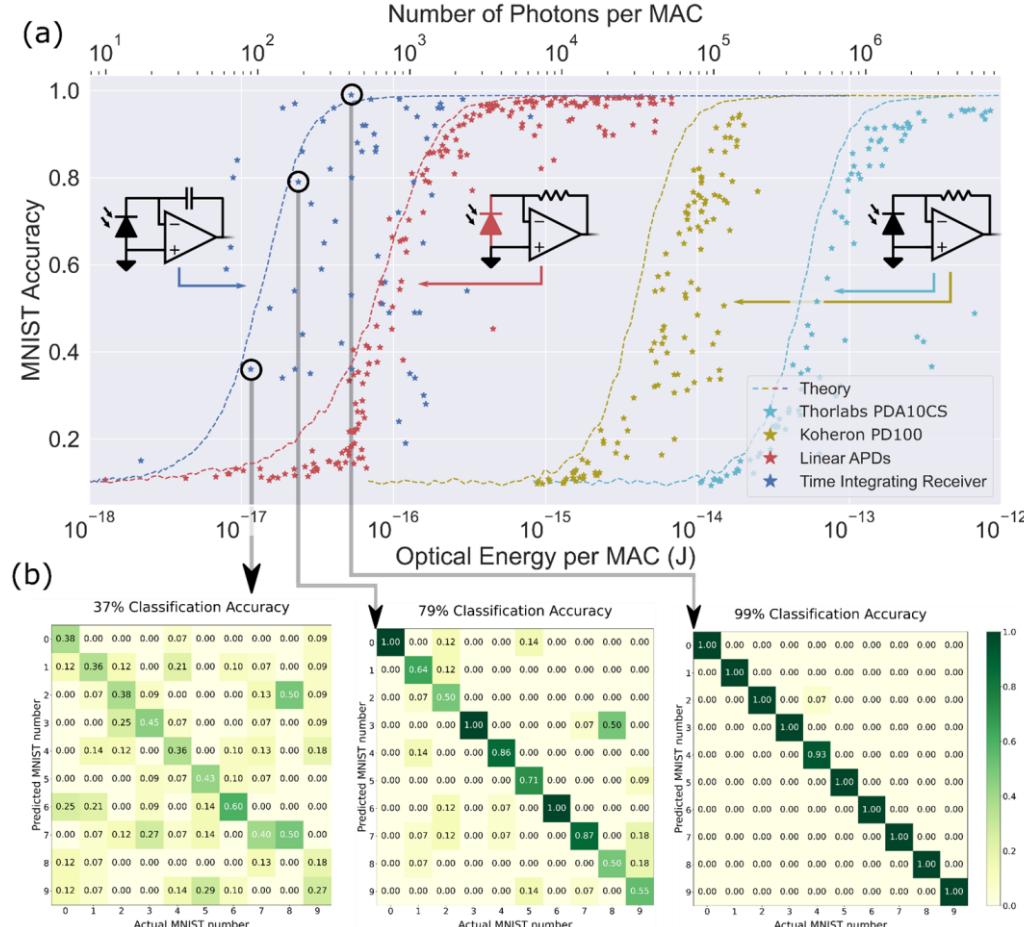


Dataflow And Accuracy



Accurate computation with 86 km deployed fiber with 3 THz bandwidth

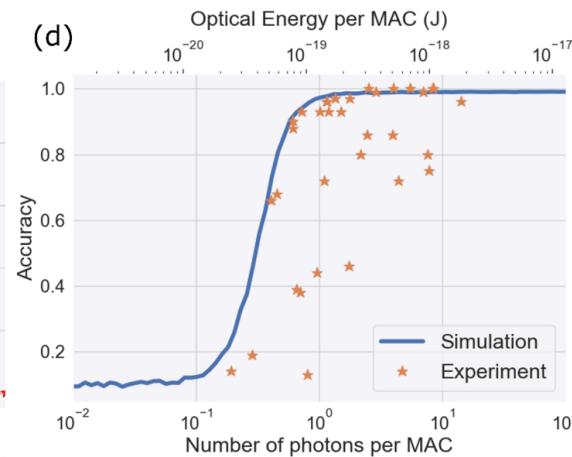
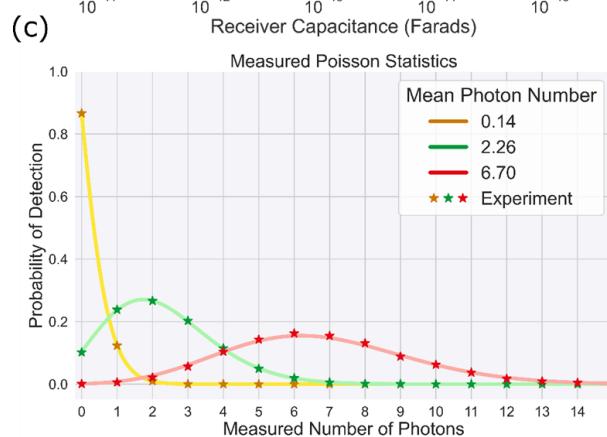
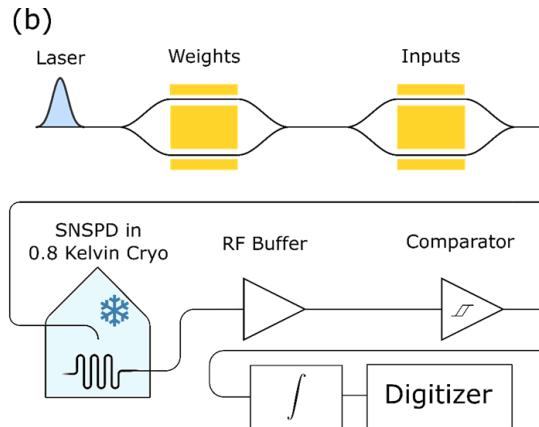
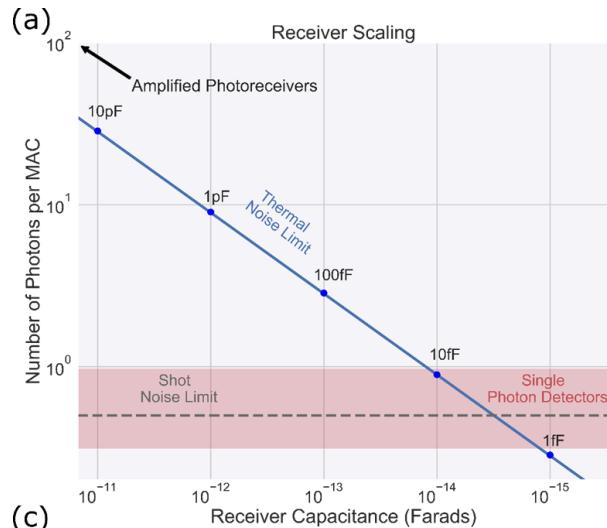
Time Integration Improves Sensitivity



Fundamental Limits To Sensitivity

Thermal noise scales as \sqrt{kTC}

Receiver capacitance



See also:

R. Hamerly, et al, PRX 2019

T. Wang et al, Nat Comms 2022

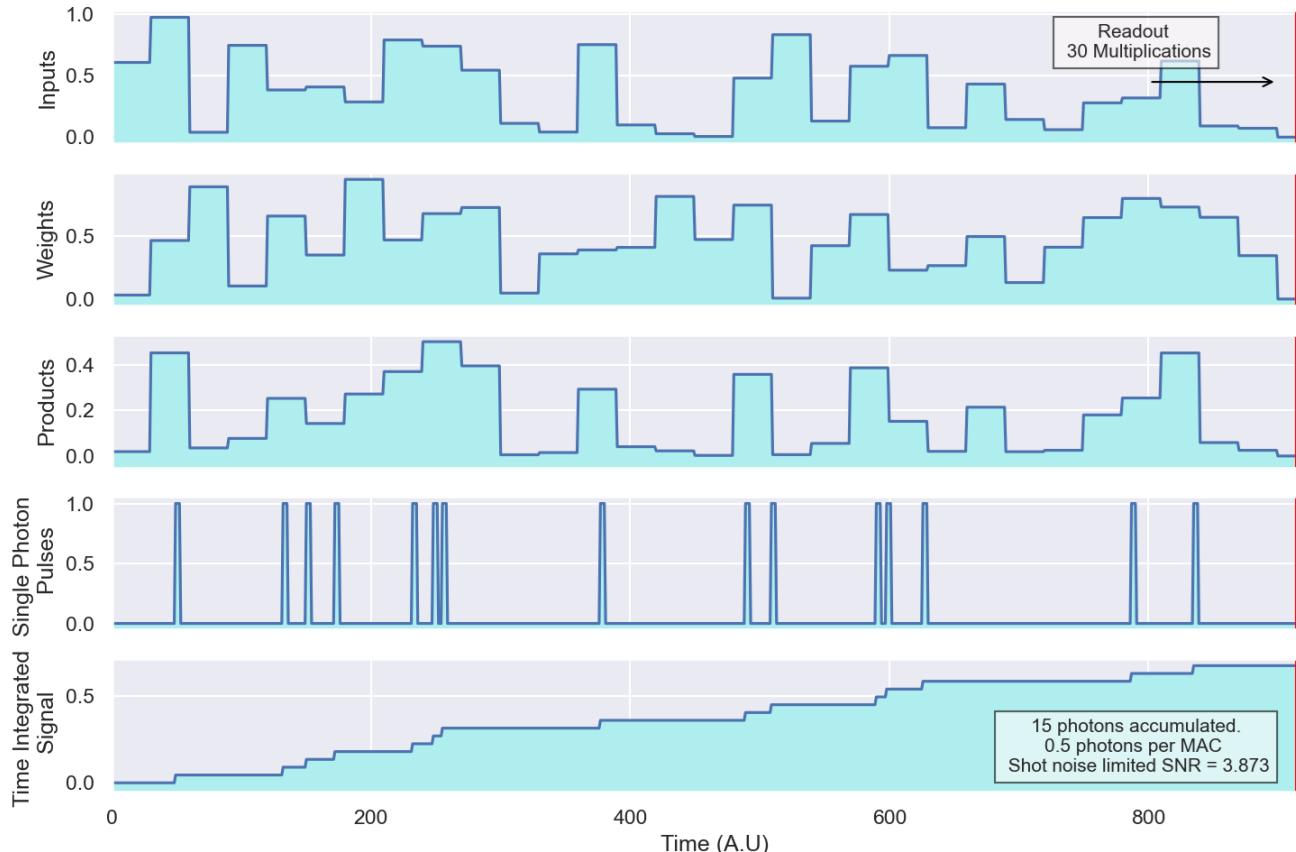
Introduction

Netcast

Discussion

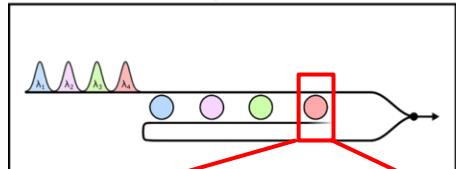
Conclusion

What Does <1 Photon Per MAC Mean?

