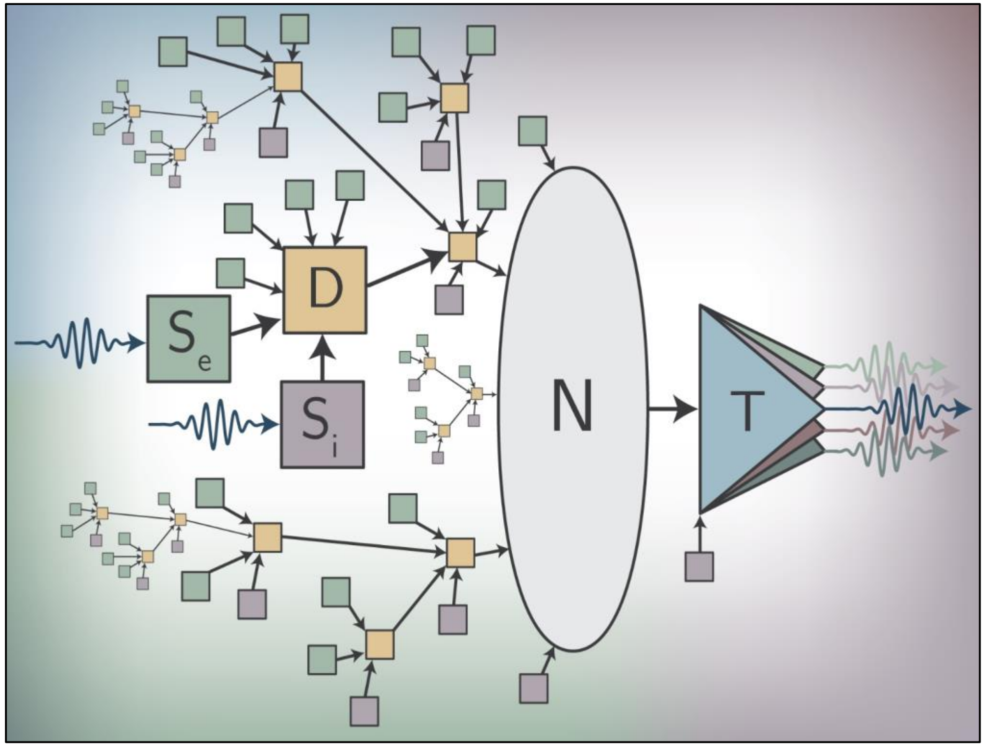
"Super" Neuromorphic Computing with Photonic and Superconducting Devices

Superconductors, light, and the "age of of cognition"



Jeff Chiles, Sonia Buckley, Adam McCaughan, Saeed Khan, Bryce Primavera,
Rich Mirin, Sae Woo Nam, Jeff Shainline
NIST, Physical Measurement Laboratory

NIST

High Performance Single Photon Detectors

- Superconducting nanowire single photon detector
- Superconducting Transition-Edge Sensors
- Microwave Kinetic Inductance Detectors

- Conventional Applications
  - Advanced sensors for astronomy and astrophysics
  - **High Energy Physics – rare particle event searches**
  - **Quantum Information / Quantum Computation**

Photonic Quantum Information Requirements

	Quantum Communication	Quantum Computing (Photonic)	Quantum Computing (atoms)	Entanglement-based random numbers
Wavelength	1550 nm	Visible, Near-IR	UV	Near-IR
Detection Efficiency	As high as possible	As high as possible	>80%	>67%

Dark / Background counts	As low as possible
Timing jitter	As low as possible
Maximum count rate	As high as possible

NIST

Conventional Single-Photon Detectors

	Wavelength Range	QE (%, max)	DCR (cps)	Jitter	Max Count Rate (cps)
PMT (visible)	400-900 nm	40	100	300 ps	10×10^6
PMT (IR)	1000-1600 nm	2	200K	300 ps	10×10^6
Silicon (thick)	400-1050 nm	65	25	400 ps	10×10^6
Silicon (thin)	400-1000 nm	49	25	35 ps	10×10^6
InGaAs APD	950-1600 nm	20	75K	350 ps	10×10^3

- Commercially available
- Relatively inexpensive

M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov, Rev. Sci. Instrum. 82, 071101 (2011)

NIST

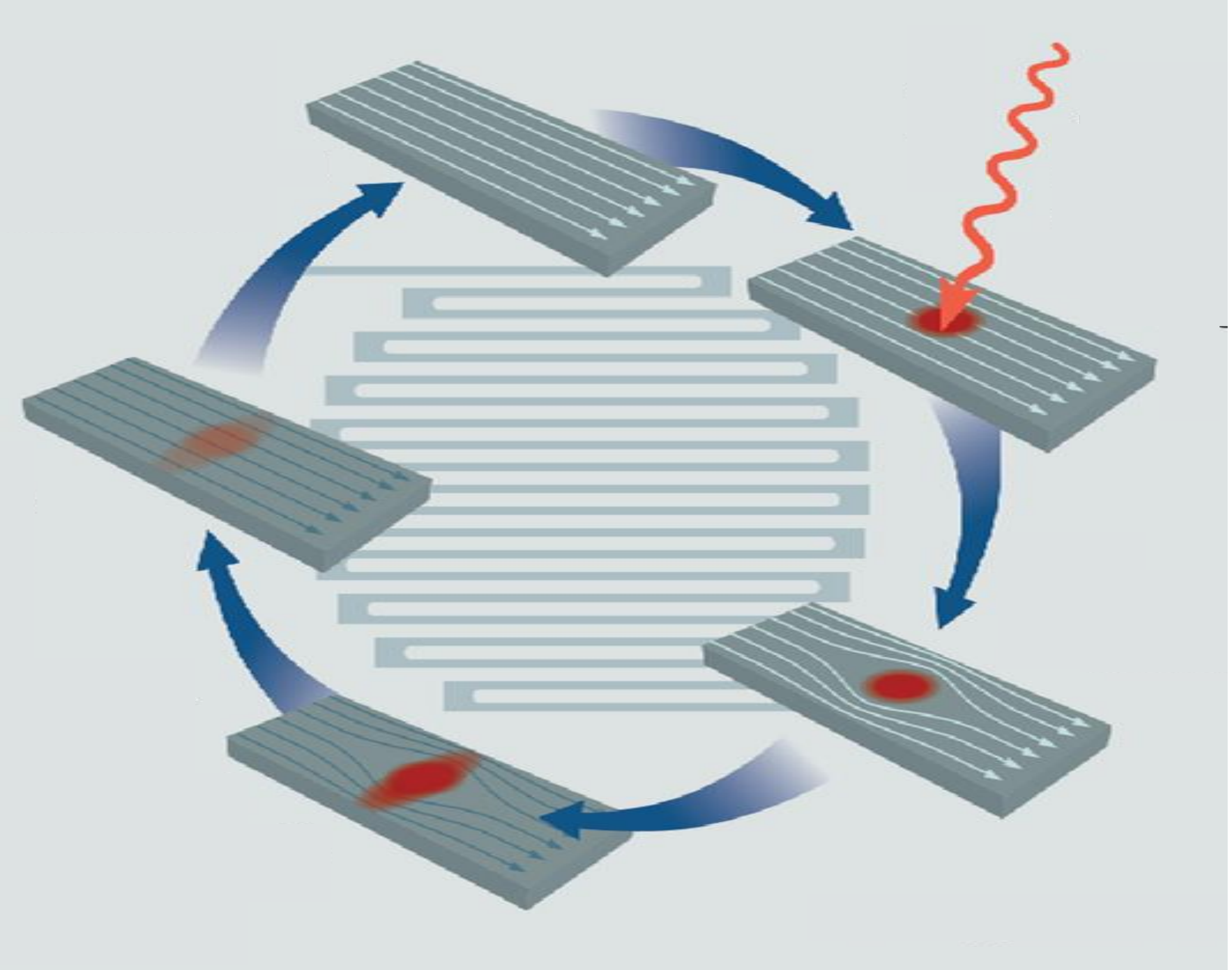
Low Temperature Physics = High Performance

	Wavelength Range	QE (%, max)	DCR (cps)	Jitter	Max Count Rate (cps)
W-TES (NIST)	UV-1850 nm+	>98%	<<1	10-100 ns	100×10^3
SNSPD: NbN	UV-5 um	>90%	100-1000	~3 ps	100×10^6
SNSPD: WSi	UV-5 um	~98%	$<<10^{-5}$	~5 ps	10×10^6

TES: Transition Edge Sensor
SNSPD: Superconducting Nanowire Single Photon Detector

- No afterpulsing problems
- Excellent prospects for longer wavelengths

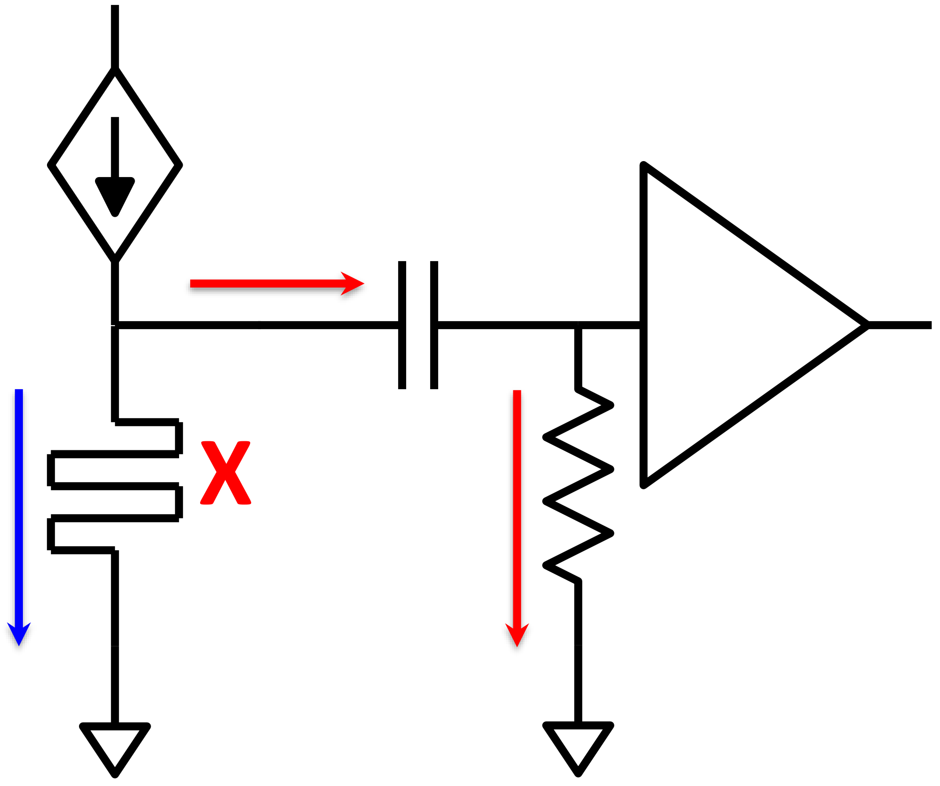
Superconducting Nanowire Single Photon Detectors:



- ultra-thin (2nm to 8nm)
- Anomolous kinetic inductance
- NbN, NbTiN
  - 2K operating temperature
  - ~80nm wide
- **New materials:**
- W-Si, Mo-Si, Mo-Ge
  - 1K operating temperature
  - ~150nm

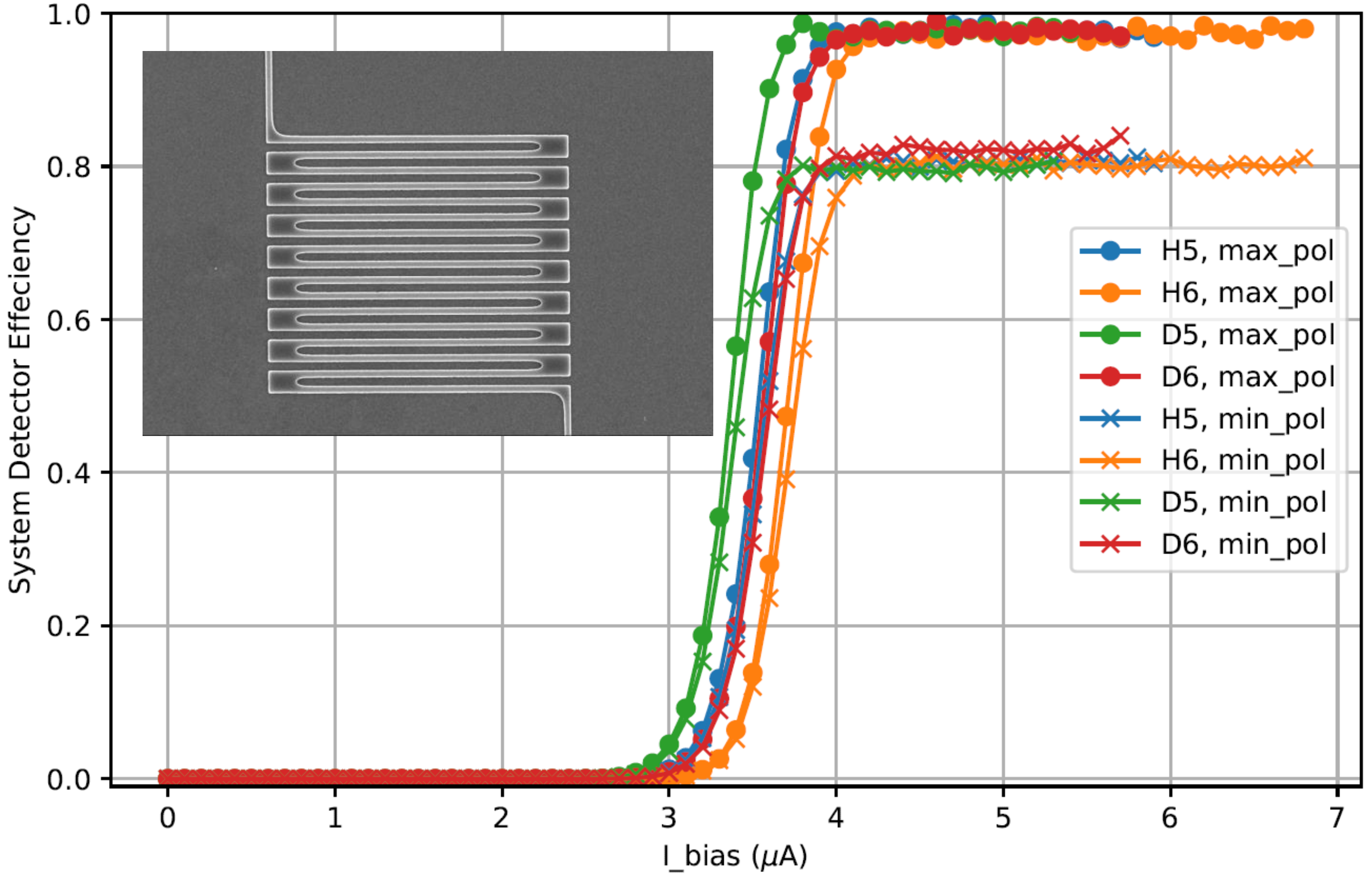
5 microns wide

Simplicity of Superconducting Nanowire Single Photoncs Detectors



Simple Readout

Bias versus Efficiency / counts



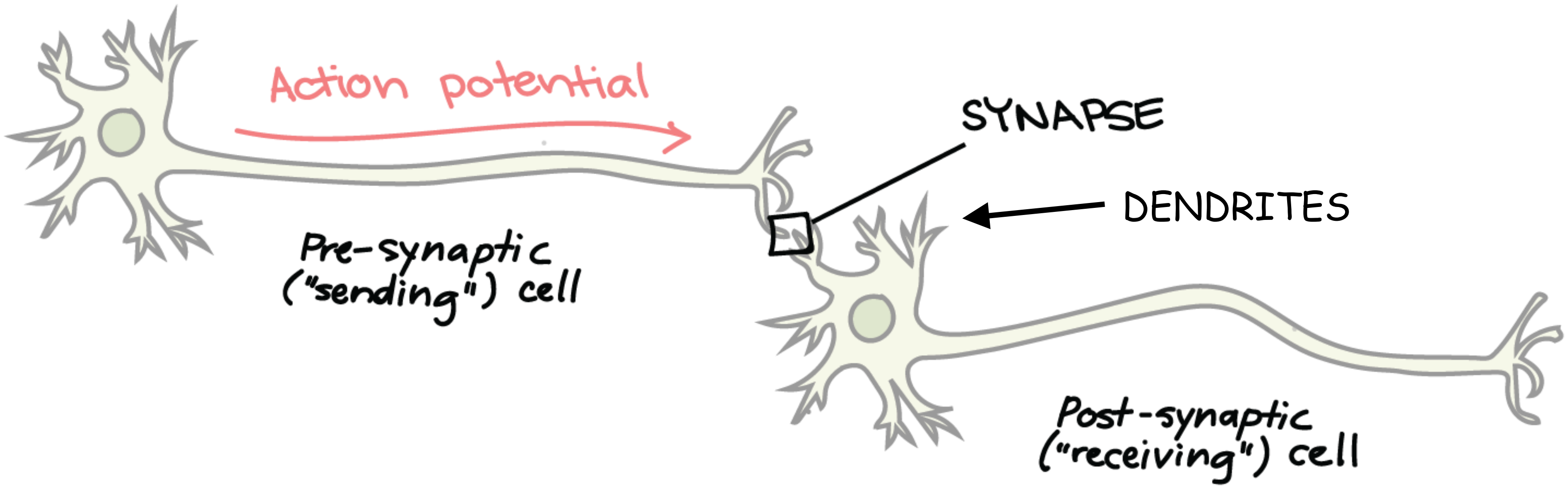
Future of Computing

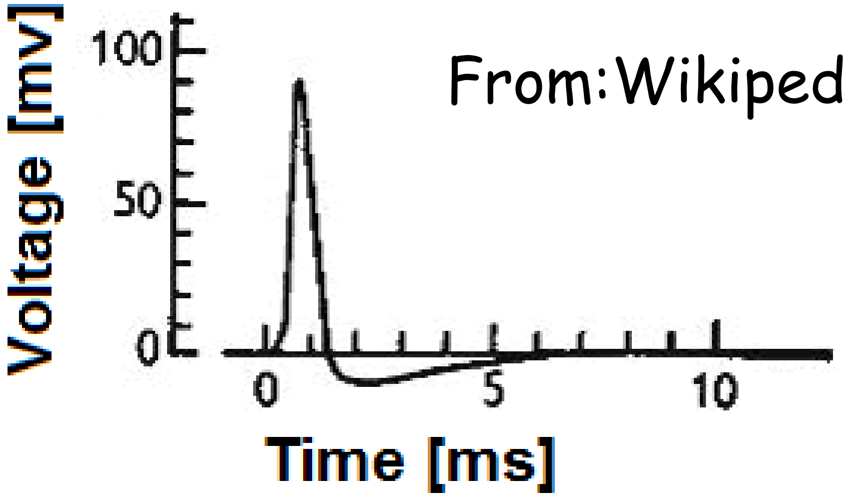
- High Performance Computing. (HPC)
  – Traditional model of computation