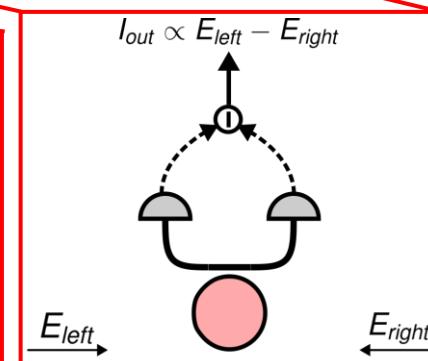
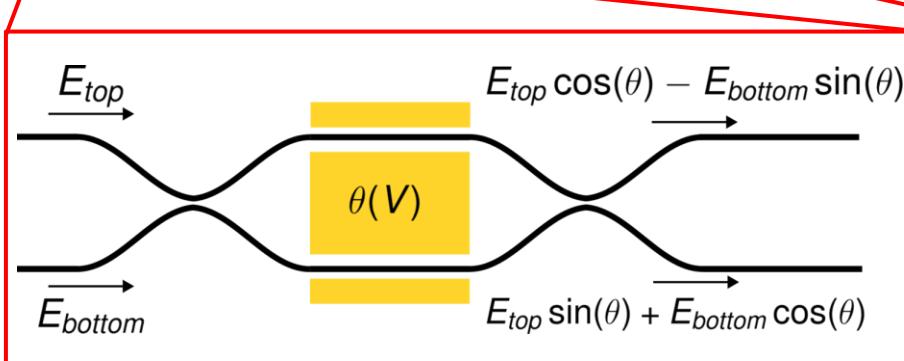
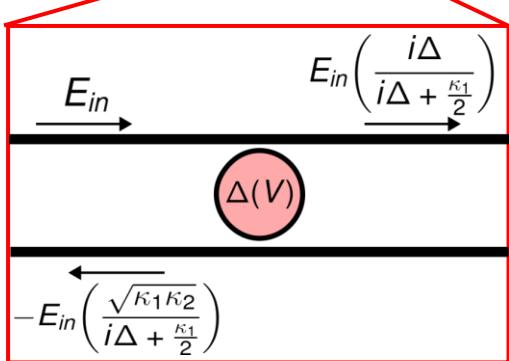
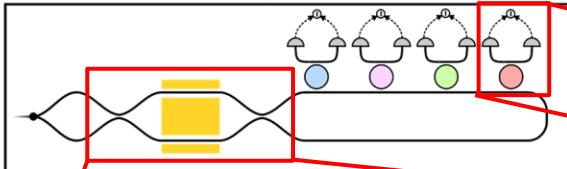


Discussion: Noise

Simple TX



Simple RX



$$I_{out} = \left[\left(\left| \frac{i\Delta}{i\Delta + \frac{\kappa_1}{2}} \right|^2 - \left| \frac{-\sqrt{\kappa_1\kappa_2}}{i\Delta + \frac{\kappa_1}{2}} \right|^2 \right) \cos(2\theta) \right] |E_{in}|^2$$

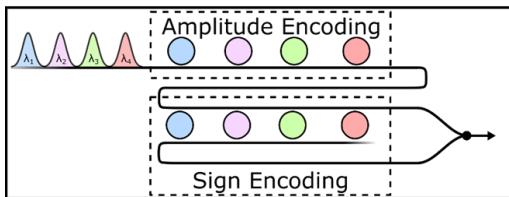
Weights Inputs

Problem: Differential encoding leads to small photocurrent for large number of input photons. Shot noise large. “Fat zero” problem.

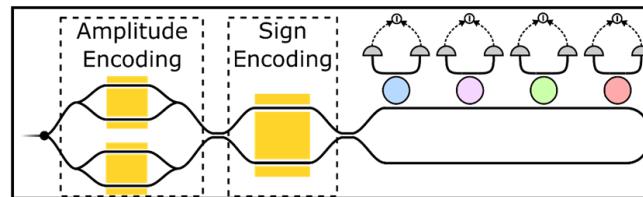
Discussion: Noise

Solution: Separate amplitude and sign encoding.

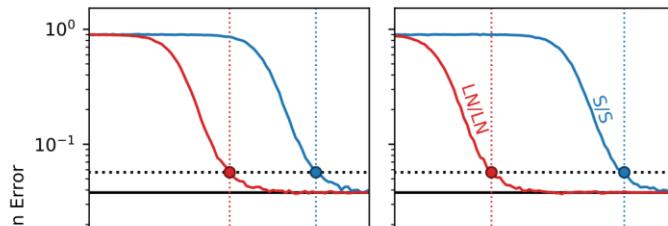
Low-Noise TX



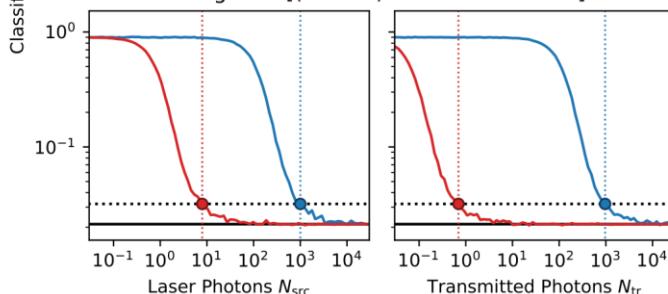
Low-Noise RX



Small NN [(28×28)→100→100→10]



Large NN [(28×28)→1000→1000→10]

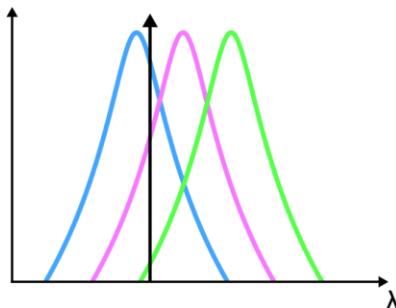


3 orders of magnitude
receiver sensitivity
improvement over naive
implementation

Discussion: Scaling And Crosstalk

Problem 1: Spectral Crosstalk

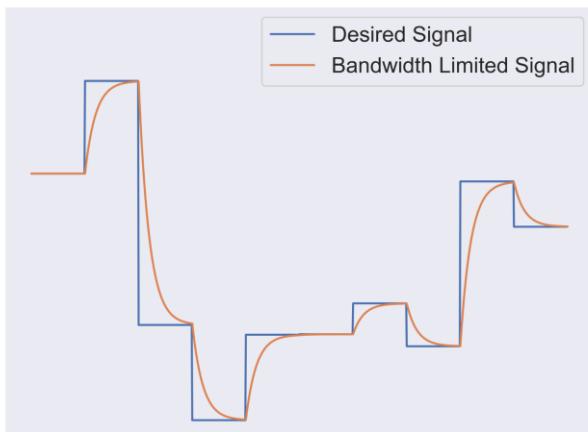
Cavities accept light from neighbors



$$\chi_\omega = \frac{\left(\frac{1}{2}\kappa\right)^2}{\Delta\omega^2 + \left(\frac{1}{2}\kappa\right)^2} \Rightarrow \Delta\omega \geq \frac{\kappa}{2\sqrt{\chi_\omega}}$$

Problem 2: Temporal Crosstalk

Data on sequential timesteps merge from finite bandwidth (photon lifetime + RC)

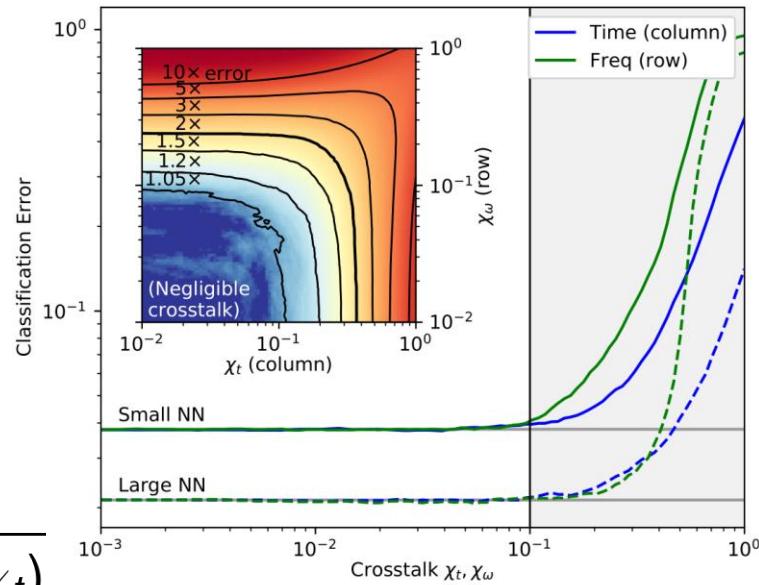


$$\chi_t = e^{-\frac{t}{\tau}}$$

$$\tau = \sqrt{\frac{1}{\kappa^2} + (\tau_{RC})^2} \approx \frac{\sqrt{2}}{\kappa}$$