- Quantum Computing and Networking (QC)
  - Can outperform HPC in some tasks

- Neuromorphic Computing (NC)
  – "Neuroscience" inspired architecture
  – Dedicated Hardware

Structure of a Biological Neuron



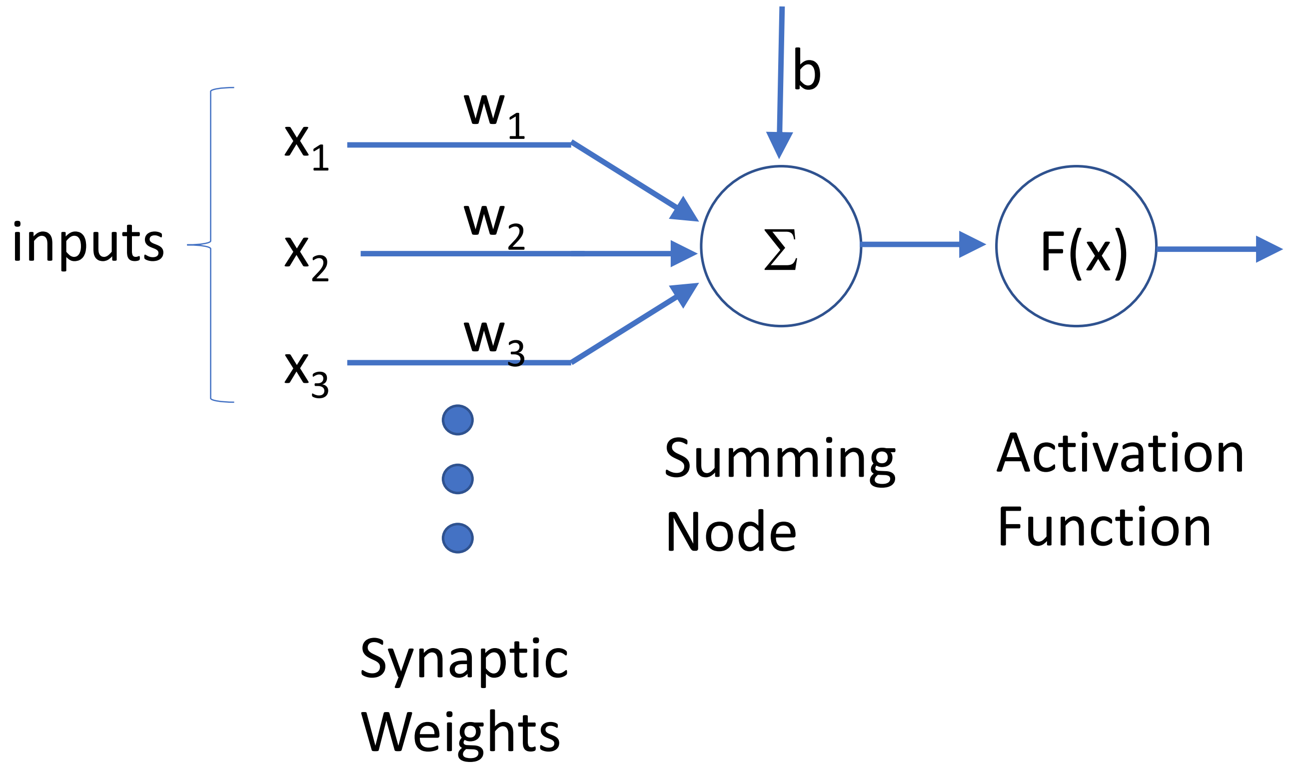


From: Wikipedia

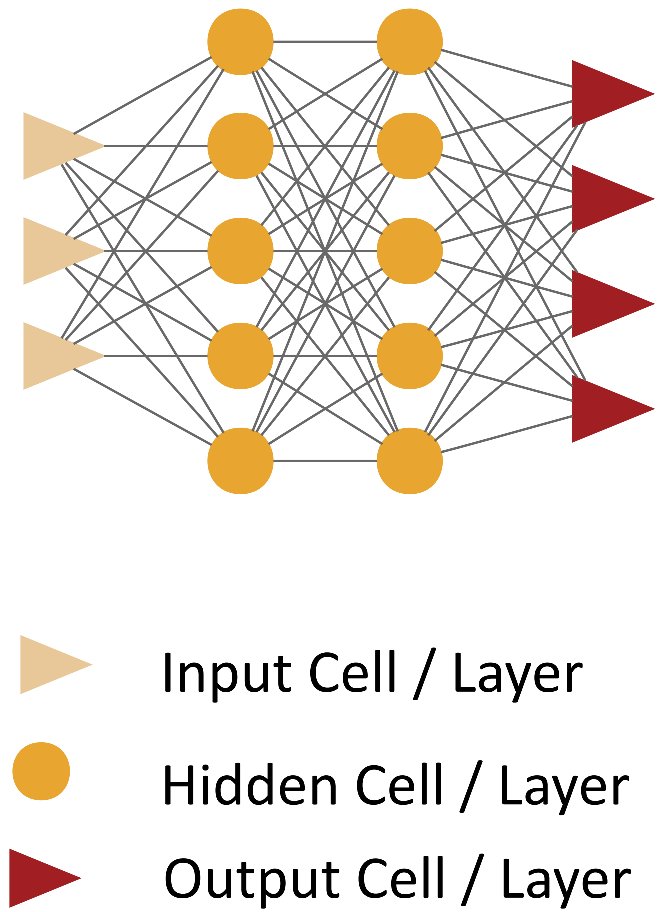
https://www.khanacademy.org/science/biology/human-biology/neuron-nervous-system/a/the-synapse

Fast Introduction to Neural Networks

Artificial Neuron



Network of Neurons



NIST

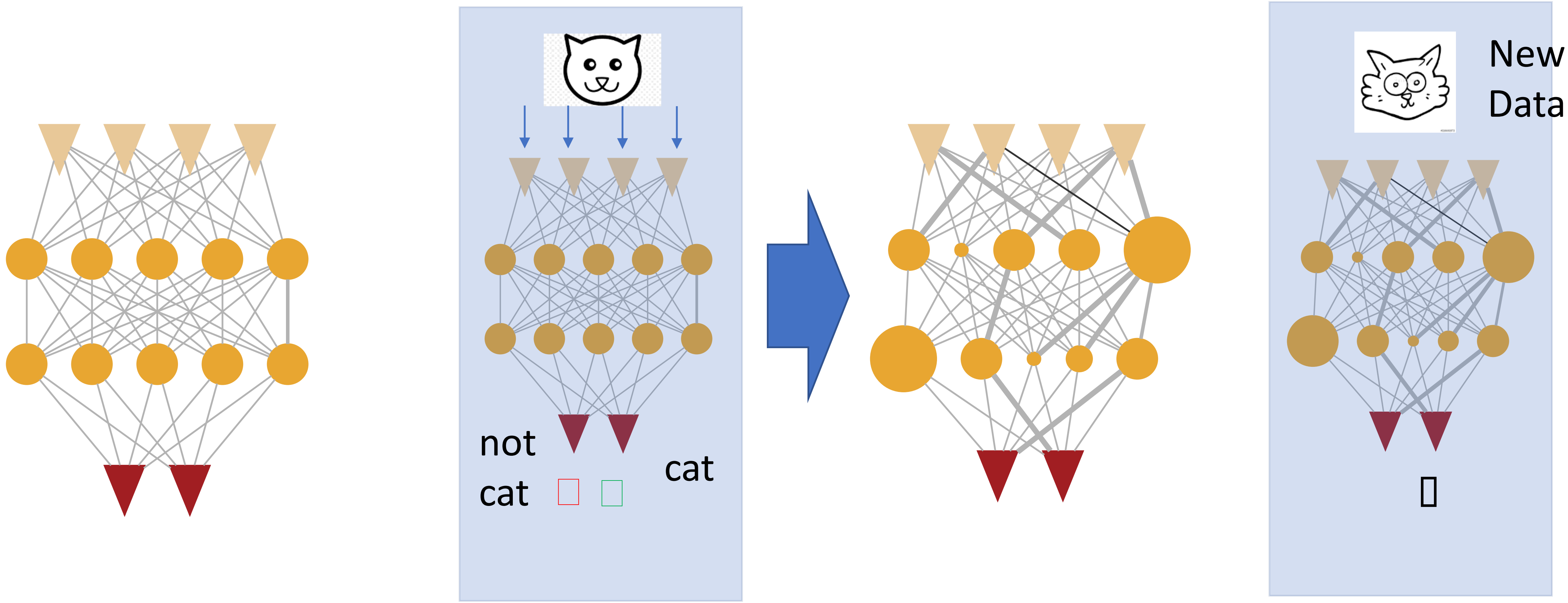
Supervised Learning

Untrained Neural Network

Training

Trained Neural Network

Inference



Machine learning hardware

- Today, there is large demand to perform "neural network" operations
  - Image Classification
  - Speech Recognition
  - Natural Language Processing: "translation"

YouTube amazon Google Uber

- Dedicated Neuromorphic Hardware

TrueNorth, IBM

see Merolla et al., Science, 2011.

IBM intel

Human Brain Project

- Other Technology platforms Neuromorphic have attracted venture interest

SPARK

GIGAFUND

Bill Gates

2nd and 3rd Generation Neural Networks

2nd Generation: “Proven”

- Limited Biological “inspiration”
- Matrix-vector multiplication
- Seeking weight matrix
- Trained by supervisory system
- Limited History
- Task specific

3rd Generation: “Less Proven”

- More Biologically inspired
- Information in dynamical state
- Spiking neurons communicate in rate and timing
- Potential for learning without supervision
- Information integrated across space and time
- General cognitive systems

Question: Why do we need Generation 3?
Answer: Energy Efficiency and Size

Consumption	CO_2e (lbs)
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000

Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

https://arxiv.org/abs/1906.02243v1

- “How do you make the largest scale artificial neural network?”
  - Human Brain:
    - 10 billion neurons
    - 1 neuron fans out to 10,000 neurons
- What are the fundamental limits?
- How do you evaluate the performance?
- How do physical characteristics of the devices relate to the performance of the neuromorphic computer?

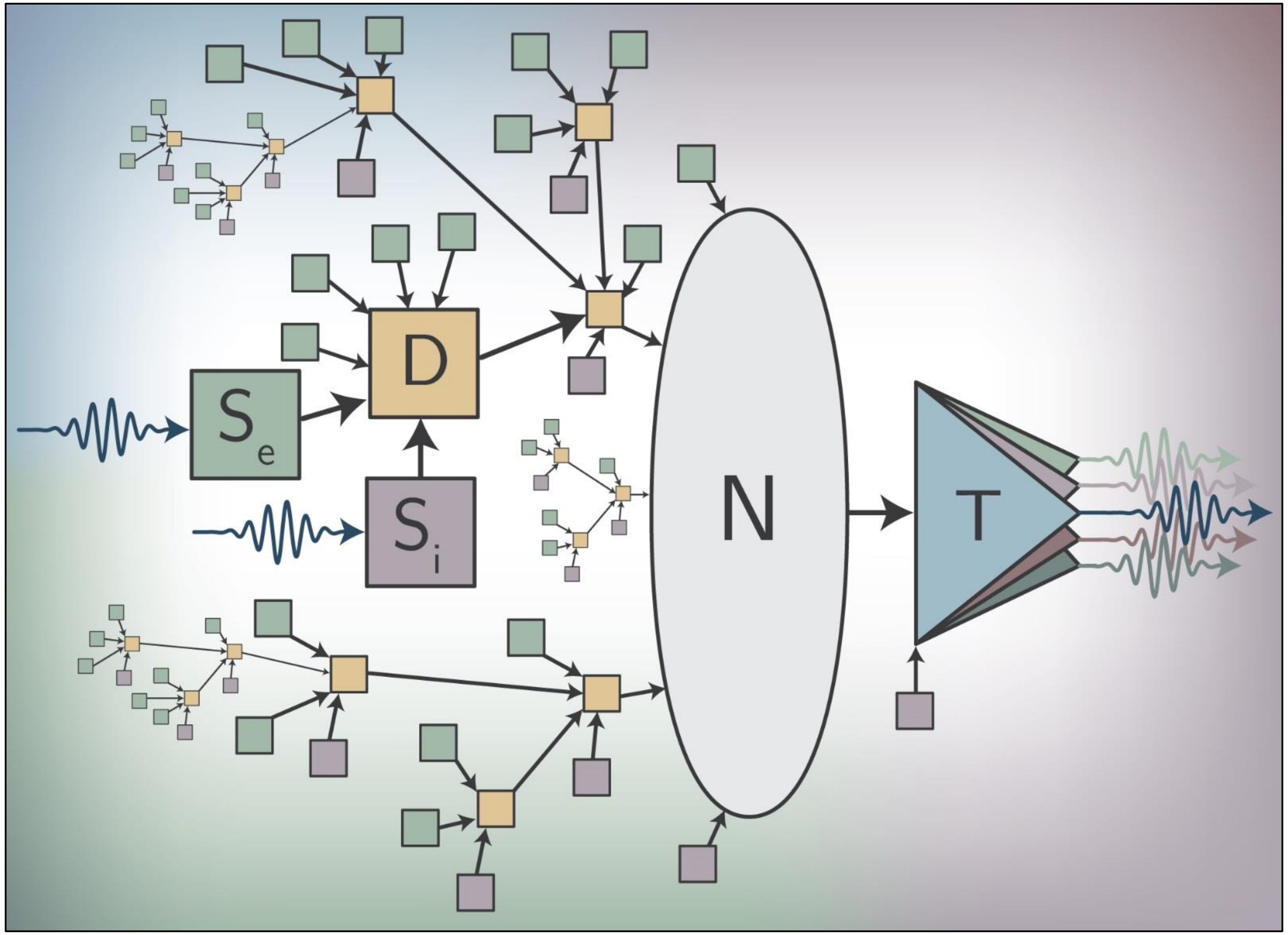
Spiking Neural Networks / Dedicated Hardware

- Biologically Inspired
  - Spiking Signals, Energy Efficient
  - Rate and Time encoding
- Differentiated local processing
- Information integration across:
  - Space (network structure)
  - Time (dynamics)
  - Experience (plasticity)

Can we get to a size scale for "Cognitive" Systems?

Energy and Training Targets

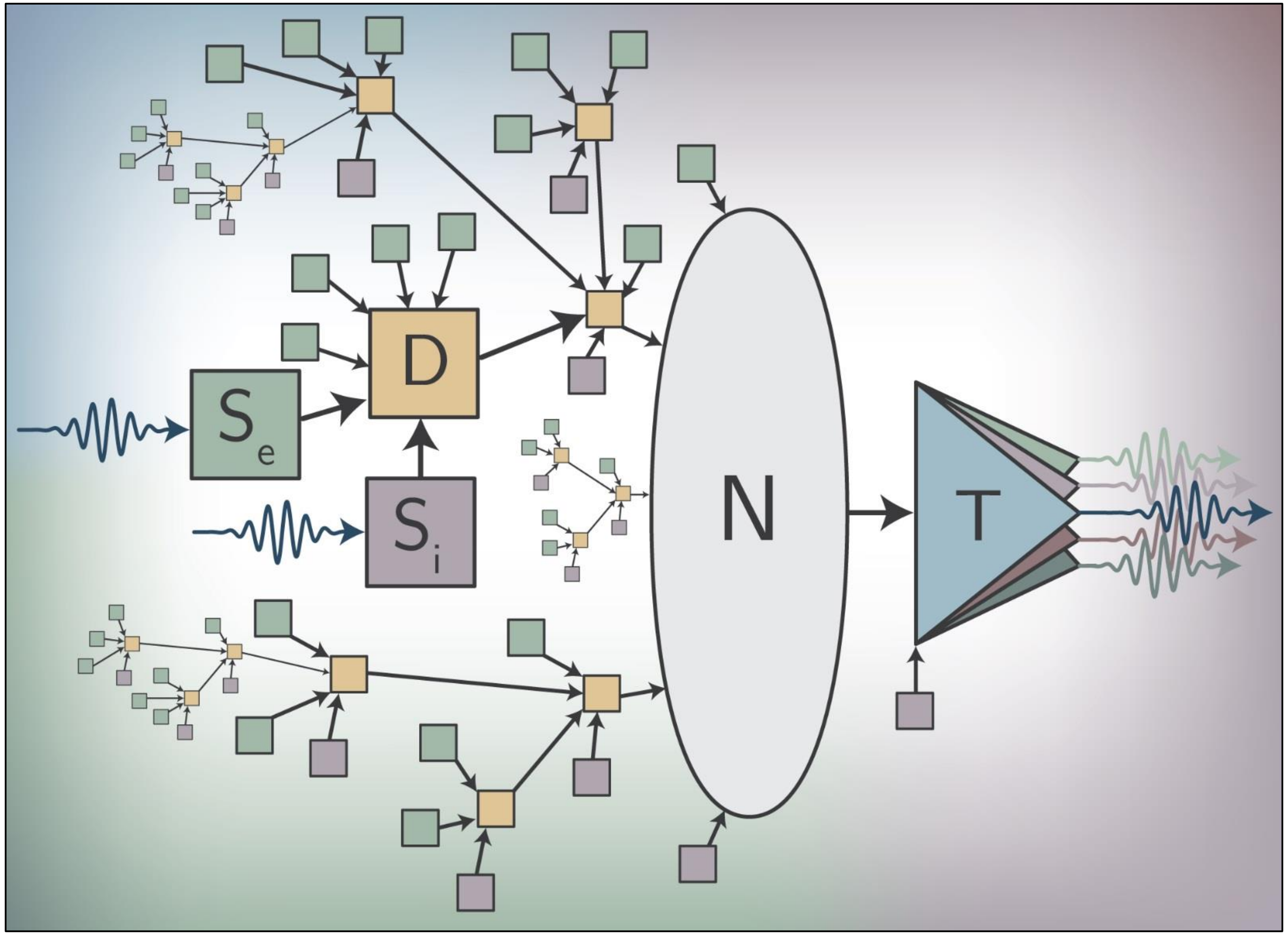
	Power	Clock speed	Training time
Human	2500 kCal/day = 100Watts	1 Hz to 1kHz	18 years
Artificial	100 kW to 1 MW	1 Mhz to 10MHz	1 week



Fluxonic processing of photonic synapse events

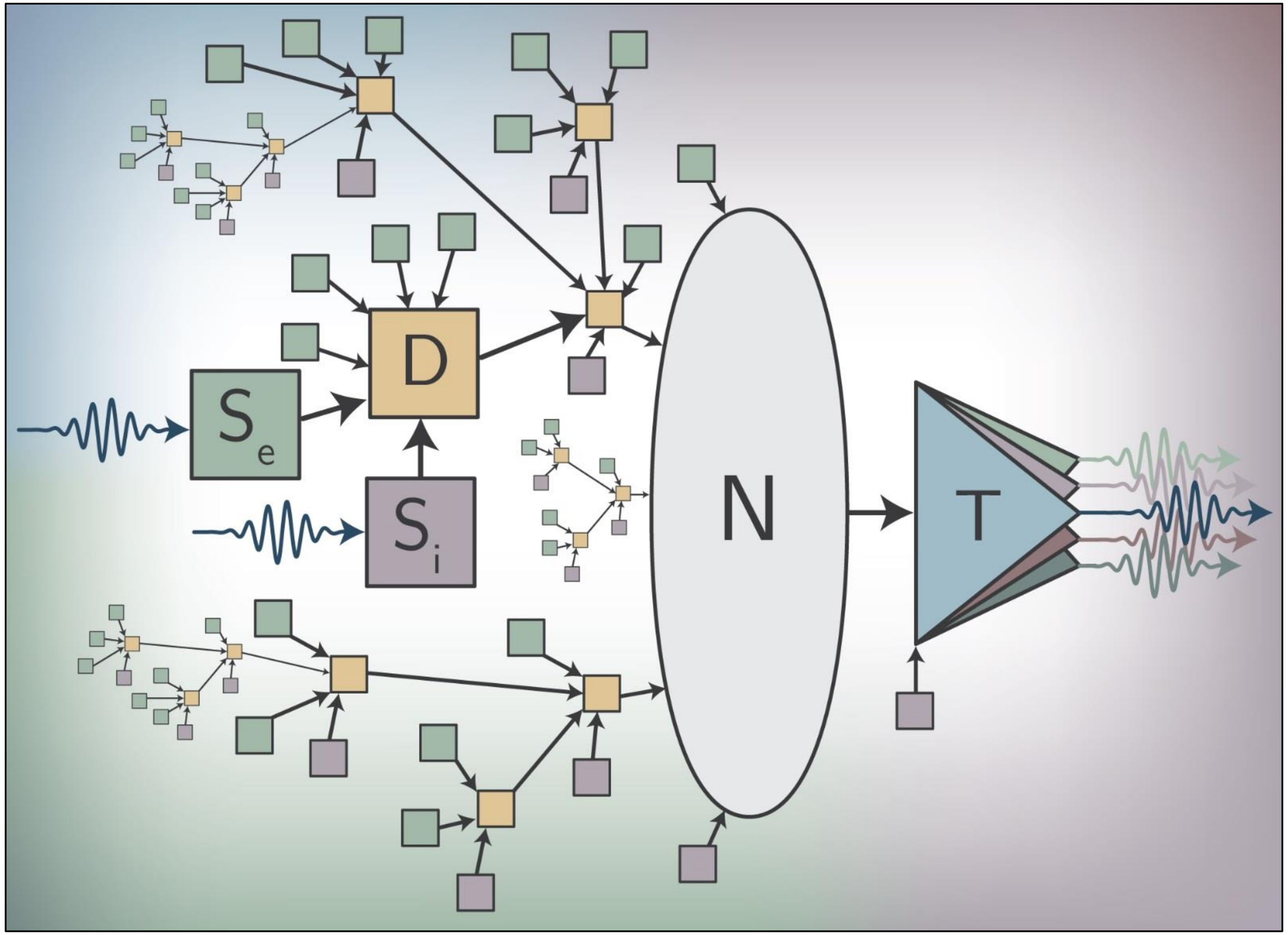
Spiking neural networks

jeffrey.shainline@nist.gov



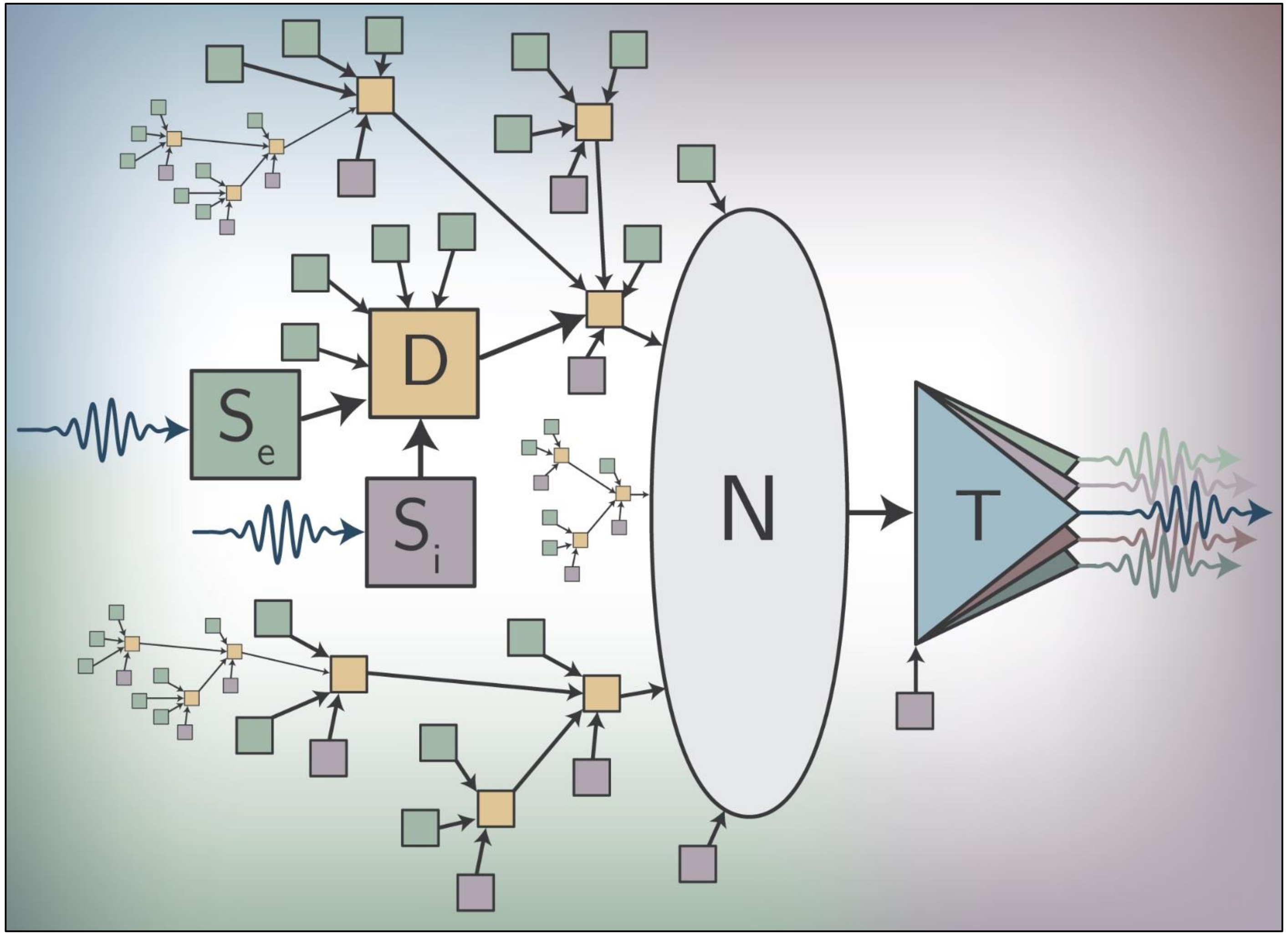
Fluxonic processing of photonic synapse events

Light for communication



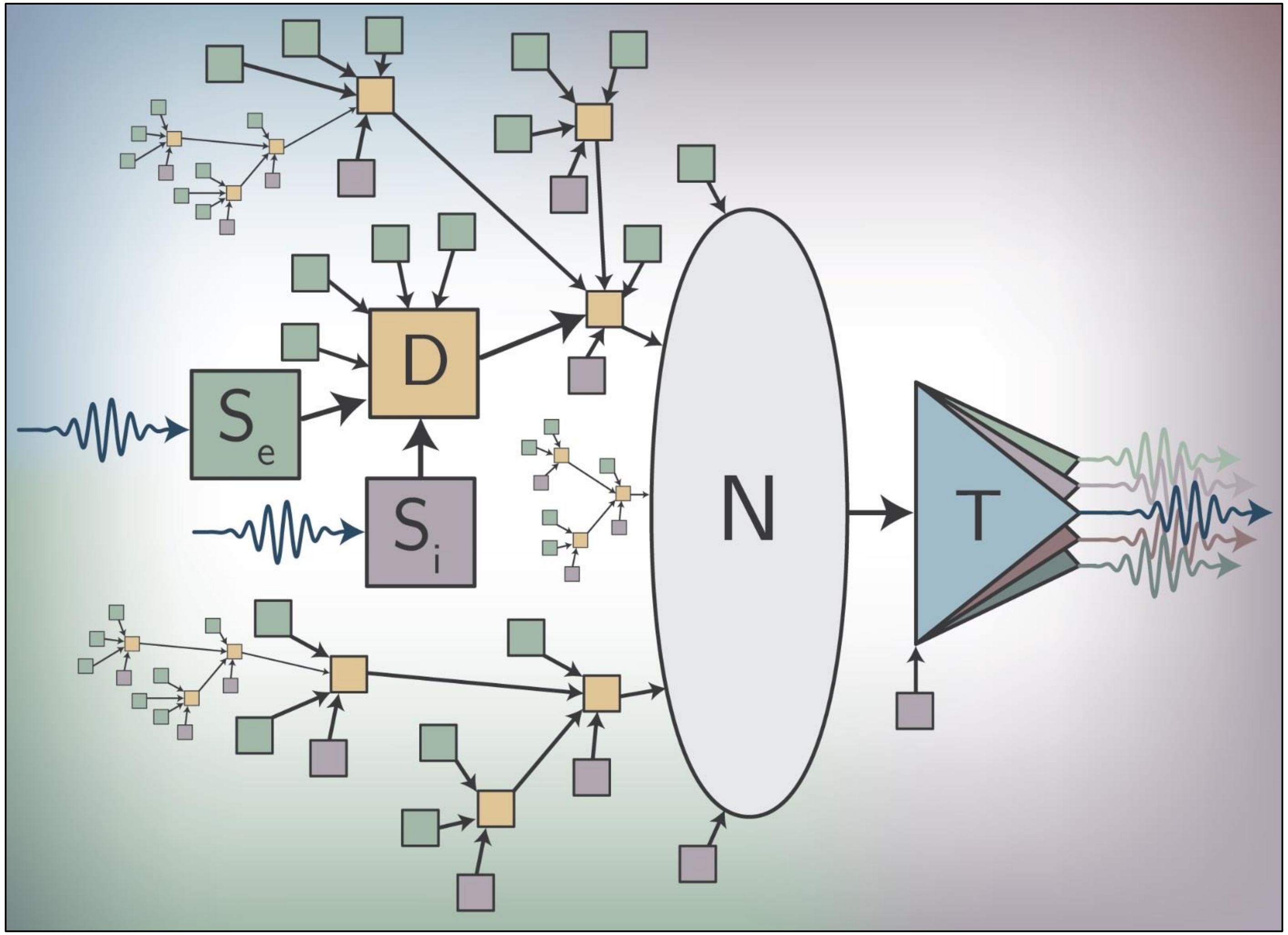
Fluxonic processing of photonic synapse events

Superconducting electronics for single-photon detection



Fluxonic processing of photonic synapse events

Superconducting electronics for neural computation



Fluxonic processing of photonic synapse events

Neuromorphic supercomputing

Device requirements for massive connectivity

Dense local fanout

Long-range connectivity

We seek very large systems

Energy efficiency is paramount

Principal conjecture:

Use light for communication

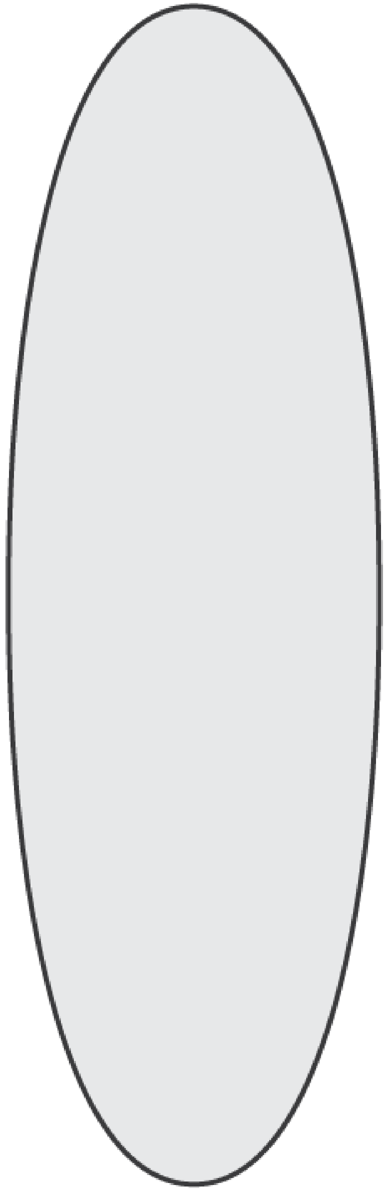
Photons don't have charge or mass

Use single-photons for communication

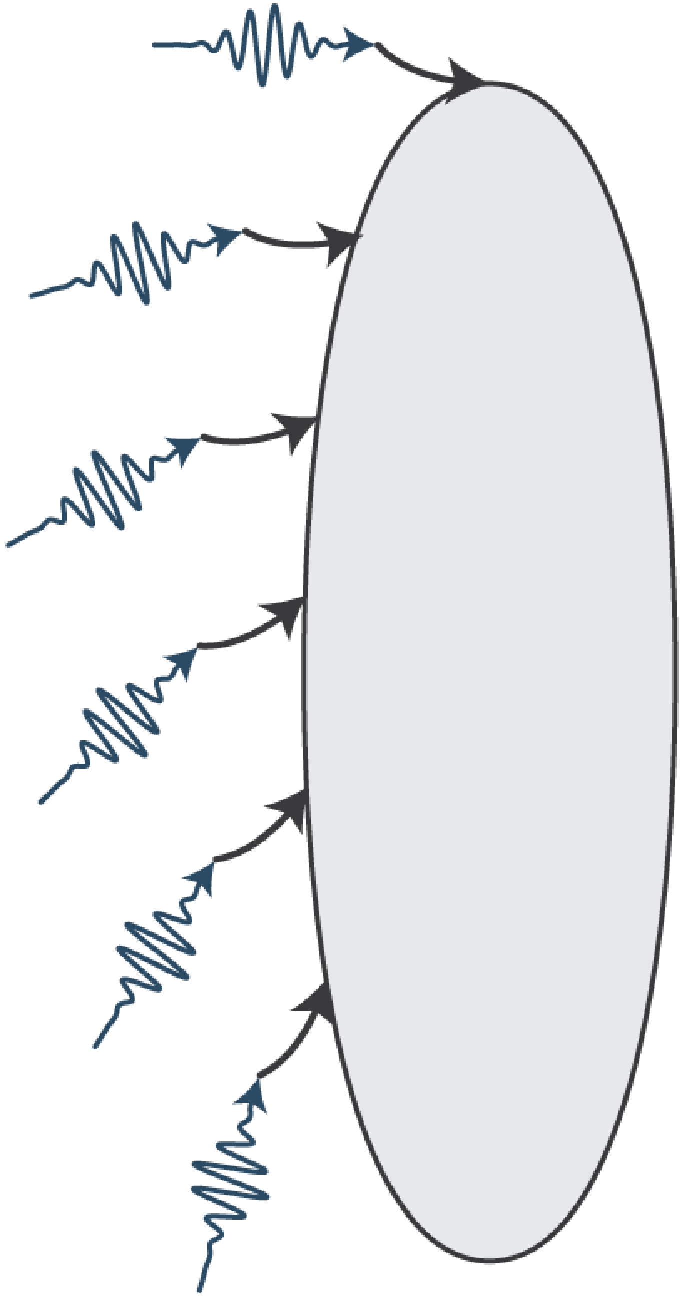
Neurons that signal with single photons

Neurons that signal with single photons

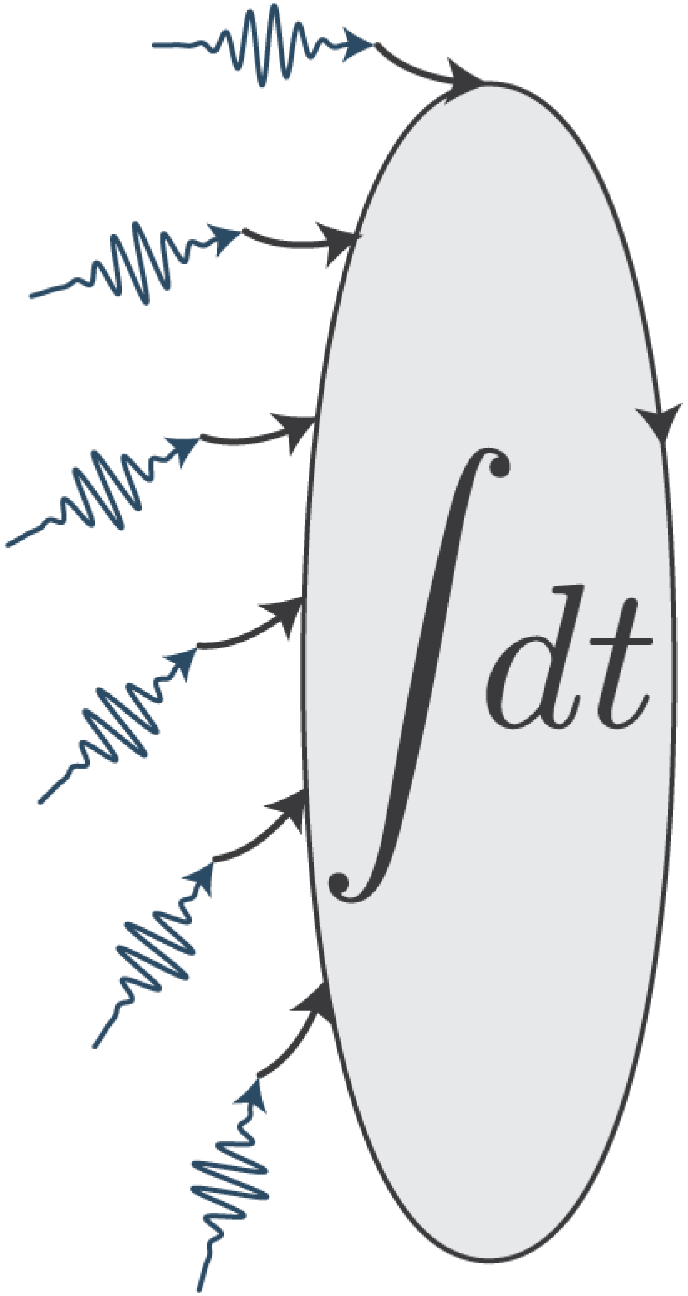
How do they work?