Solving:

$$R = \frac{1}{T} \leq \frac{\kappa}{\sqrt{2} \log(1/\chi_t)}$$



Discussion: Scaling And Crosstalk

Temporal Crosstalk limits us to a symbol rate of:

$$R = \frac{1}{T} \leq \frac{\kappa}{\sqrt{2 \log(1/\chi_t)}}$$

Spectral Crosstalk limits us to a cavity spacing of:

$$\Delta\omega \geq \frac{\kappa}{2\sqrt{\chi_\omega}}$$

Assuming $\chi_t = \chi_\omega = \chi$
channel capacity is:

our $C = R \frac{2\pi B}{\Delta\omega} \leq \frac{2\pi\sqrt{2\chi}}{\log(1/\chi)} B$

Across the telecom C band (1530nm-1565nm, B=4.4 THz):

Crosstalk χ	Capacity C (C-band)	$\times 8$ b/wt
0.1	5.3 Twt/s	43 Tbps
0.05	2.9 Twt/s	23 Tbps
0.01	850 Gwt/s	6.8 Tbps
0.005	520 Gwt/s	4.2 Tbps
0.001	180 Gwt/s	1.2 Tbps

Netcast + Discussion Summary

Demonstrated photonic edge computing

Using a 2.4 Tbps silicon smart transceiver

Made in a commercial CMOS foundry

8-bit accurate computing

Deployment over 86 km of real fiber

Across 3 THz of bandwidth

Time integrating receivers at ~10aJ/MAC

Sensitivity of <1 photon per MAC with modern technology at room temperature

Enabling TeraMAC/s computing on edge devices with mW power consumption

Acknowledgements



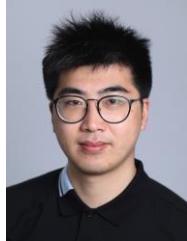
Prof. Dirk Englund



Dr. Ryan Hamerly



Dr. Saumil Bandyopadhyay



Prof. Zaijun Chen



Liane Bernstein



NOKIA



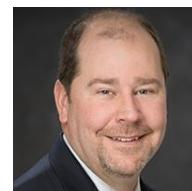
Dr. Tom Baehr-Jones



Dr. Michael Hochberg



Dr. Ben P Dixon



Dr. Scott Hamilton



Dr. Nicholas Harris

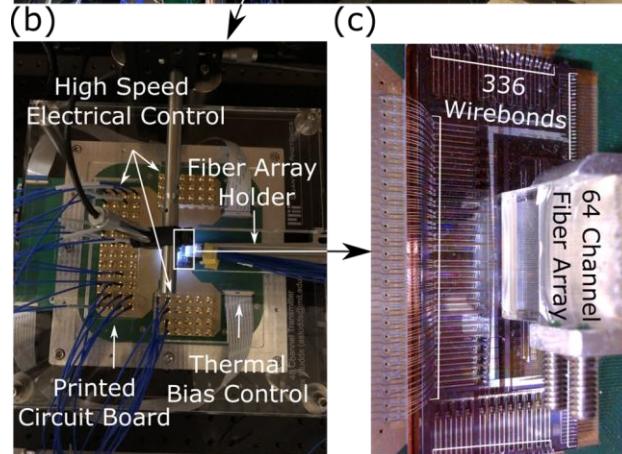
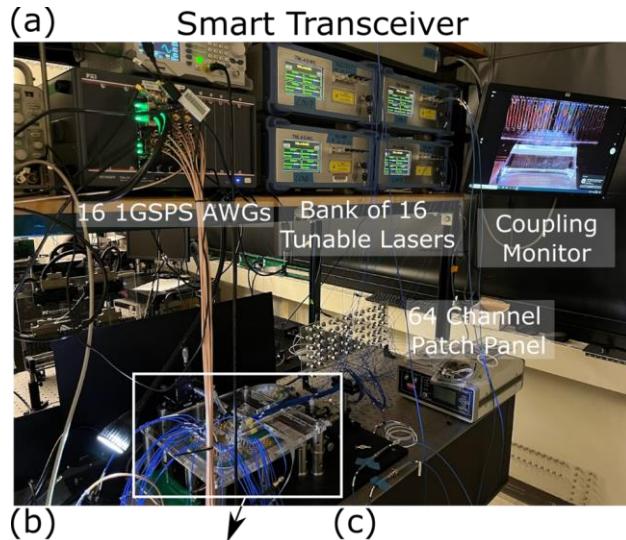


Dr. Darius Banundar

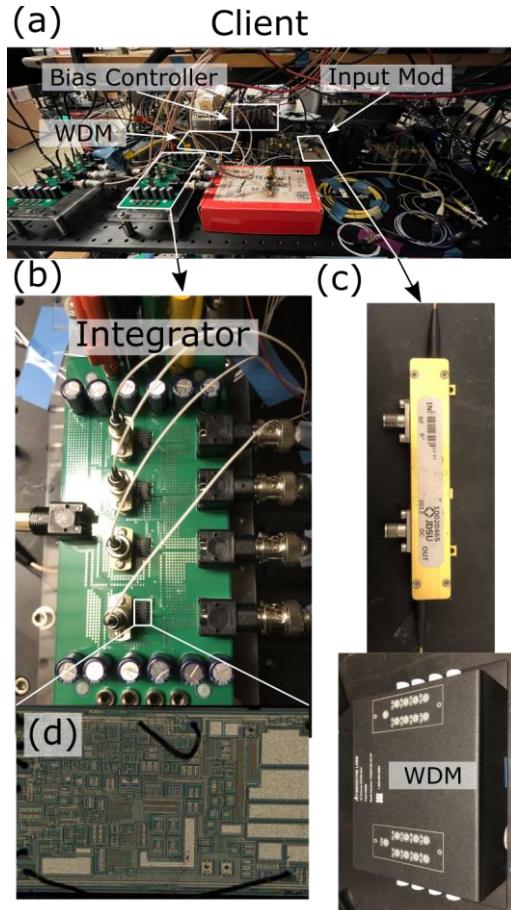
Questions?