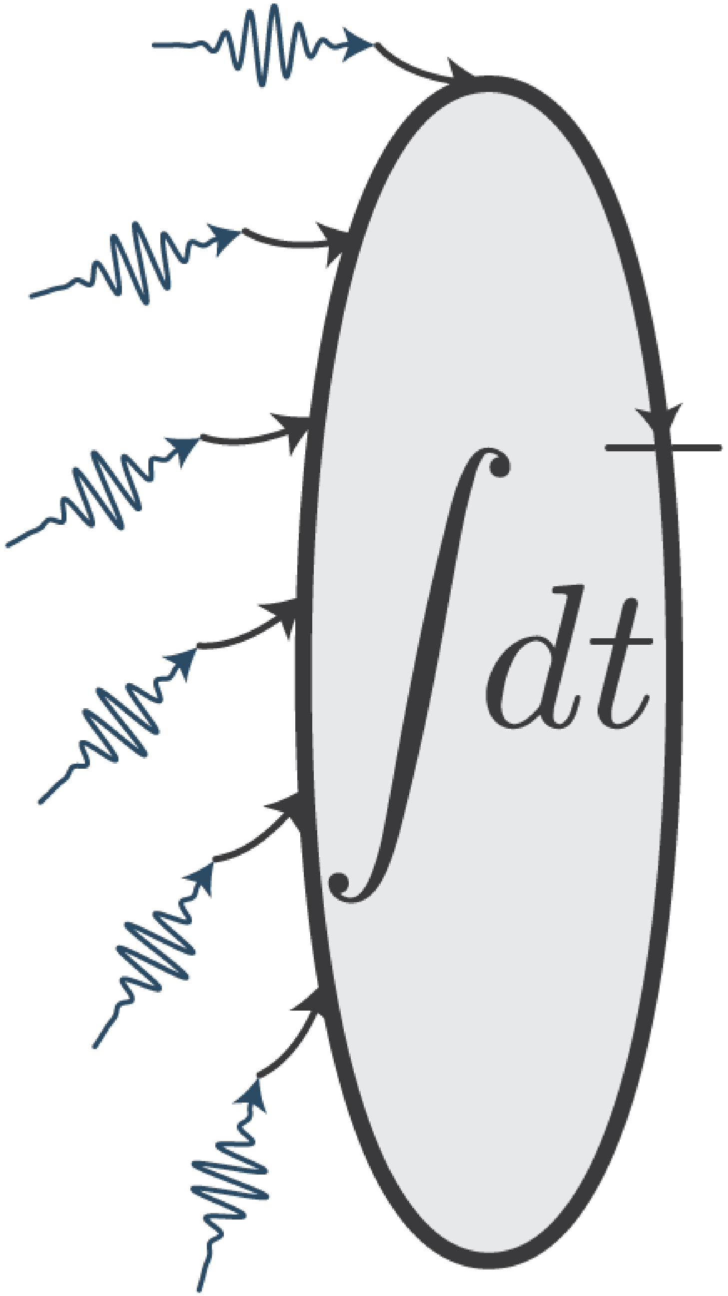
Superconducting loop



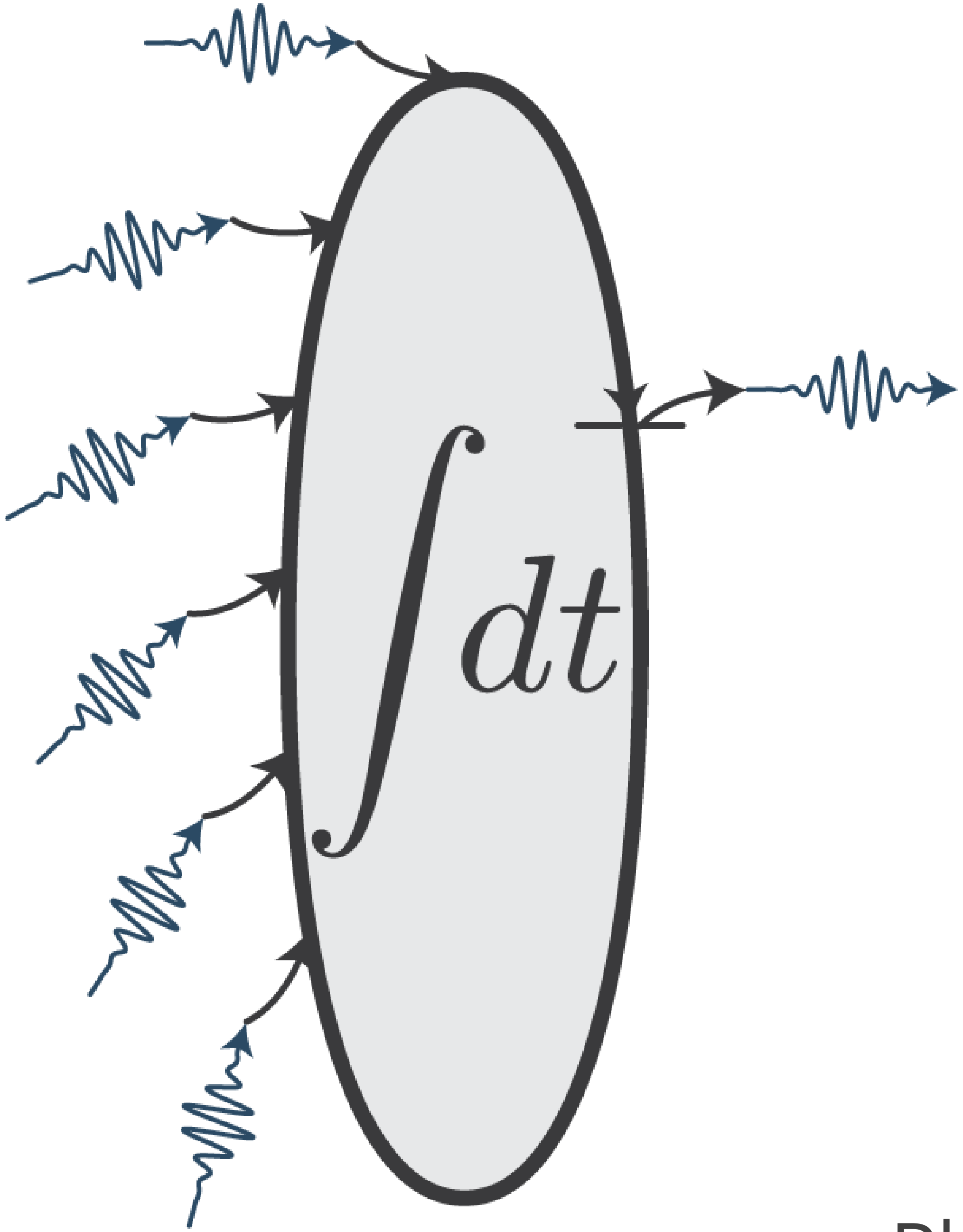
Photons add current



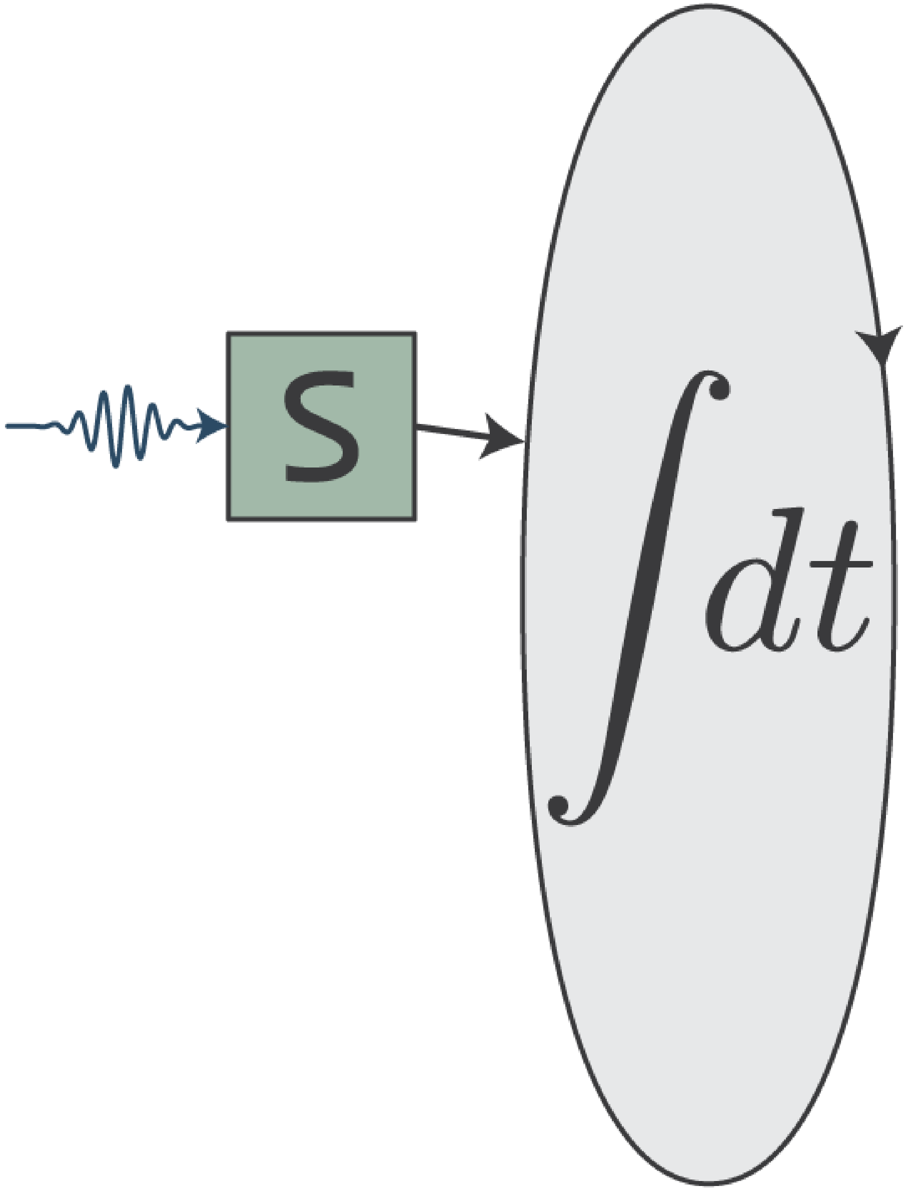
Current gets integrated



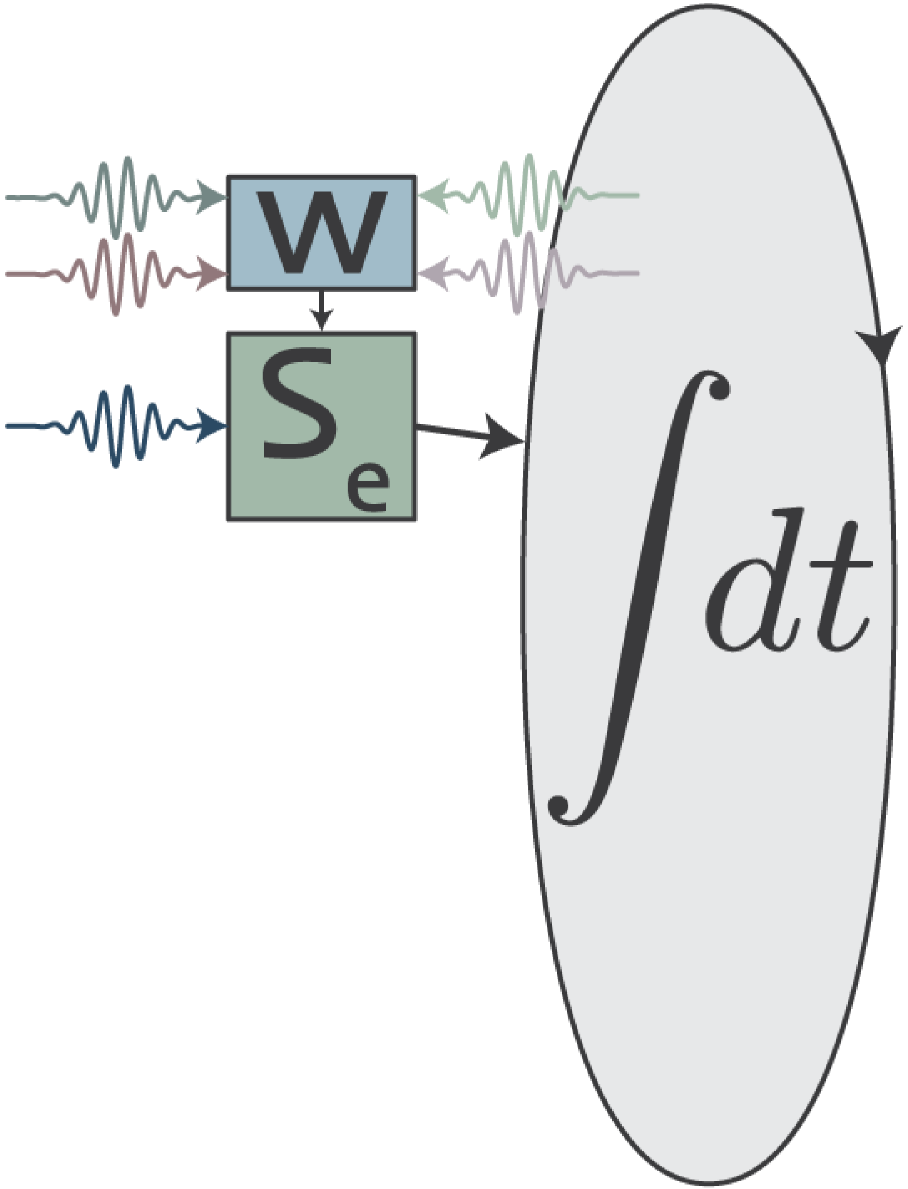
Current threshold



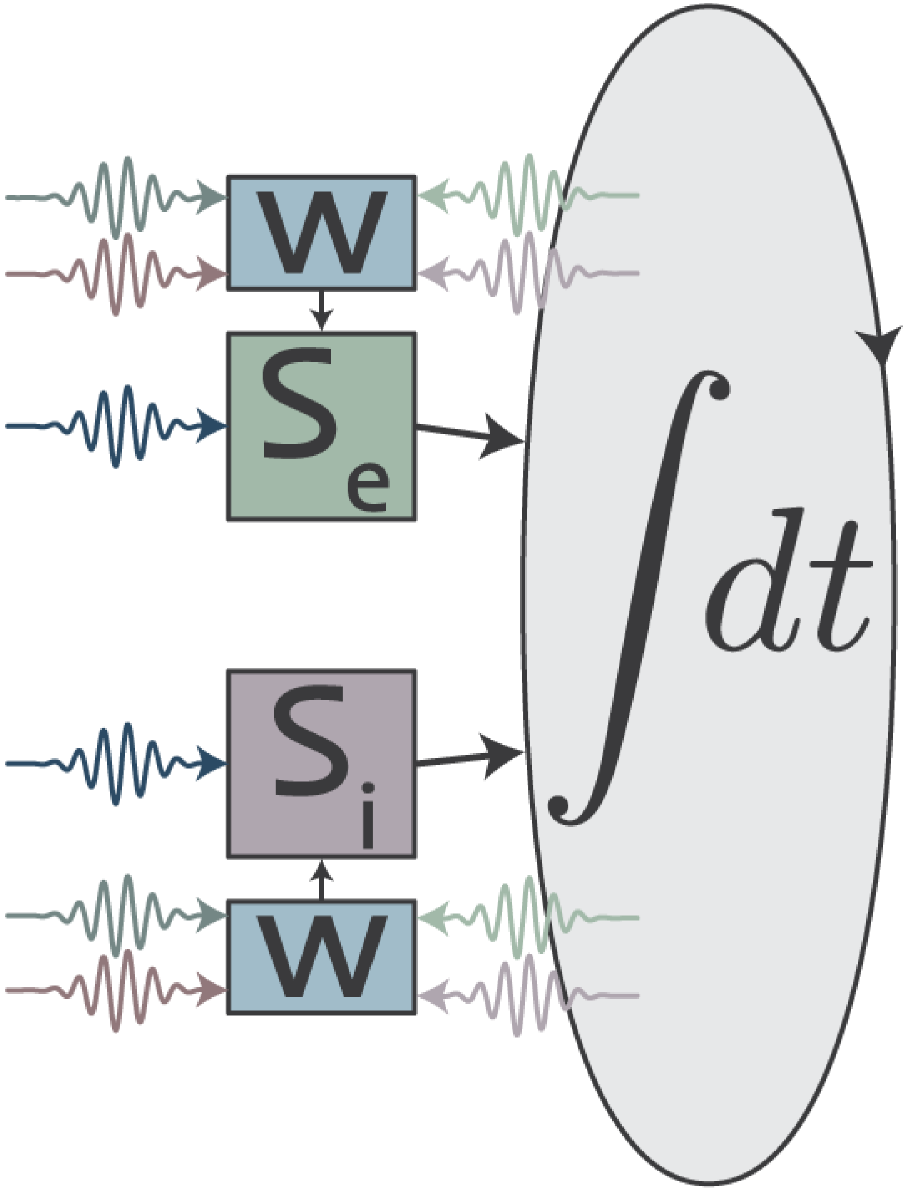
Photons produced



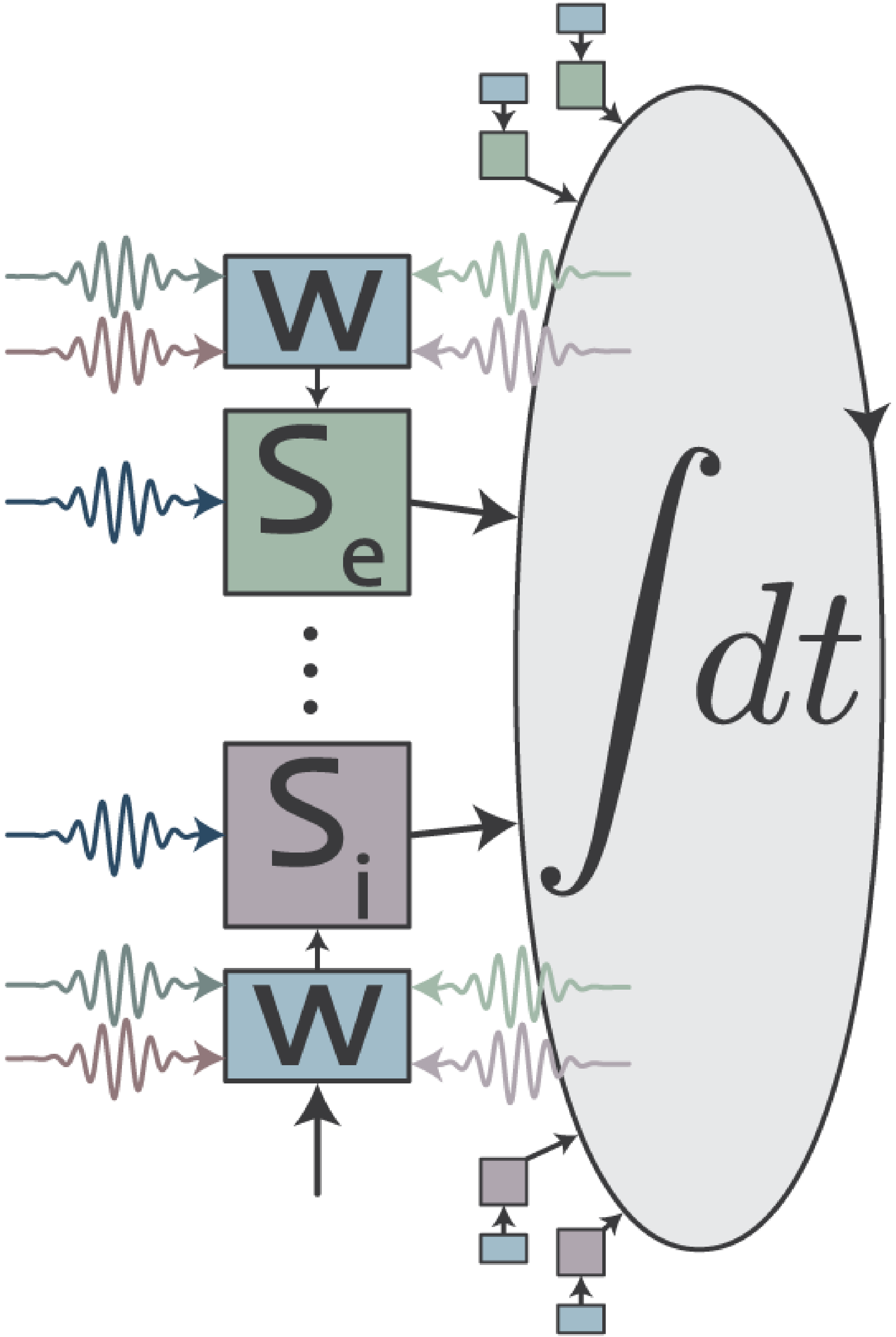
Synapse transduces



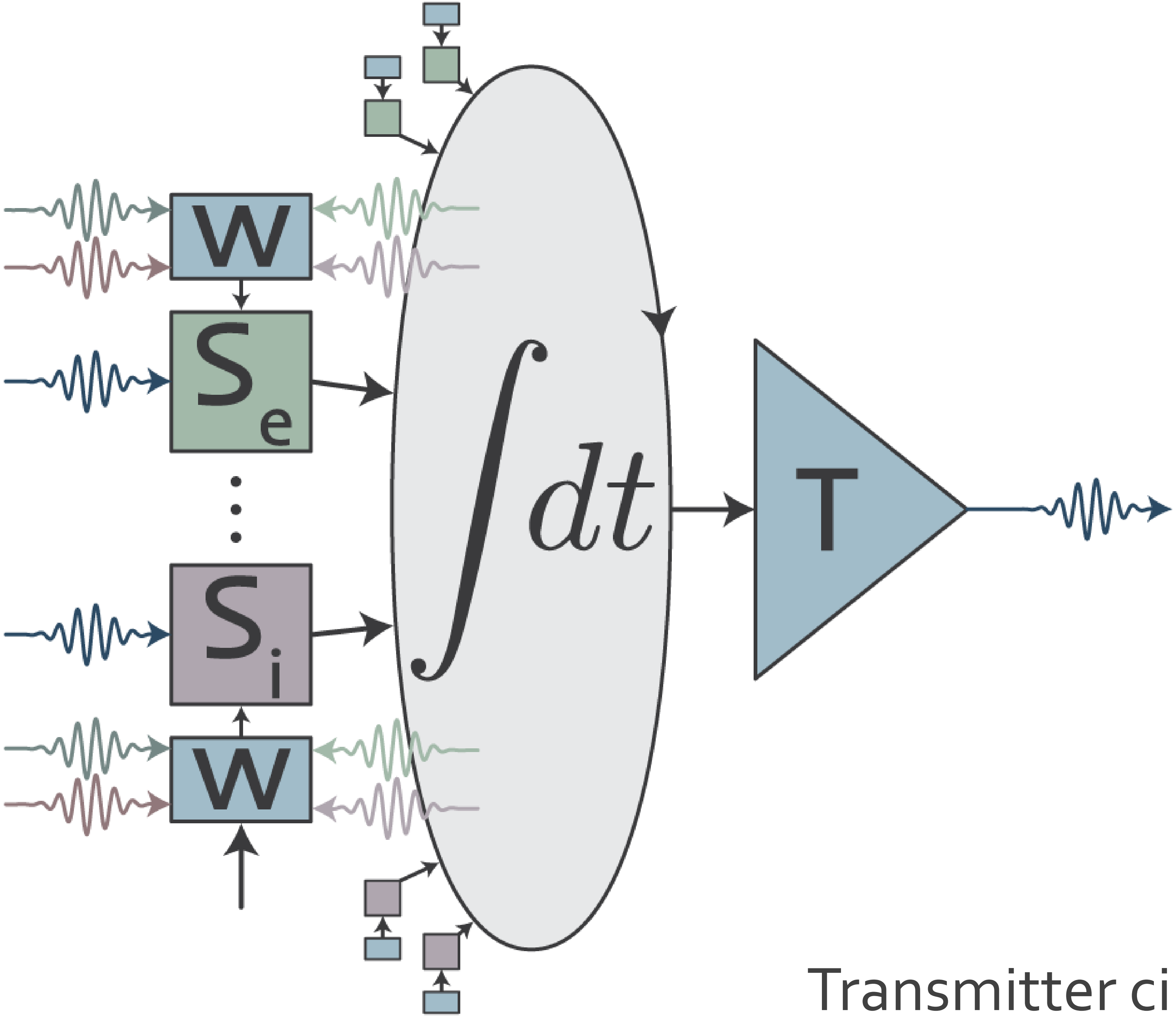
Photons update weight



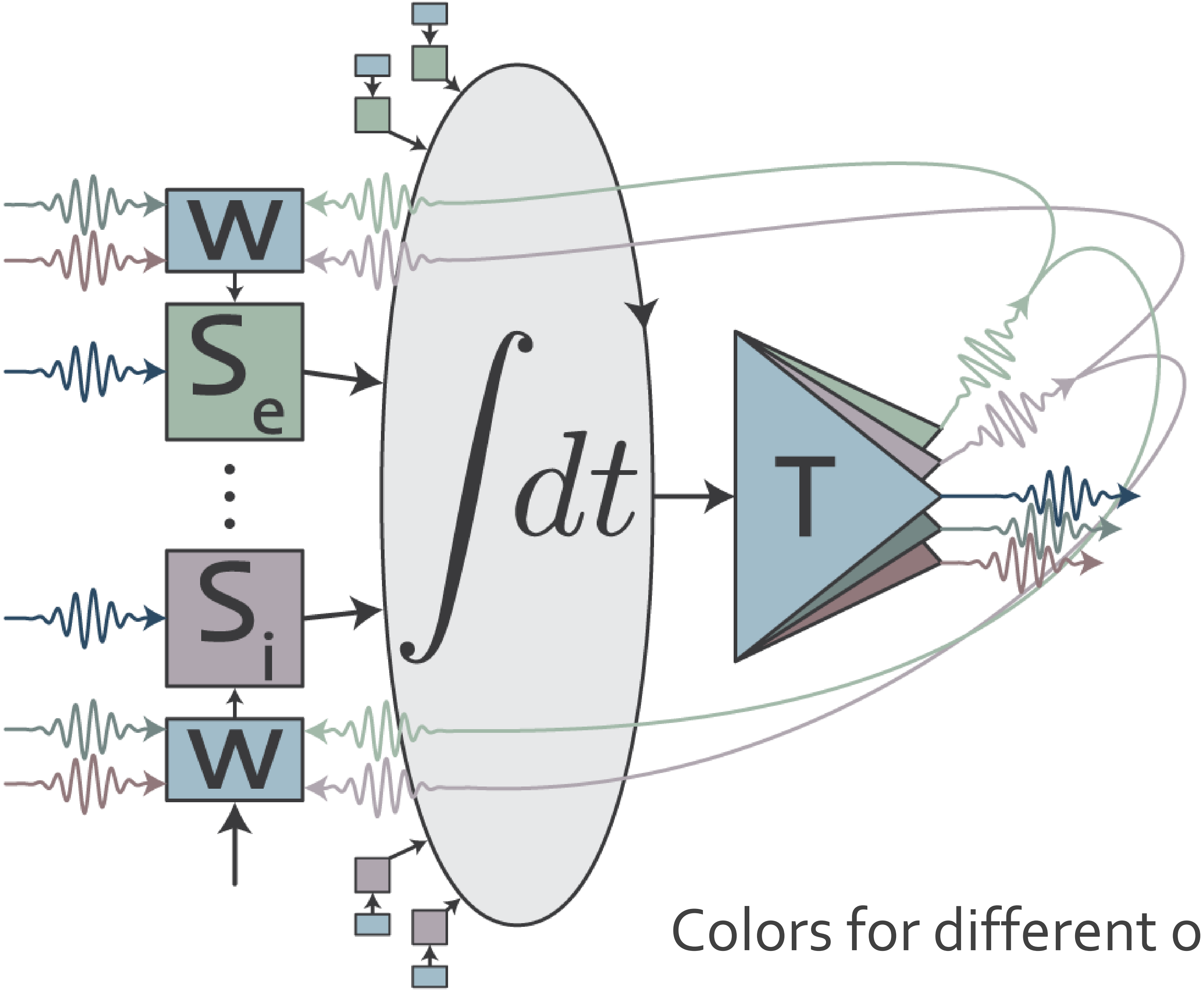
Inhibitory synapses



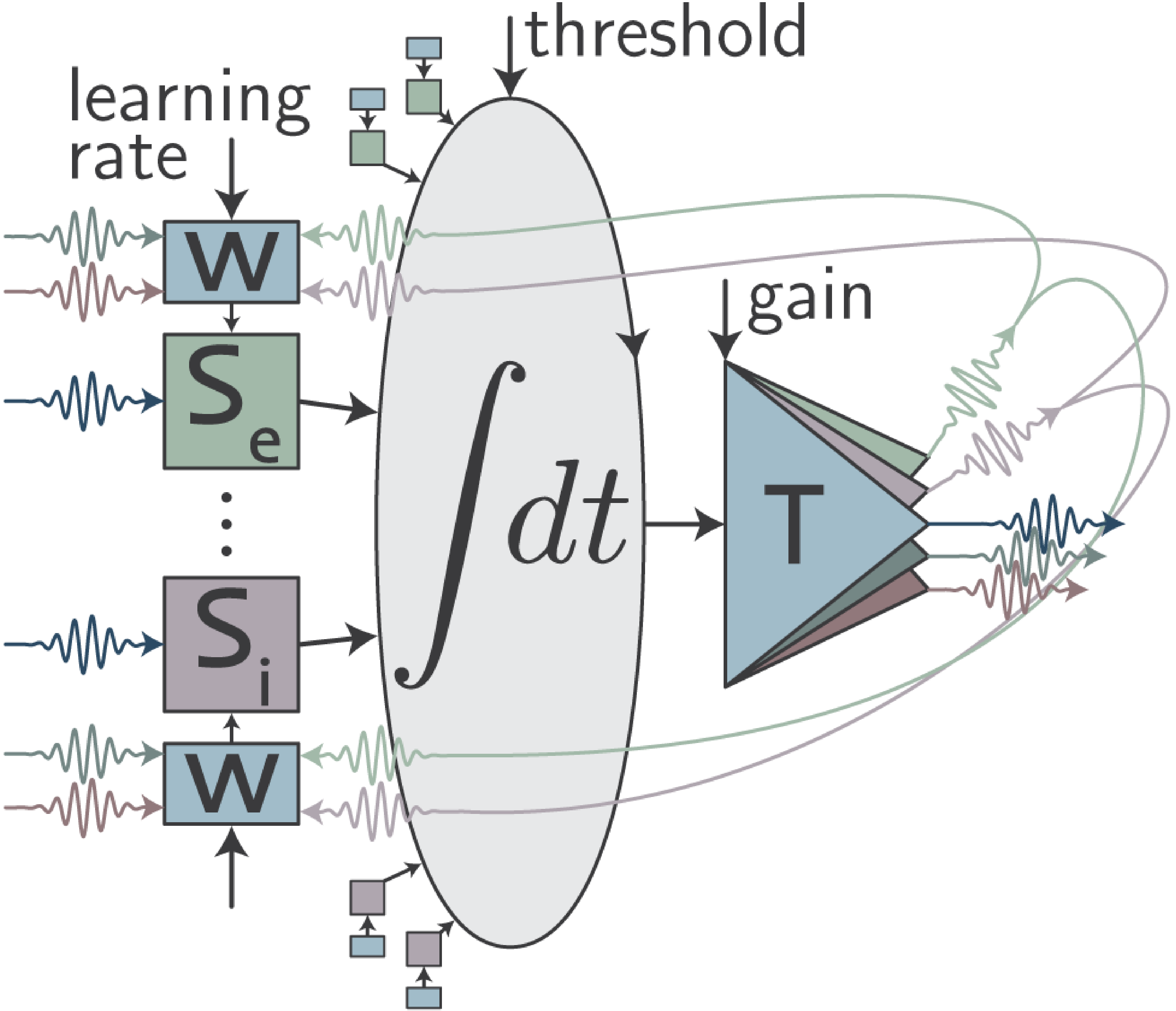
Many synapses



Transcribter circuits

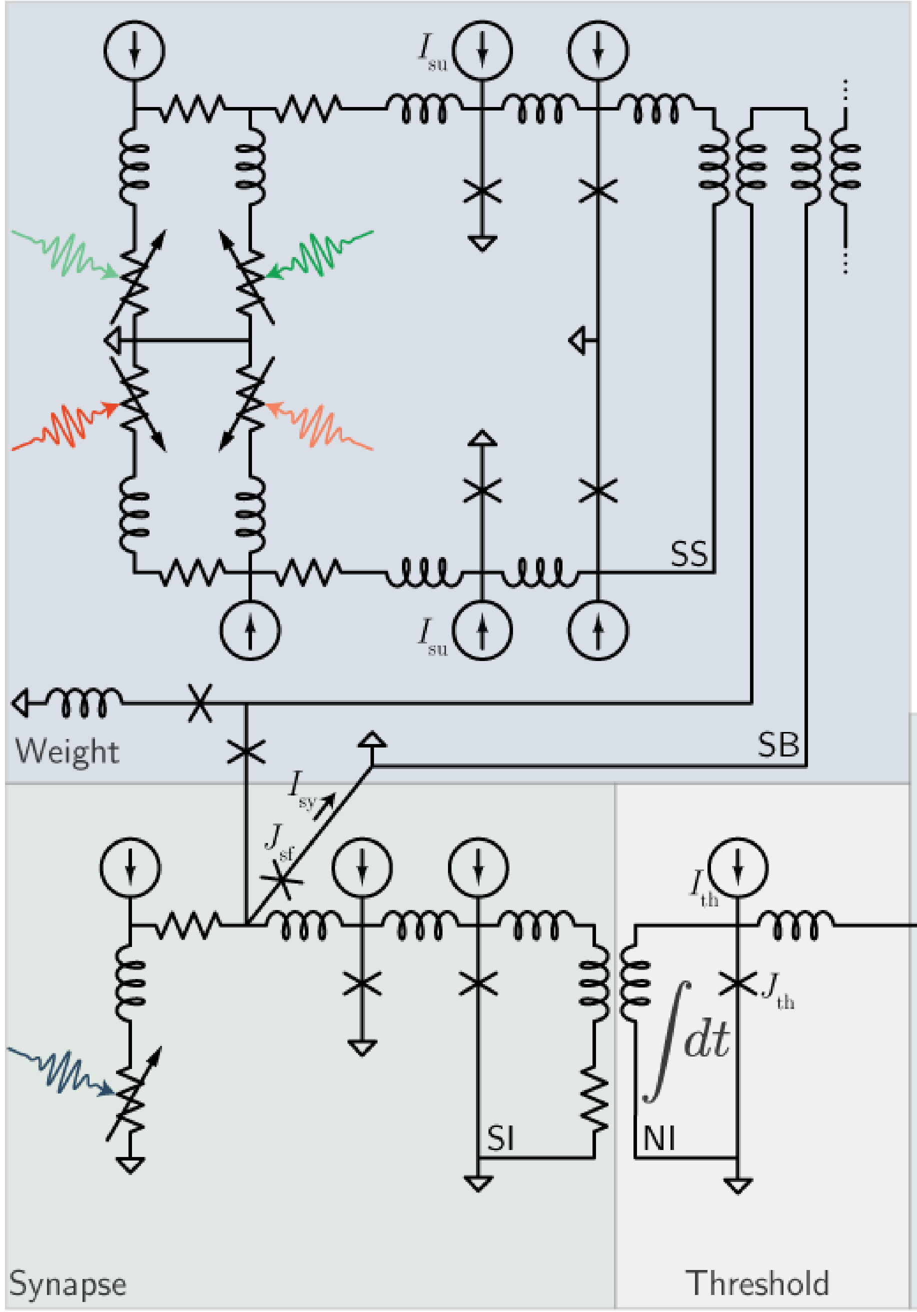


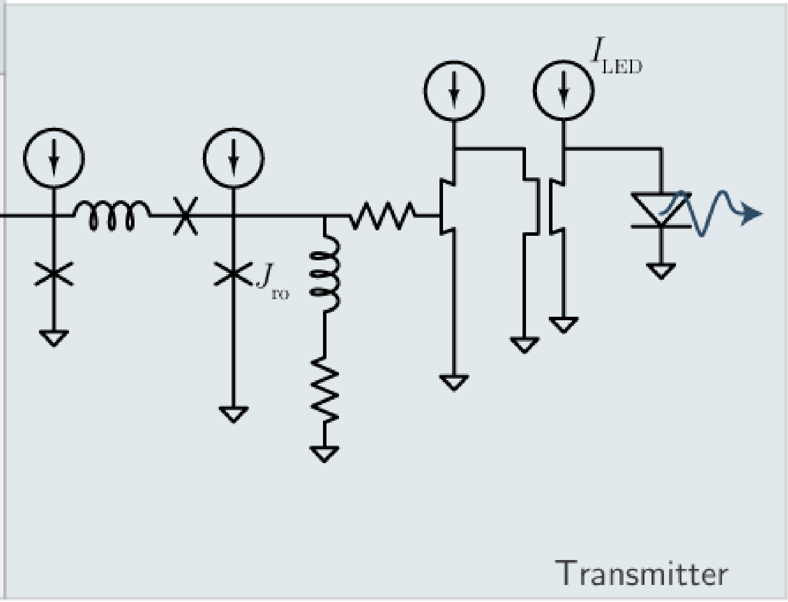
Colors for different operations

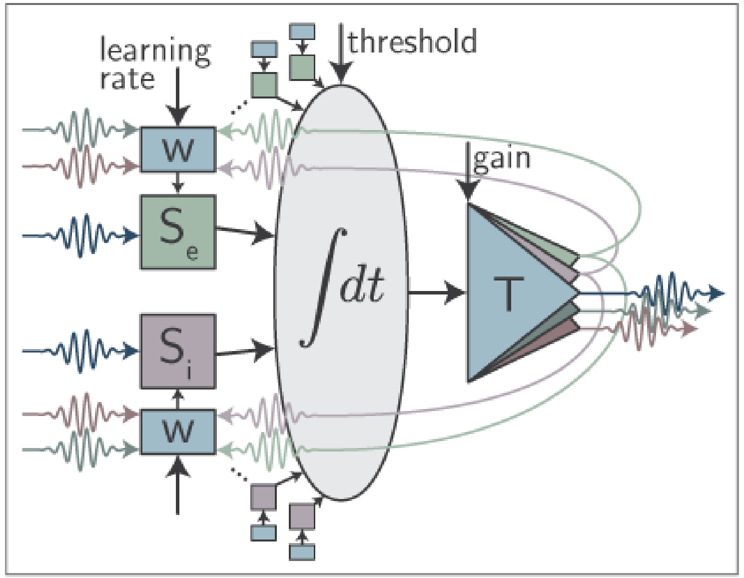

threshold
learning
rate
W
S_e
S_i
W
$\int dt$
gain
T

Loop neuron

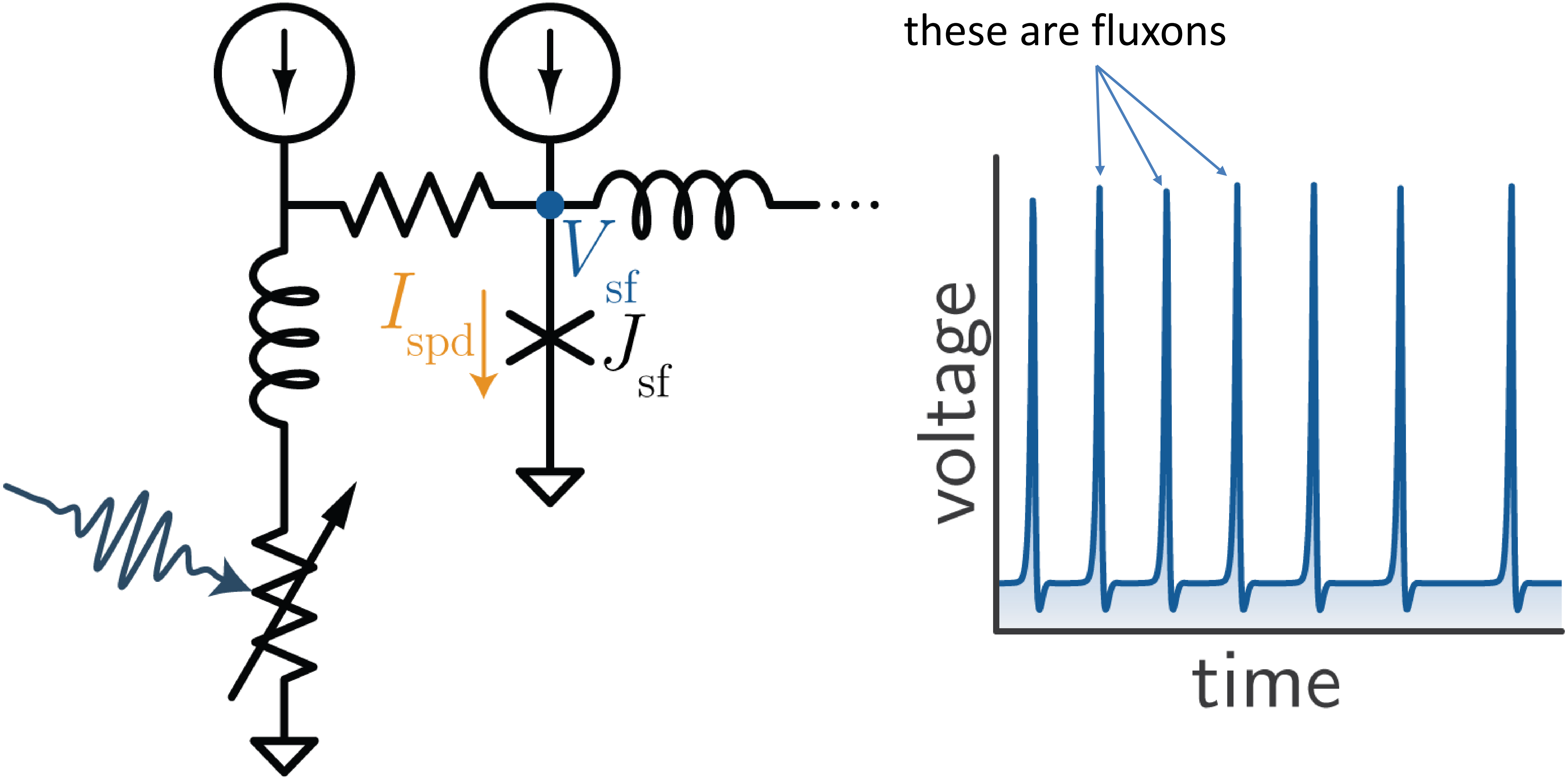
Shainline et al., arXiv:1805.01929 (2018).
Shaineline, arXiv:1904.02807 (2019)



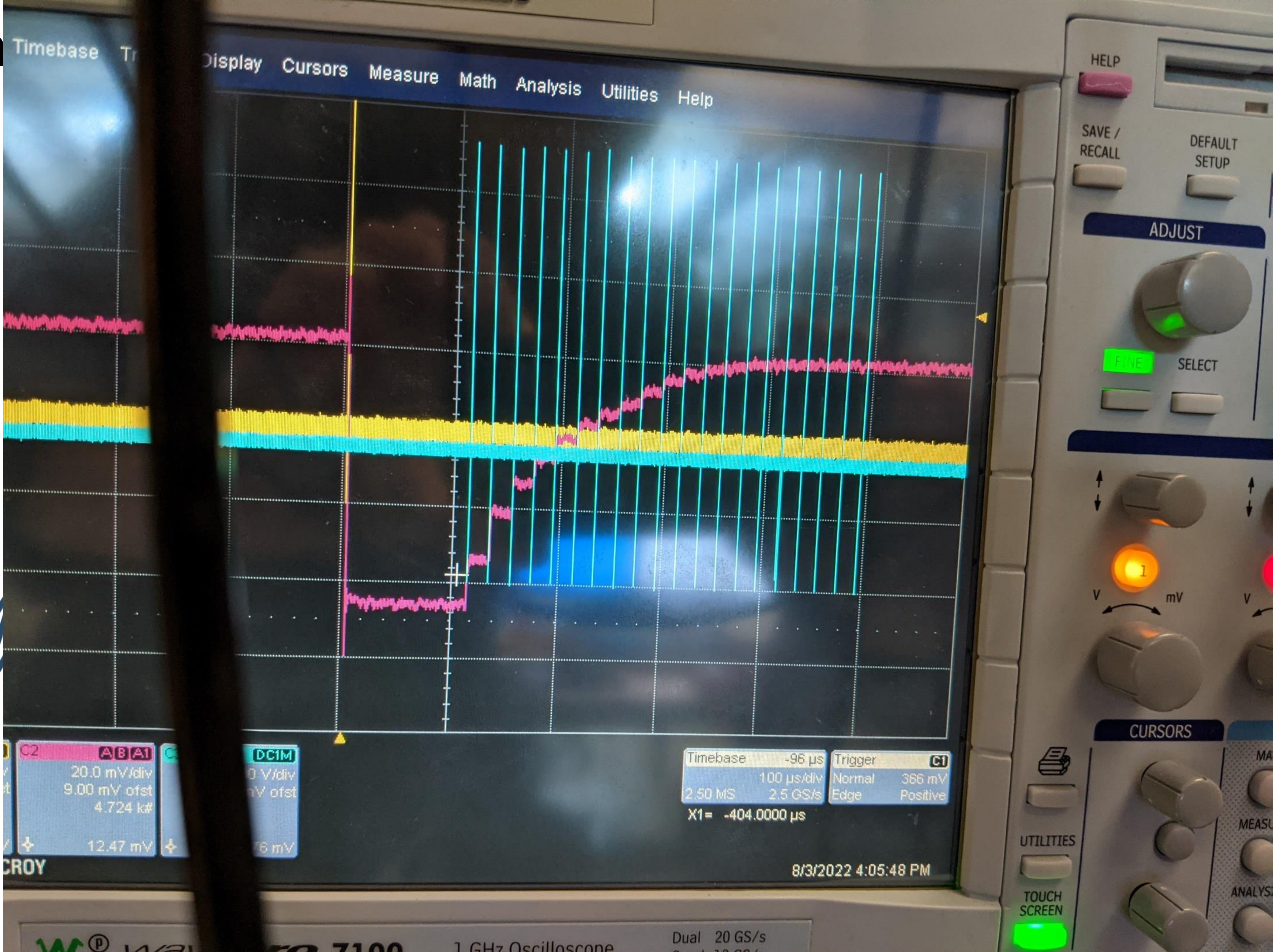




Photon-to-fluxon transducer



The syn


Timebase
Display
Cursors
Measure
Math
Analysis
Utilities
Help
HELP
SAVE / RECALL
DEFAULT SETUP
ADJUST
SELECT
C2
20.0 mV/div
9.00 mV ofst
4.724 k#
12.47 mV
DC1M
Timebase -96 µs
100 µs/div
2.50 MS 2.5 GS/s
Trigger
Normal 366 mV
Edge Positive
X1= -404.0000 µs
8/3/2022 4:05:48 PM
CURSORS
UTILITIES
TOUCH SCREEN
1 GHz Oscilloscope
Dual 20 GS/s

Long-range = photons
Short-range = electronics / SFQ

Pulsed neural networks consisting of single-flux-quantum spiking neurons
T. Hirose, T. Asai, and Y. Amemiya
Physica C, 463:1072, 2007.

PHYSICAL REVIEW E **82**, 011914 (2010)

Josephson junction simulation of neurons

Patrick Crotty,[1] Dan Schult,[2] and Ken Segall[1]

SCIENCE ADVANCES | RESEARCH ARTICLE

APPLIED SCIENCES AND ENGINEERING

Ultralow power artificial synapses using nanotextured magnetic Josephson junctions

Michael L. Schneider,* Christine A. Donnelly, Stephen E. Russek, Burm Baek, Matthew R. Pufall, Peter F. Hopkins, Paul D. Dresselhaus, Samuel P. Benz, William H. Rippard

https://arxiv.org/abs/1907.00263

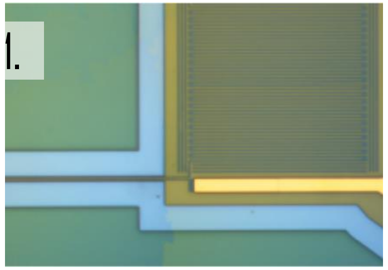
A Power Efficient Artificial Neuron Using Superconducting Nanowires

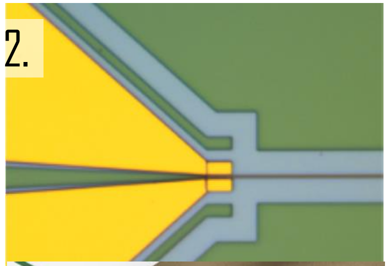
Emily Toomey, Ken Segall, Karl Berggren

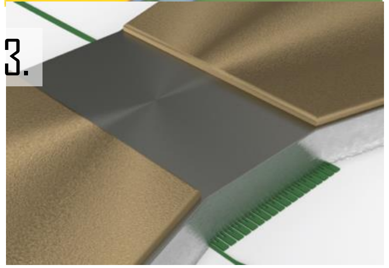
SOEN technical approach

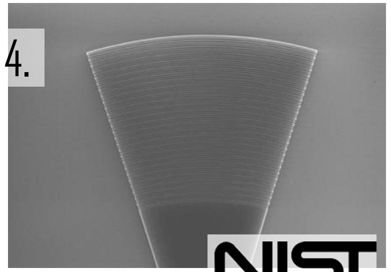
superconducting optoelectronic networks

1. Single photons for minimum spike energy
   (superconducting single-photon detectors)
2. Cold optoelectronics for monolithic integration
   (all-silicon light-emitting diodes)
3. Cold electronics (SFQ) for nonlinear processing
   (Josephson junctions, cryotron switches)
4. Light for interconnects
   (silicon photonics)