Slides Past This Point Are Supplemental

Smart Transceiver Experimental Setup

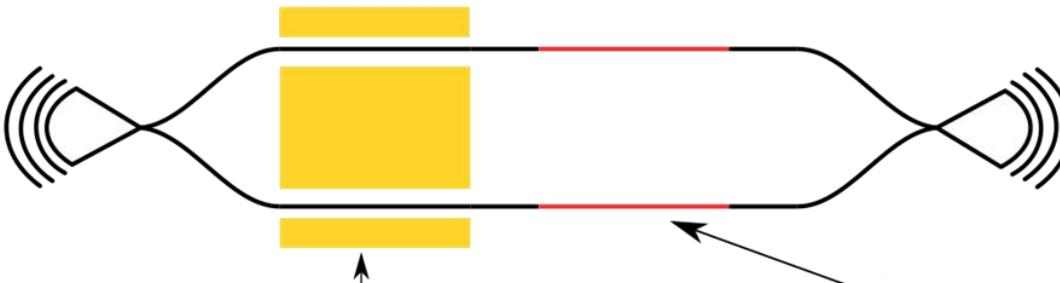


Client Experimental Setup

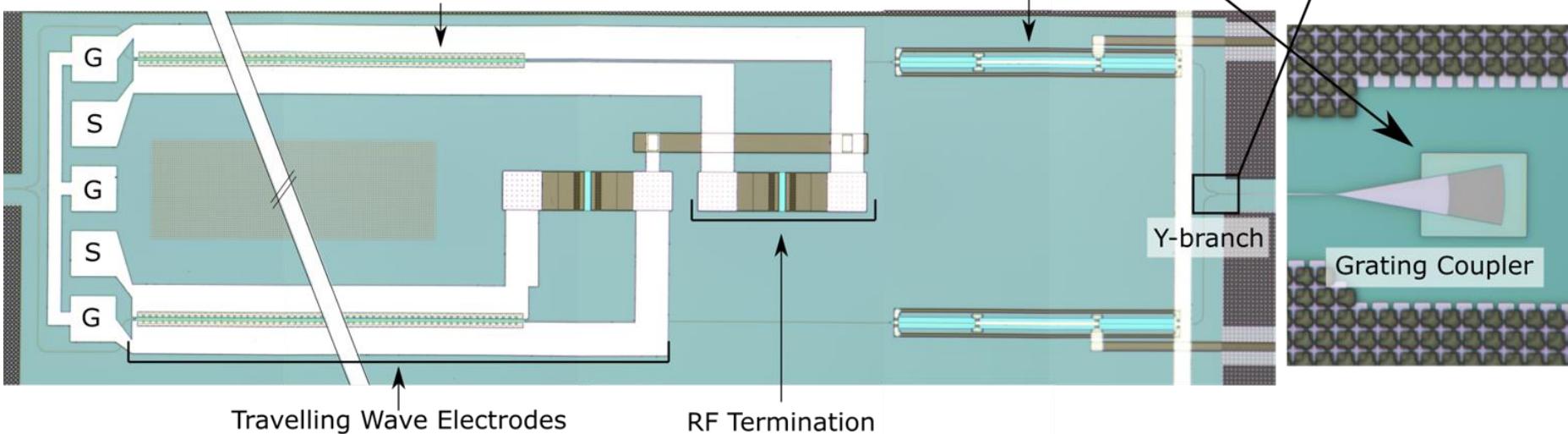


Components In Smart Transceiver

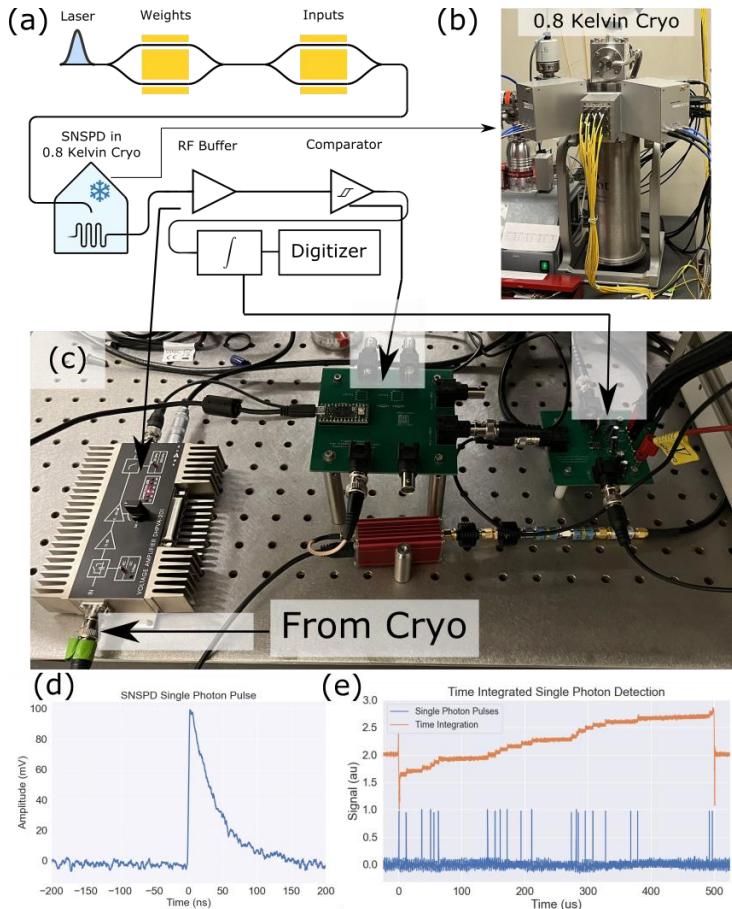
(a)



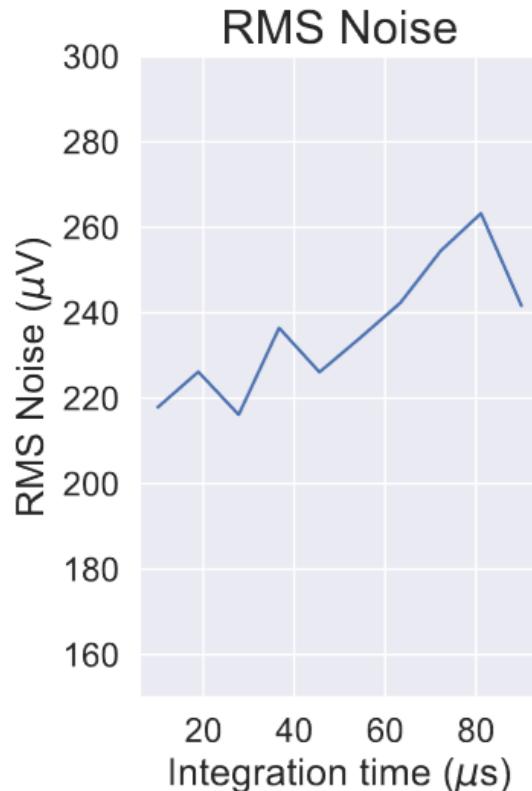
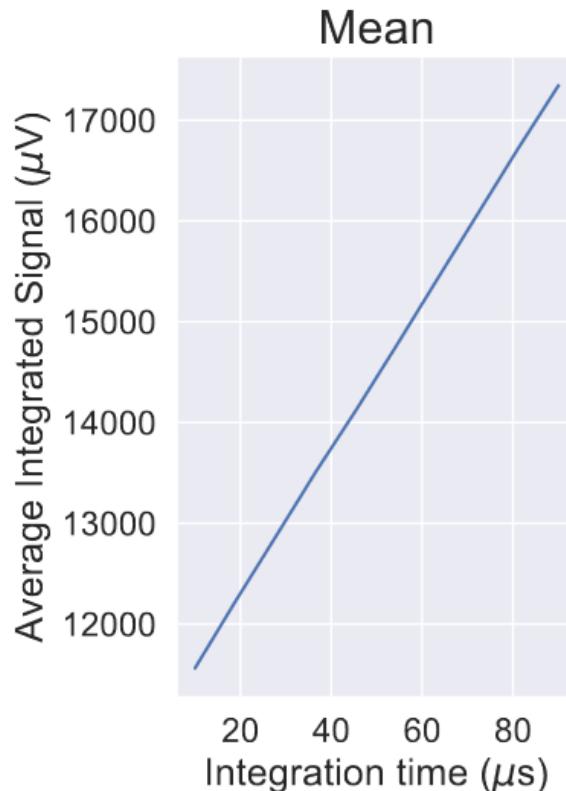
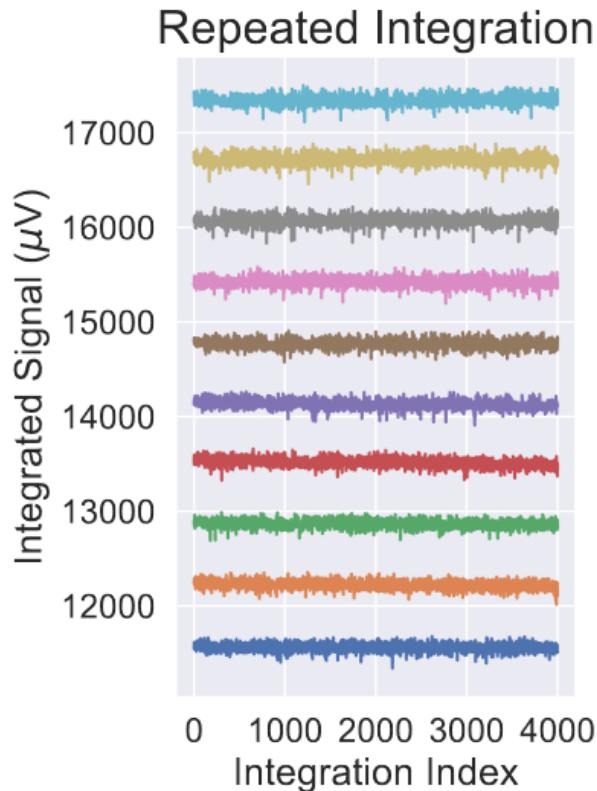
(b)



SNSPD Experimental Setup



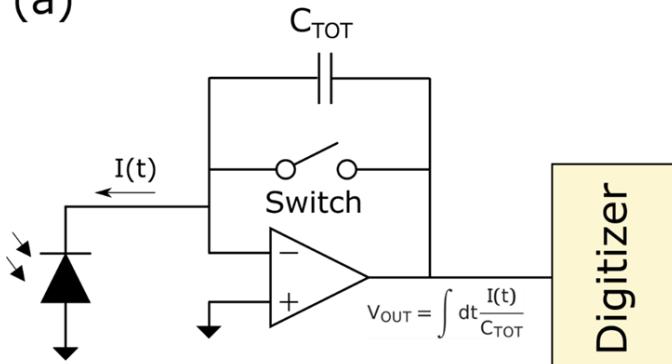
Time Integrator Noise



Time Integrating Receiver Explained

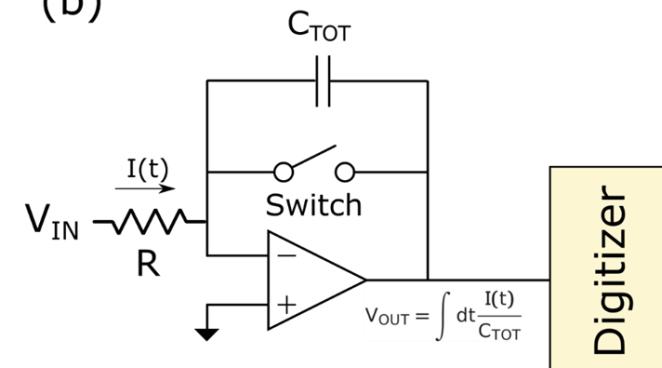
Current Integrator

(a)



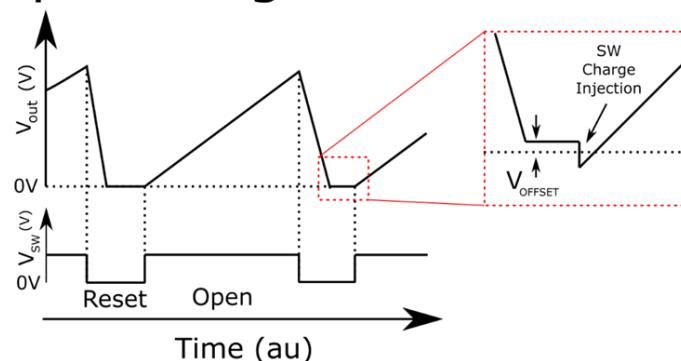
Voltage Integrator

(b)

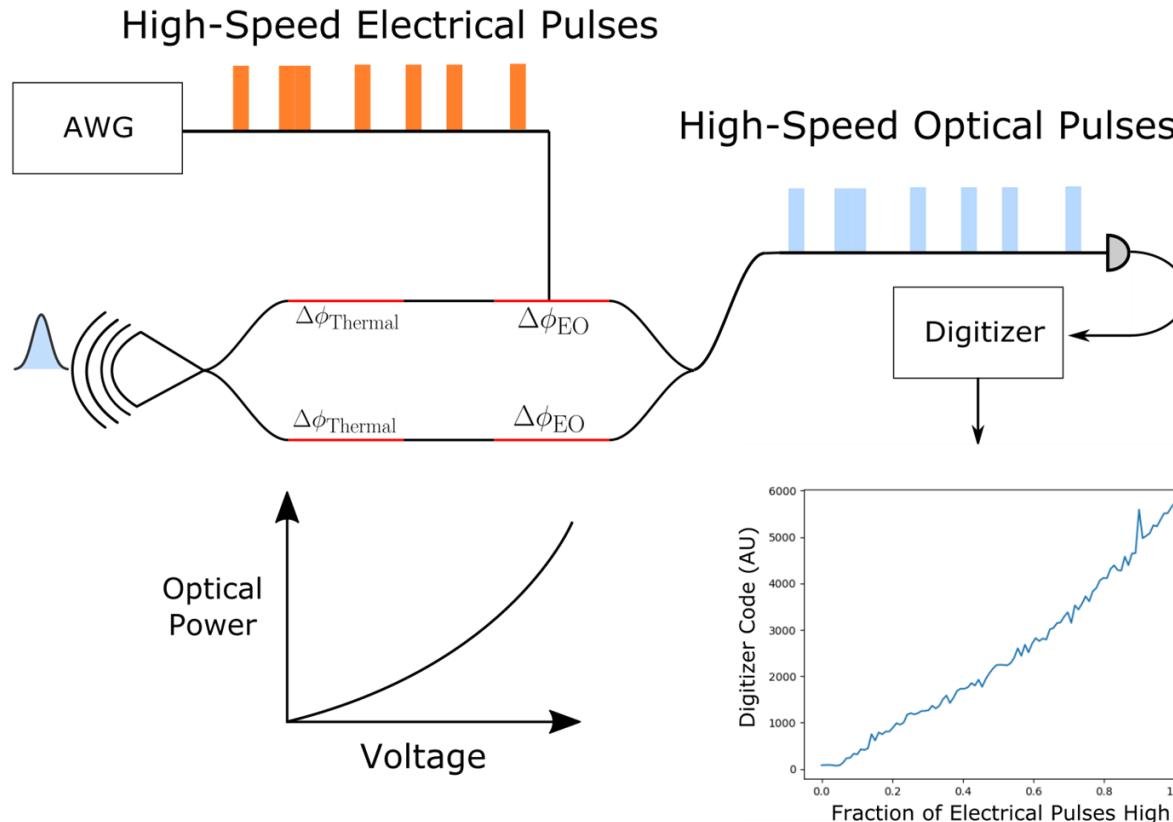


Example Integrated Waveform

(c)



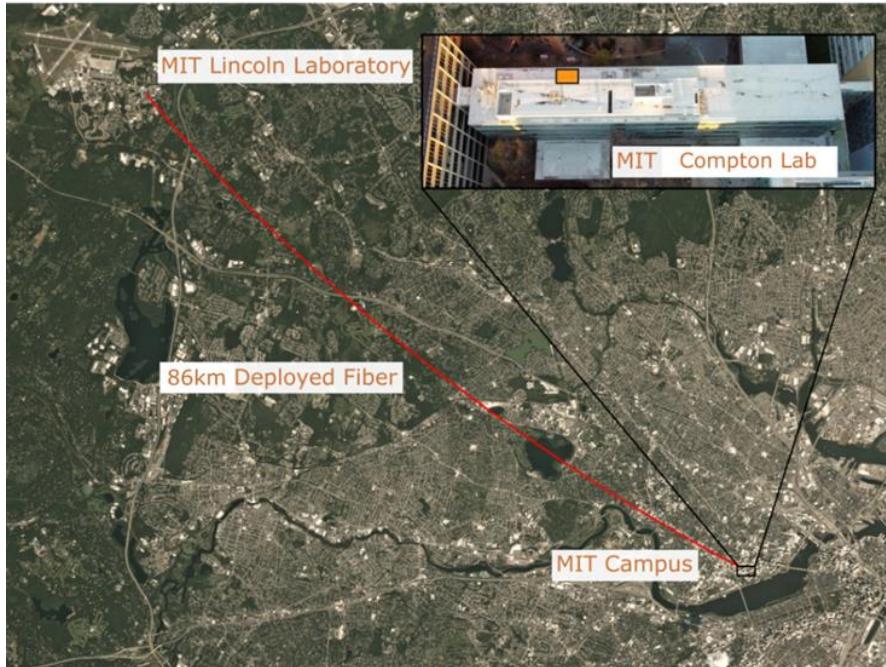
Time Integrator Bandwidth Measurement



Deployed Fiber OTDR Measurement

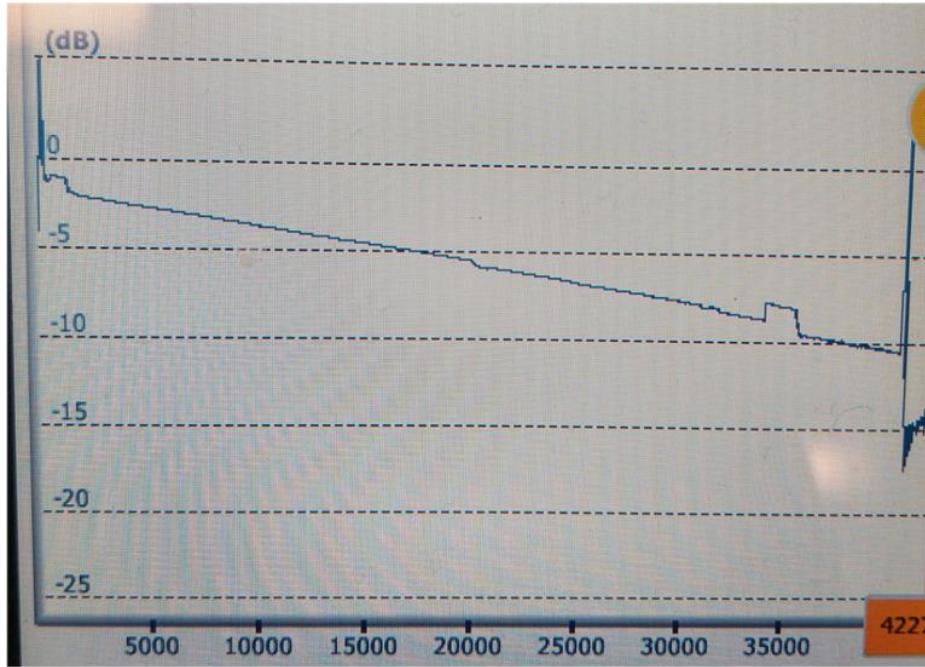
(a)

86km Deployed Fiber

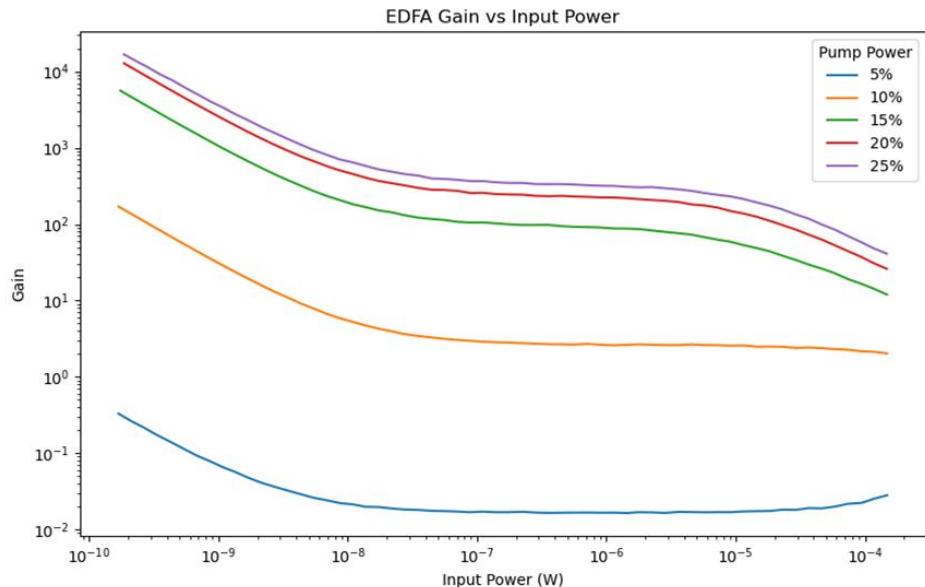
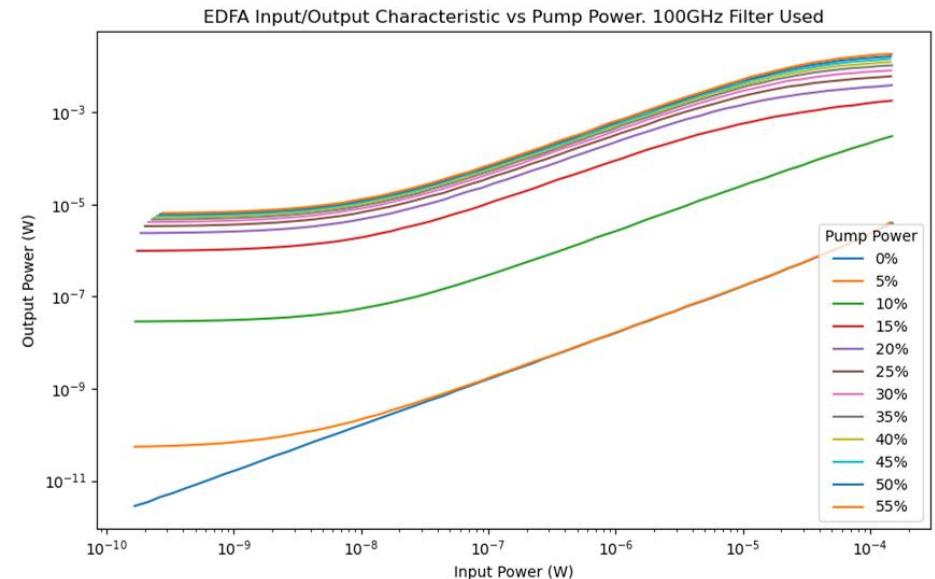


(b)