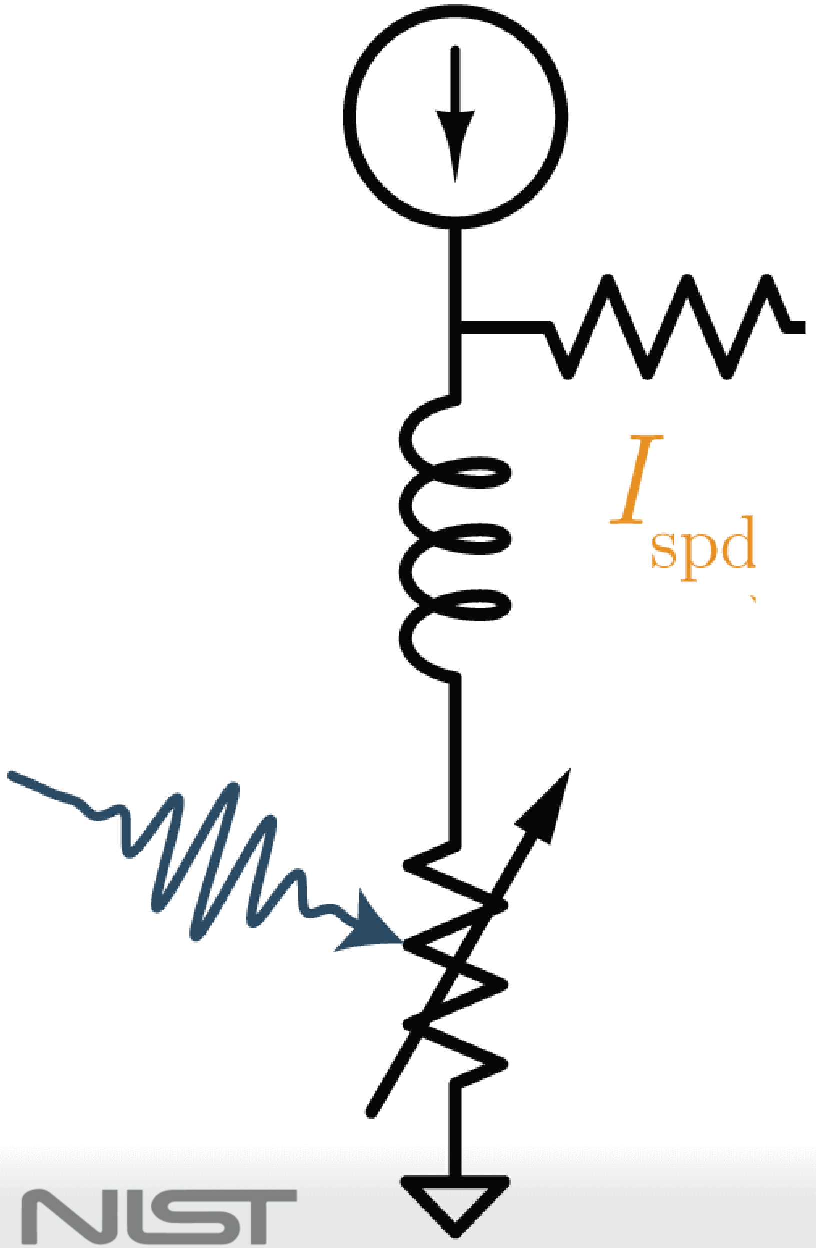


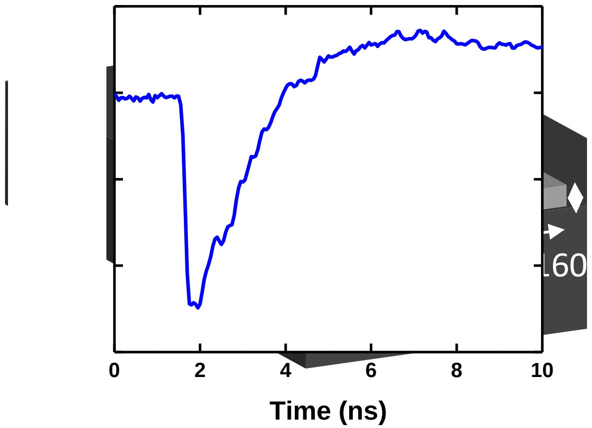


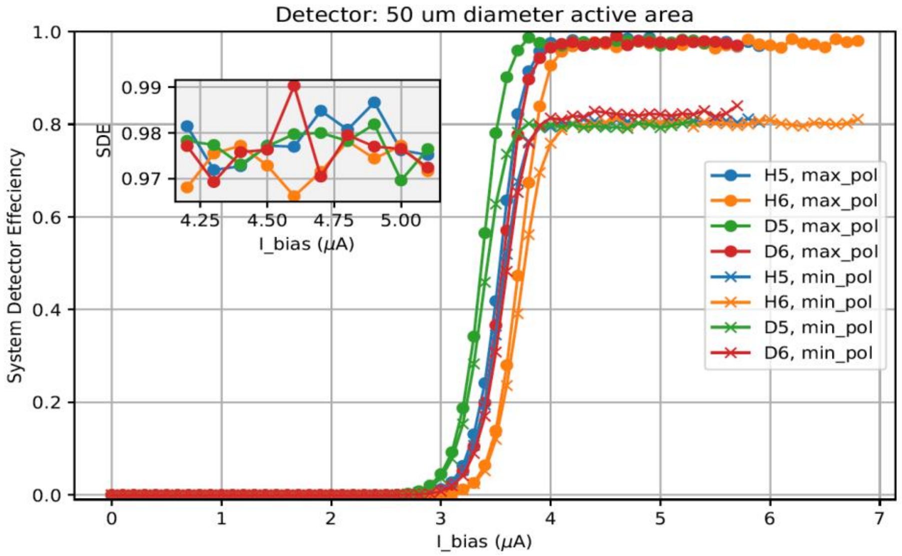
J. M. Shainline, S. M. Buckley, R. P. Mirin, and S. W. Nam, “Superconducting optoelectronic circuits for neuromorphic computing,” Phys. Rev. Applied, Mar 2017.

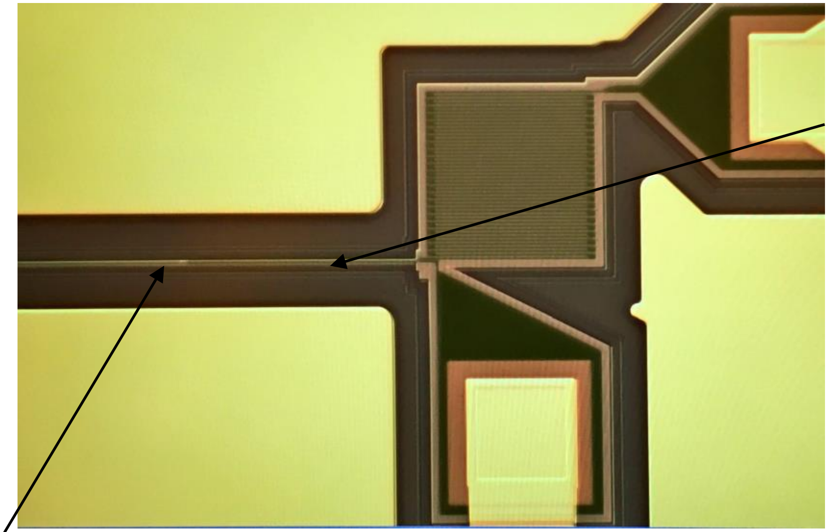
cryogenic silicon photonics platform: recent results

Superconducting Nanowire Detector







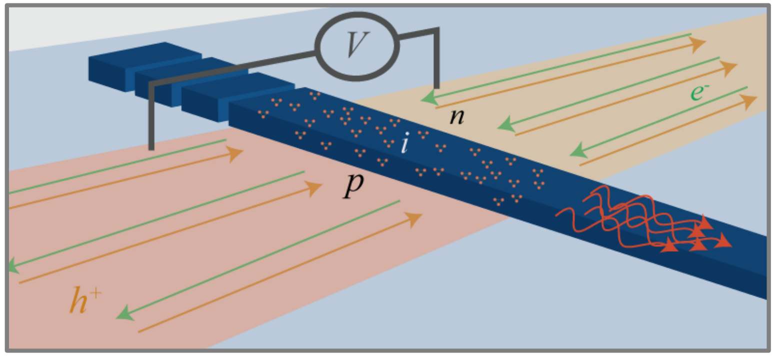


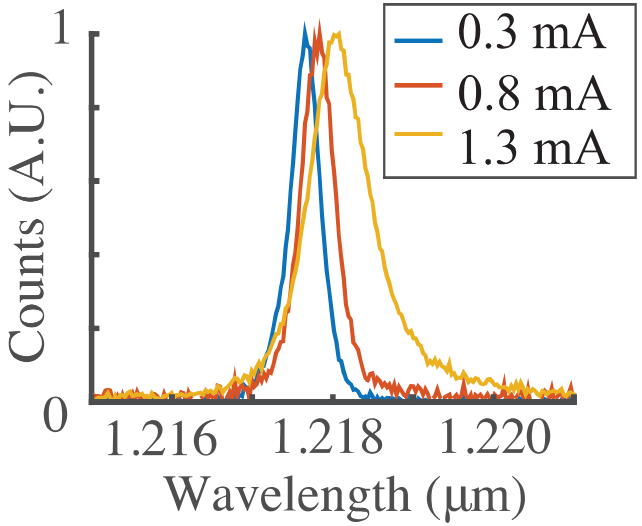
Nanowire on waveguide

Waveguide

All-silicon light emitting diodes

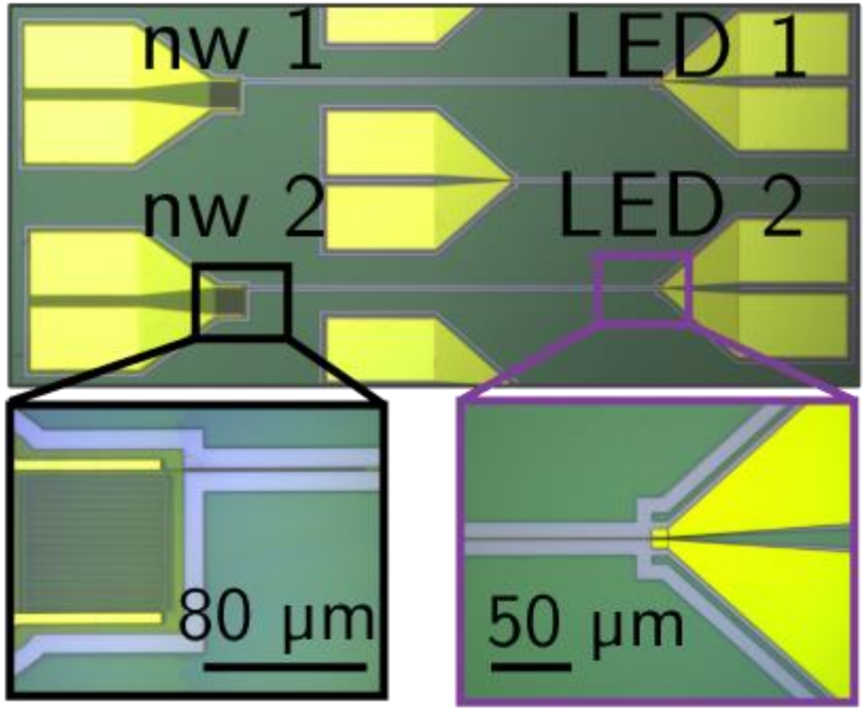
- Si defect centers have optical transitions
- Low temp. inhibits non-radiative pathways
- Electrical pumping with PN junction
- W-centers: 1220nm emission

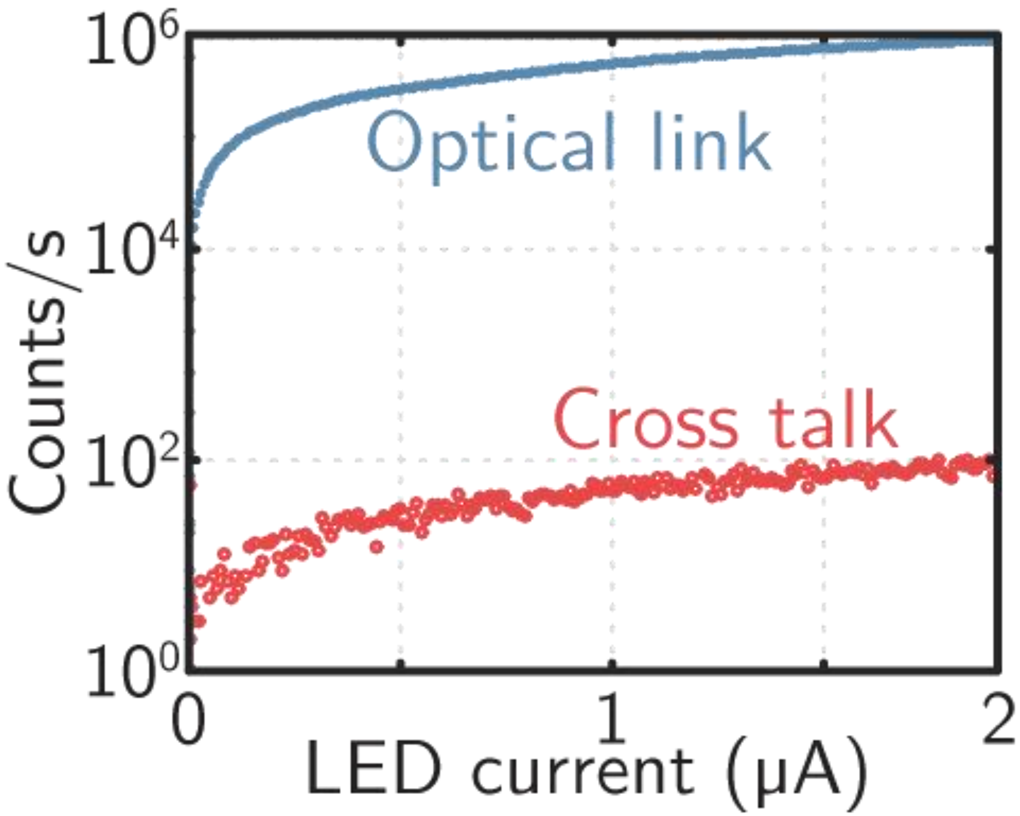




S. M. Buckley et al. "All-silicon light-emitting diodes waveguide-integrated..." APL, 2017.

Co-integration: Silicon LEDs + SNSPDs

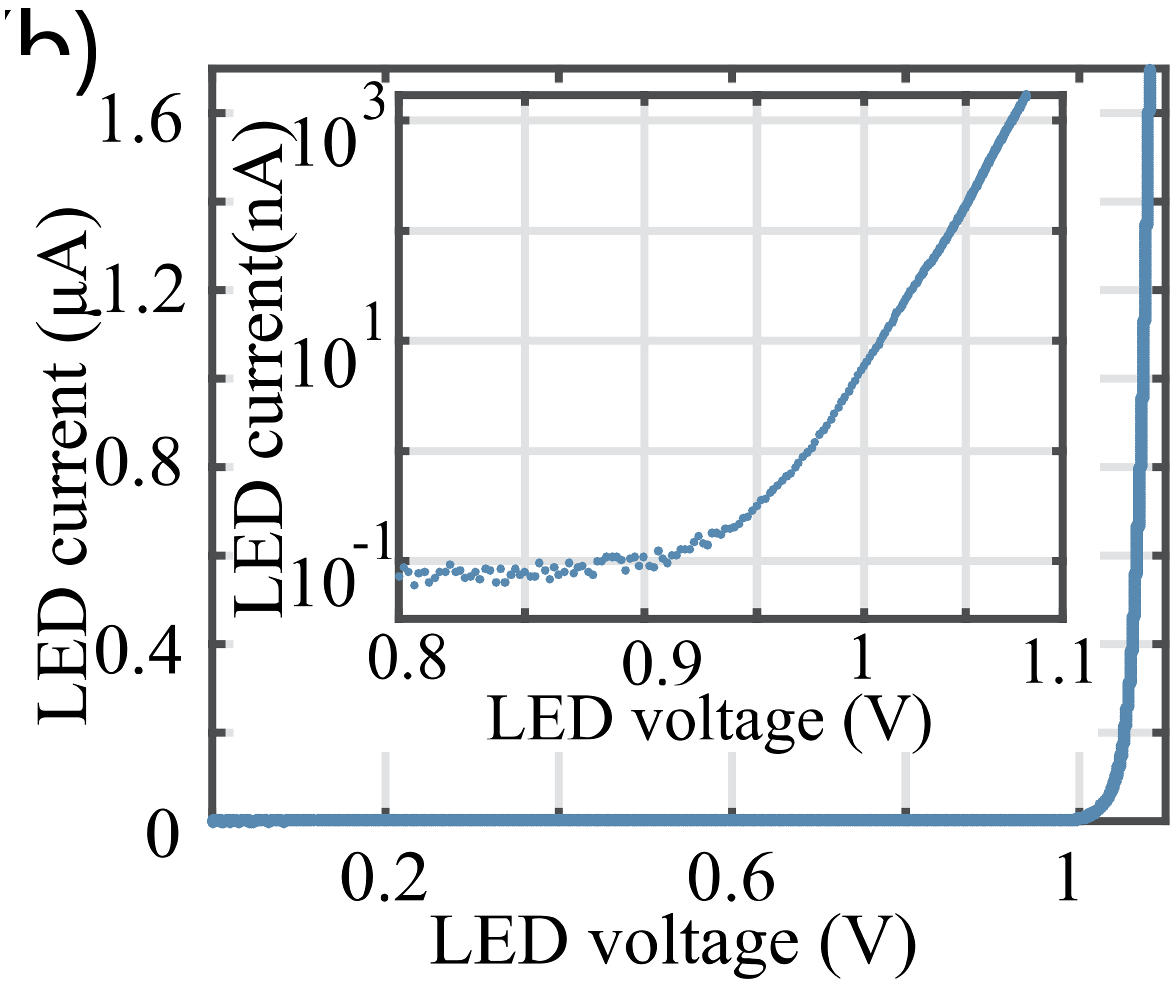




S. M. Buckley et al. "All-silicon light-emitting diodes waveguide-integrated…" APL, 2017.

All-silicon light emitting diodes

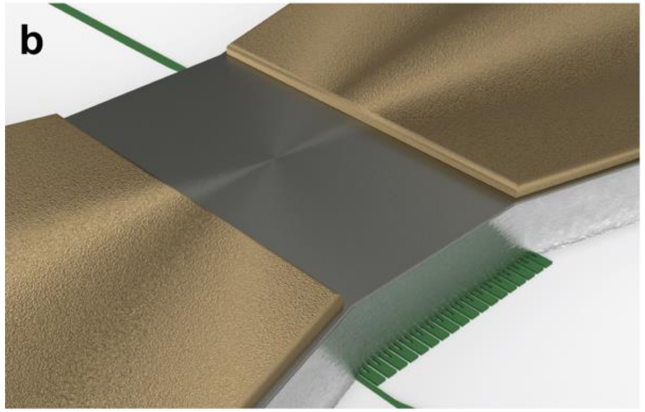
- LED turn-on voltage is ~ 1 V
- SNSPD can output ~ 1 mV