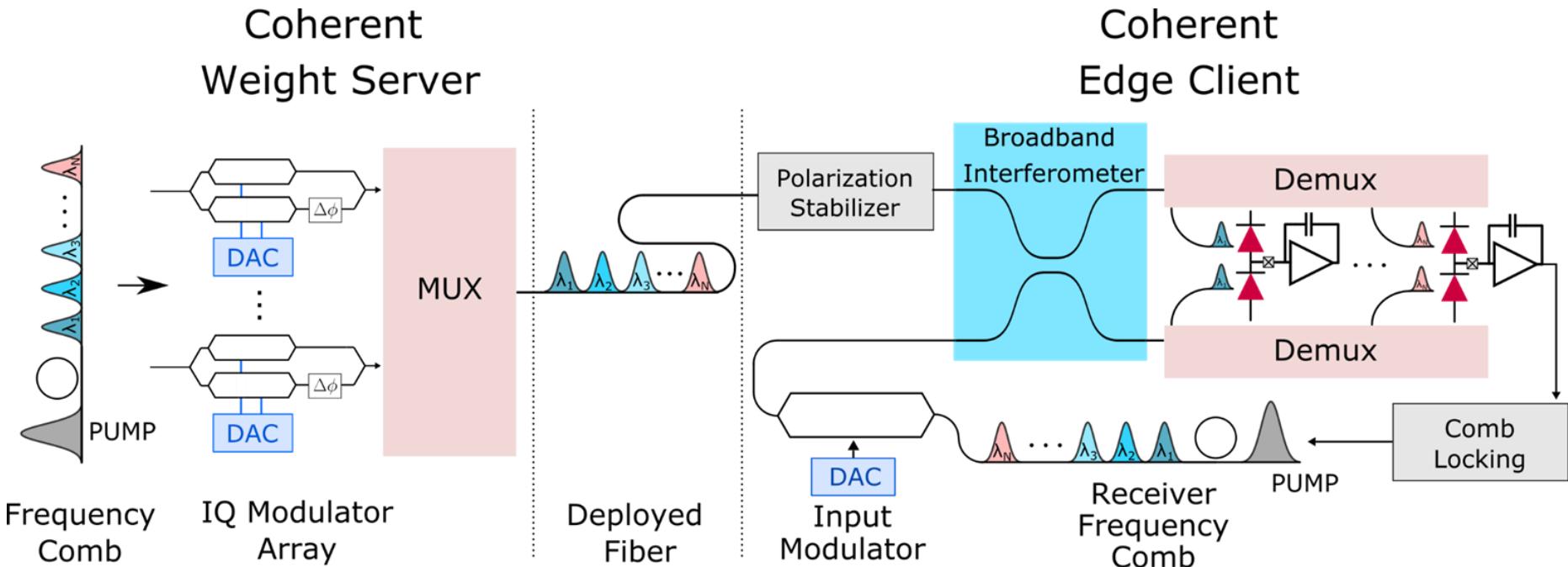
Optical Time Domain Reflectometry



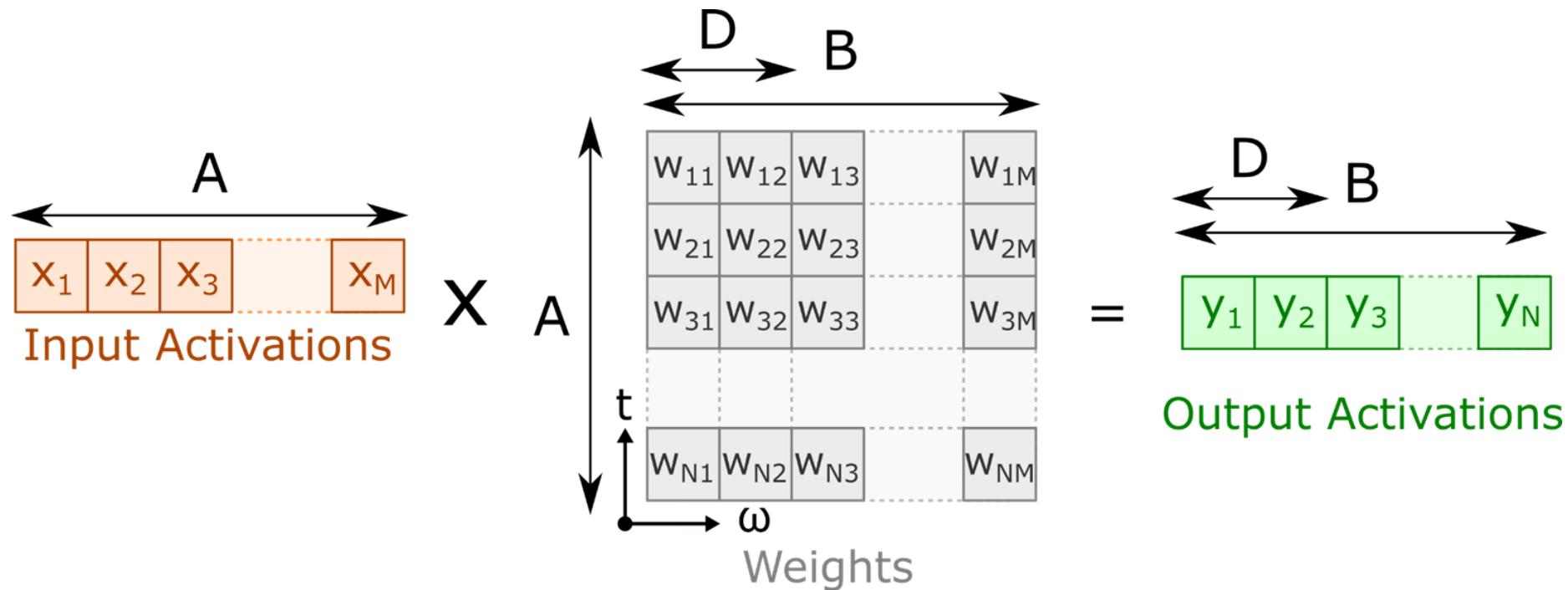
EDFA Gain/IO Characteristics



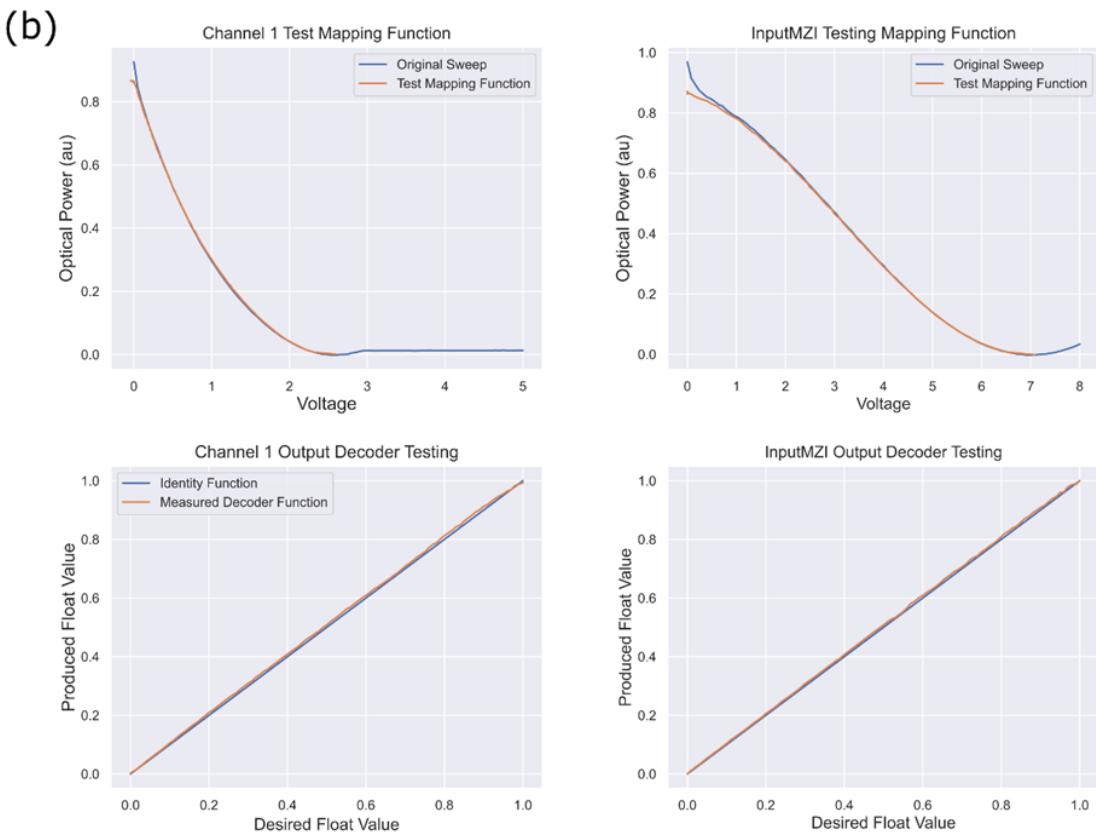
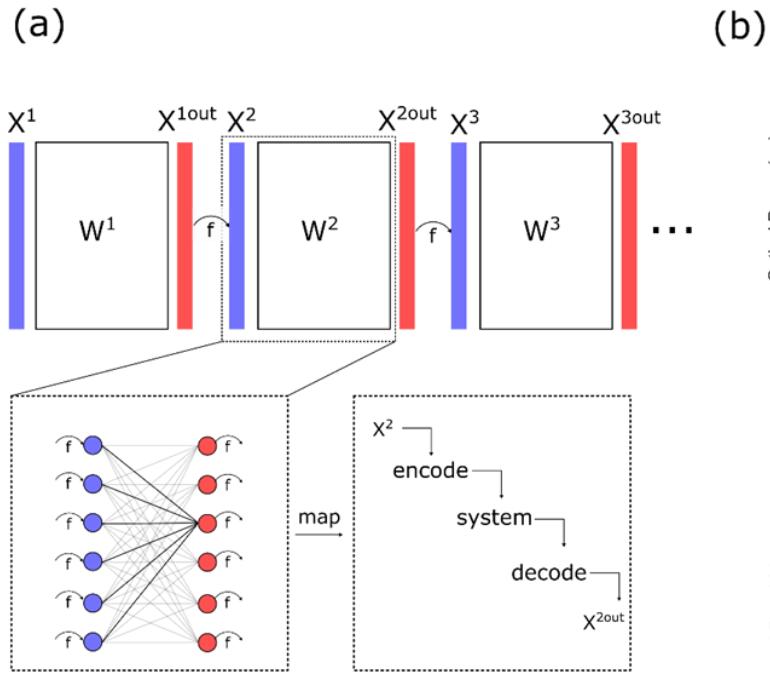
Coherent Netcast



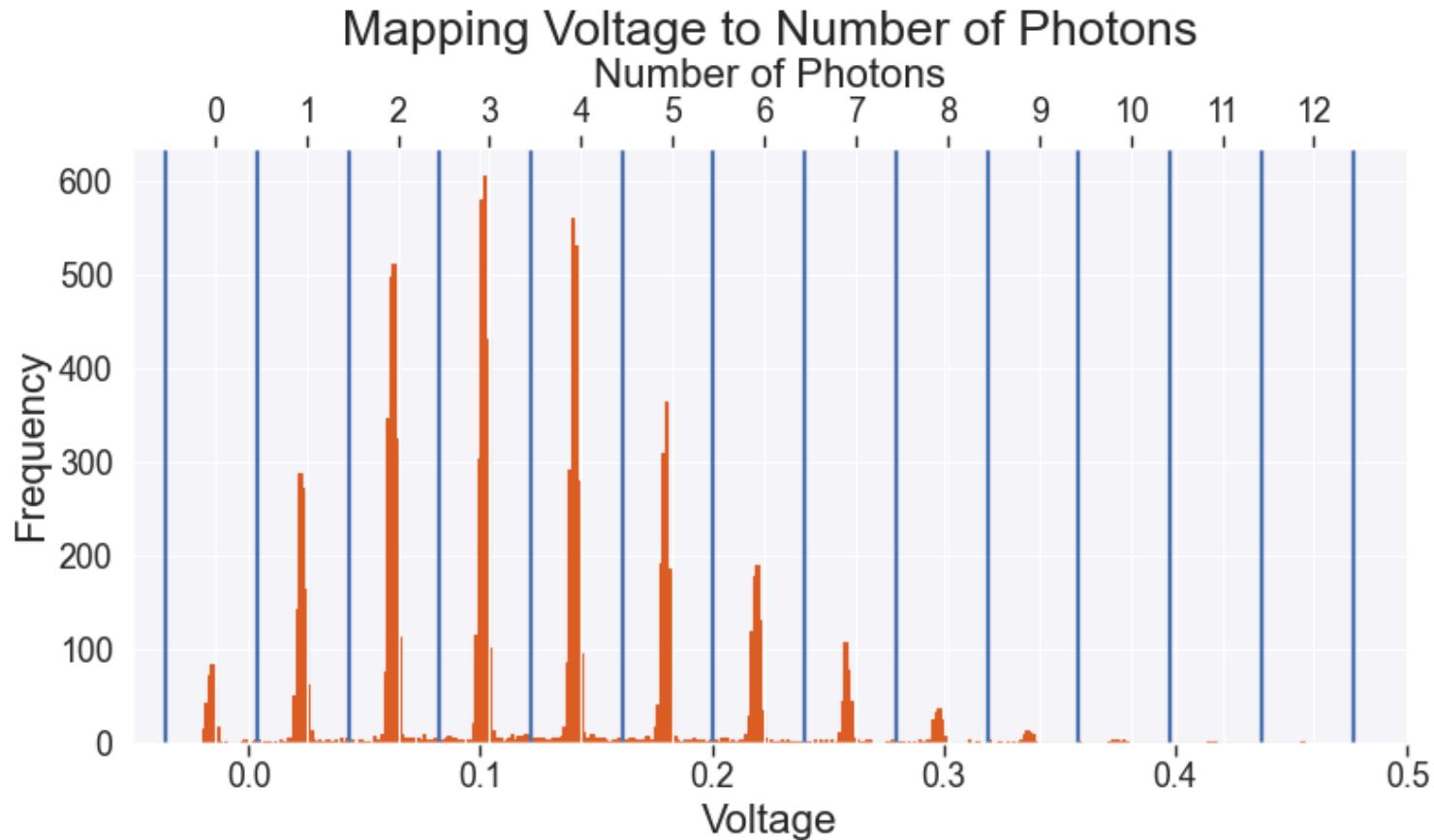
Mapping Matrix-Vector To Hardware



Calibration Of Netcast Hardware

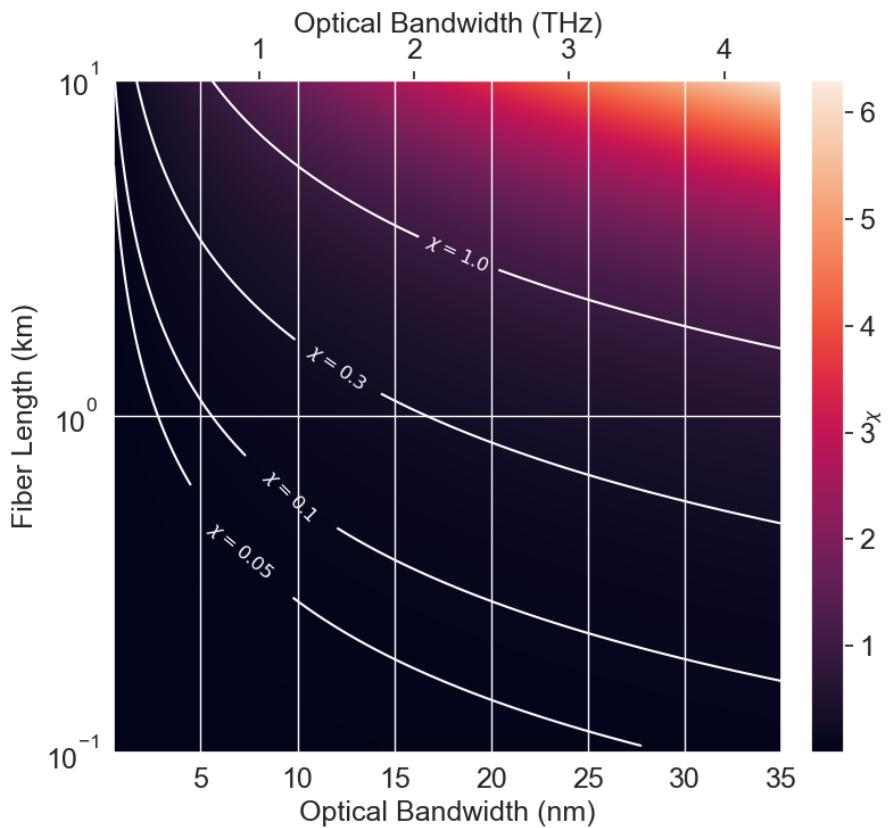


SNSPD Output Calibration

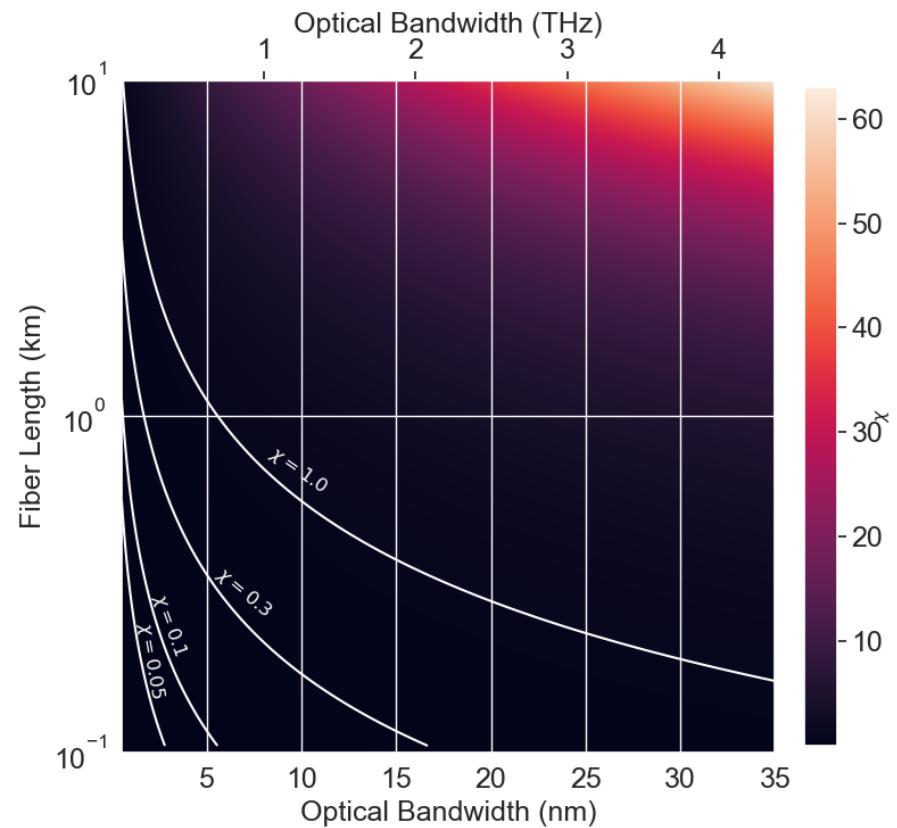


Dispersion Limit In C Band (SMF-28)

Dispersion Limited Temporal Crosstalk in C band (1GHz)

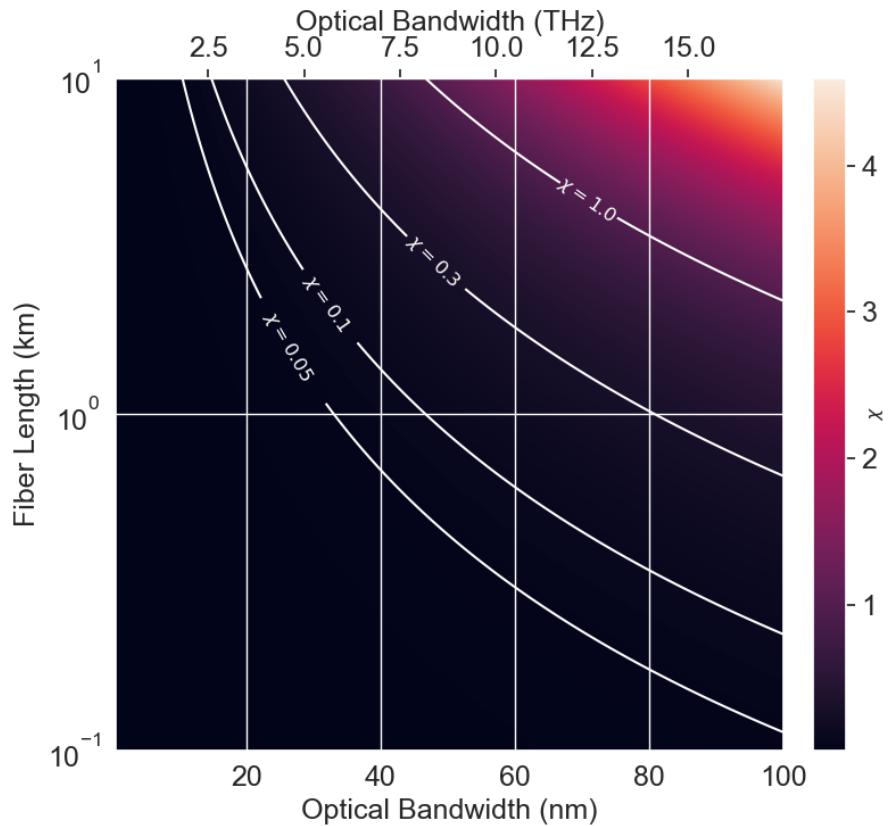


Dispersion Limited Temporal Crosstalk in C band (10GHz)

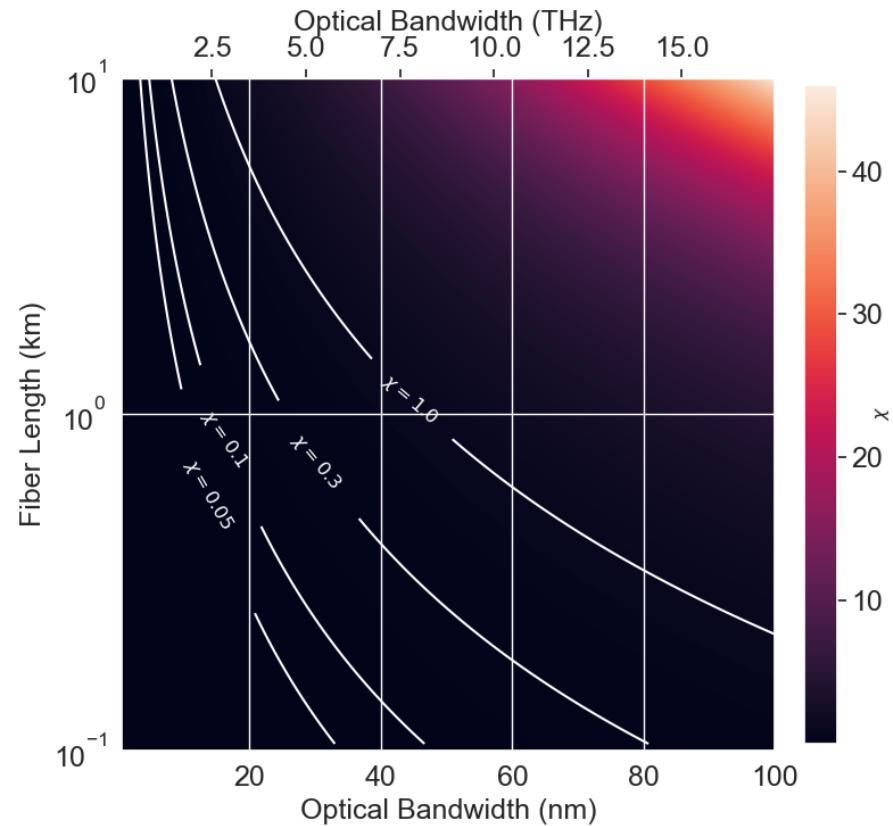


Dispersion Limit In O Band (SMF-28)

Dispersion Limited Temporal Crosstalk in O band (1GHz)

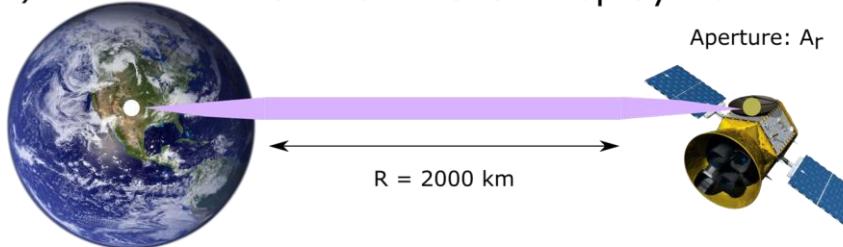


Dispersion Limited Temporal Crosstalk in O band (10GHz)

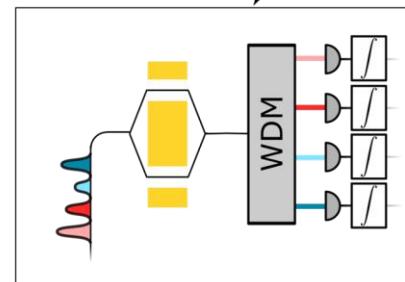
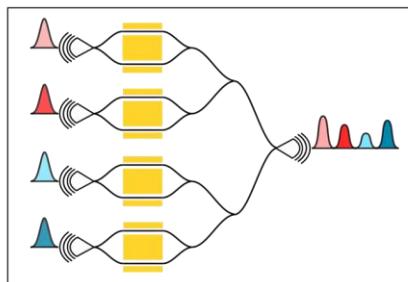
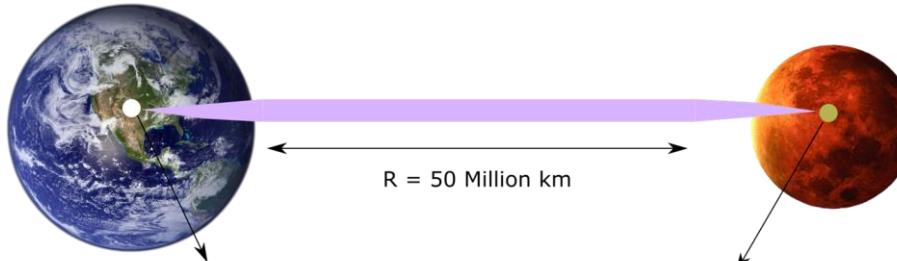


Example Space Deployment Of Netcast

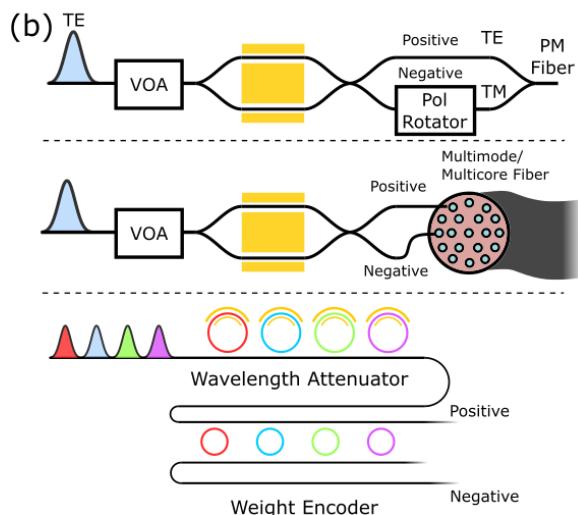
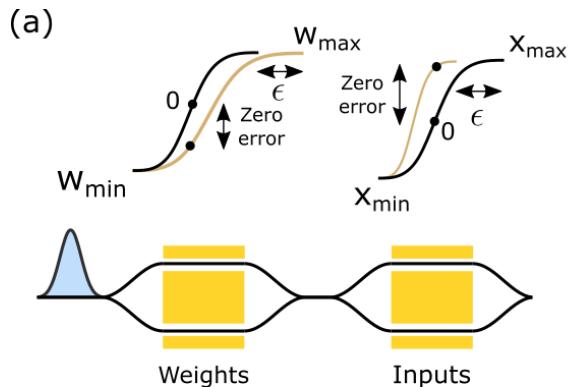
(a) Aperture: A_t Low Earth Orbit Deployment



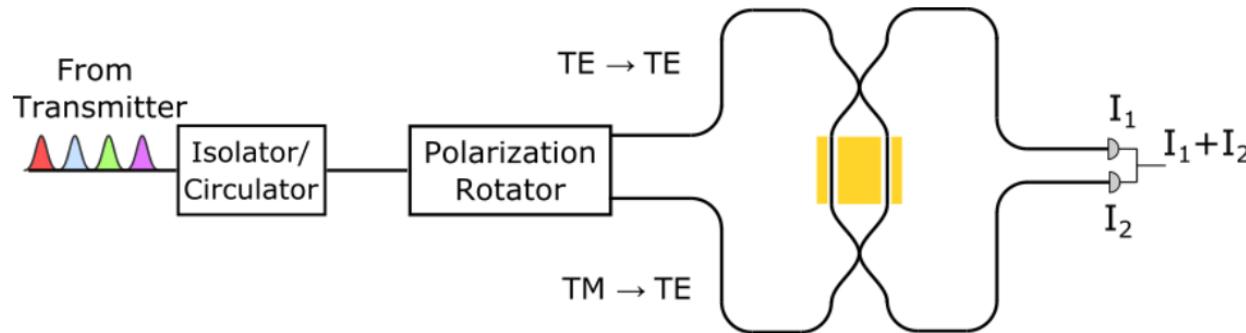
(b) Martian Deployment



Encode Negative Numbers In Netcast



Overcoming Polarization Mode Dispersion + Polarization Drift At Client



Phase Insensitive Coherent Detection