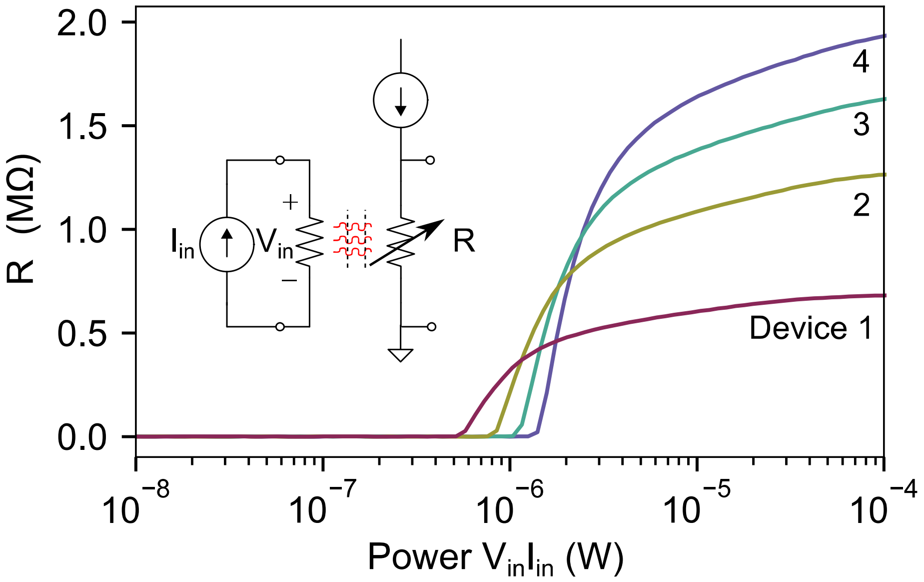


S. M. Buckley et al. "All-silicon light-emitting diodes waveguide-integrated…" APL, 2017.

Nano-cryotron thermal switch

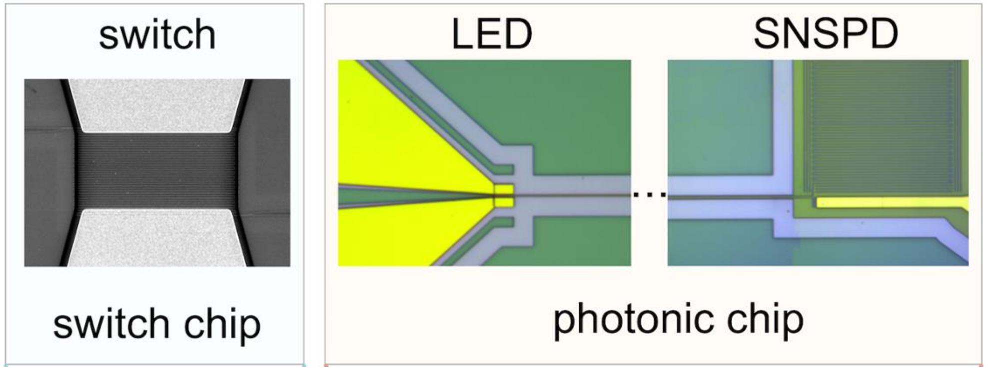
- millivolt input -> Volt output
- No Josephson Junctions needed
- Reset about 10ns

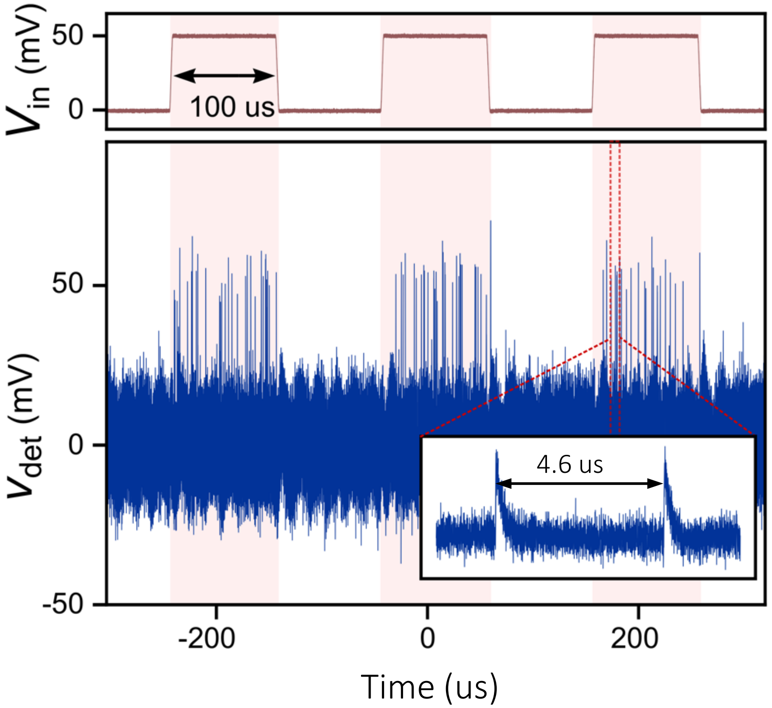




A. McCaughan et al. "A compact, ultrahigh impedance superconducting thermal switch for interfacing superconductors with semiconductors and optoelectronics," arXiv:1903.10461, 2019.

Superconductor-to-semiconductor interface

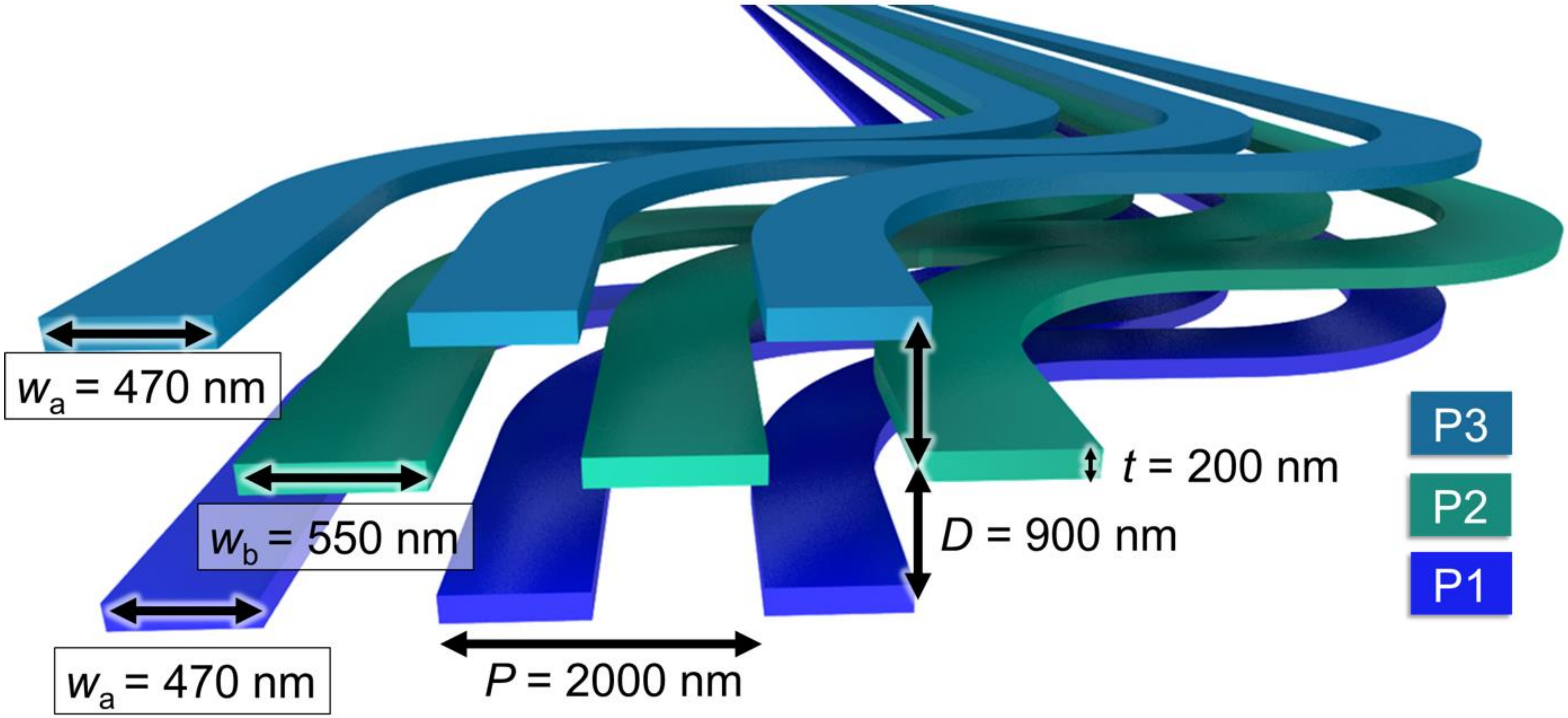




A. McCaughan et al. "A compact, ultrahigh impedance superconducting thermal switch for interfacing superconductors with semiconductors and optoelectronics," arXiv:1903.10461, 2019.

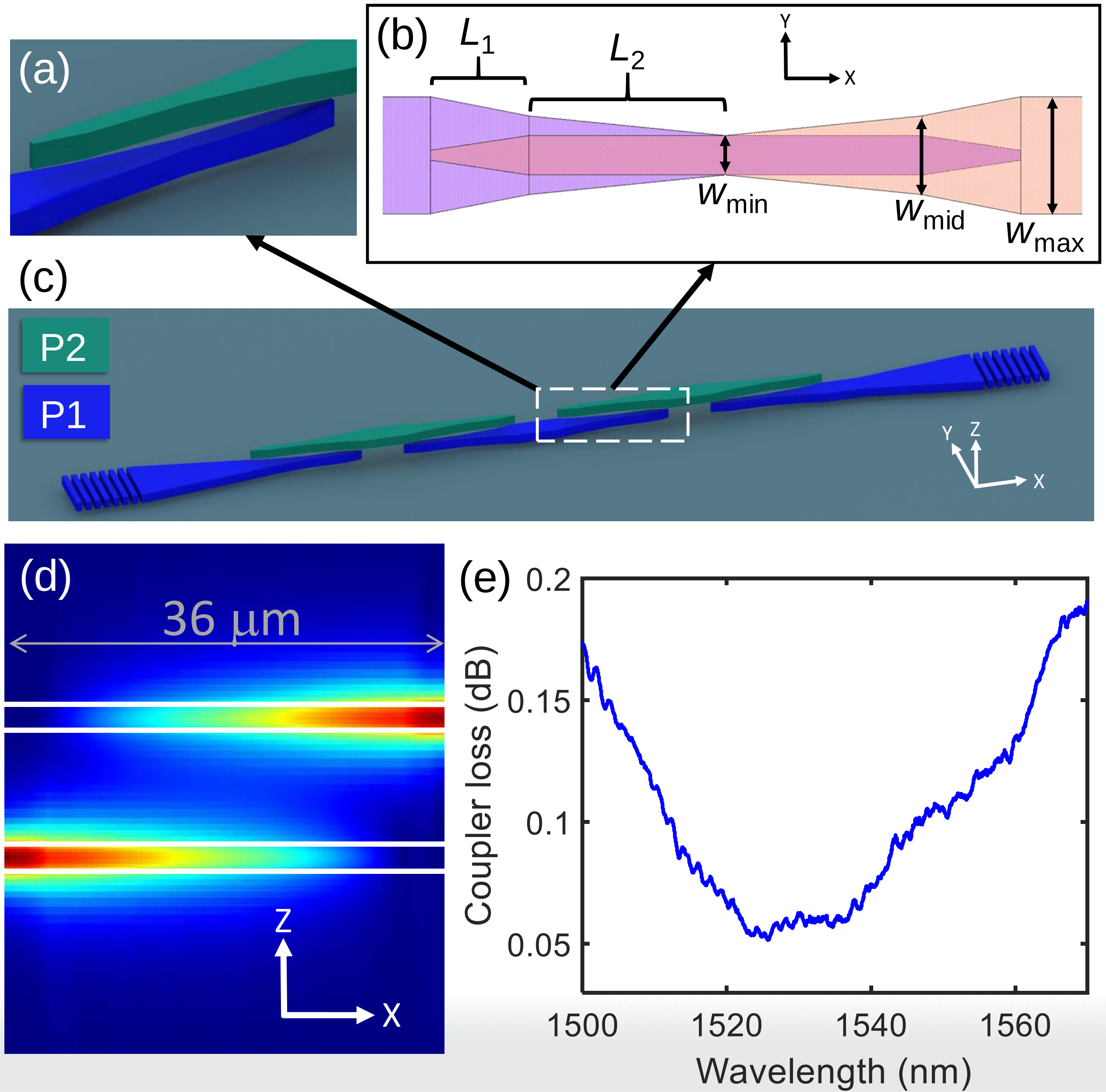
First steps toward photonic networks

Three planes of amorphous silicon waveguides



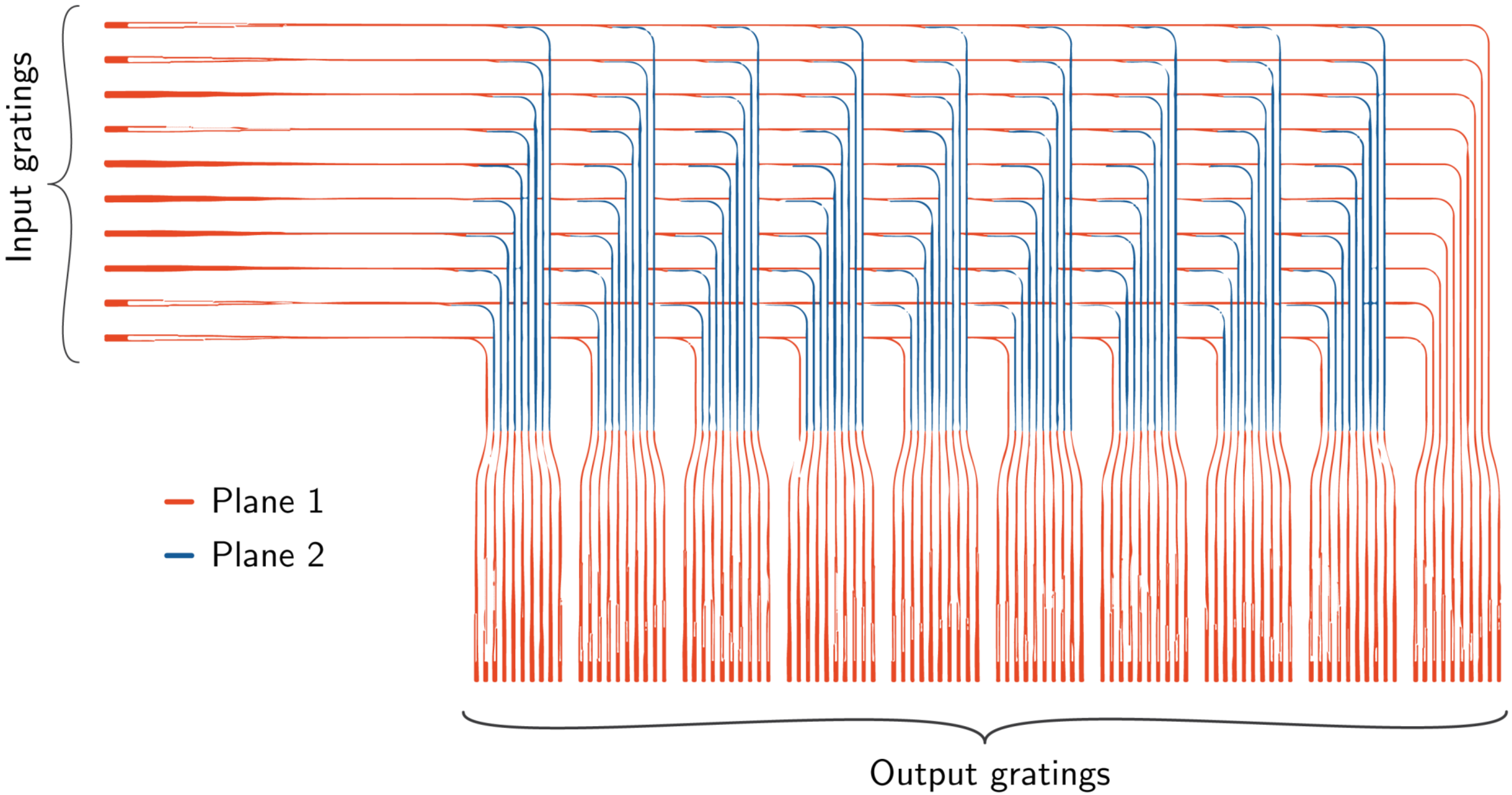
Chiles et al., APL Photonics **2**, 116101 (2017)

Inter-planar couplers

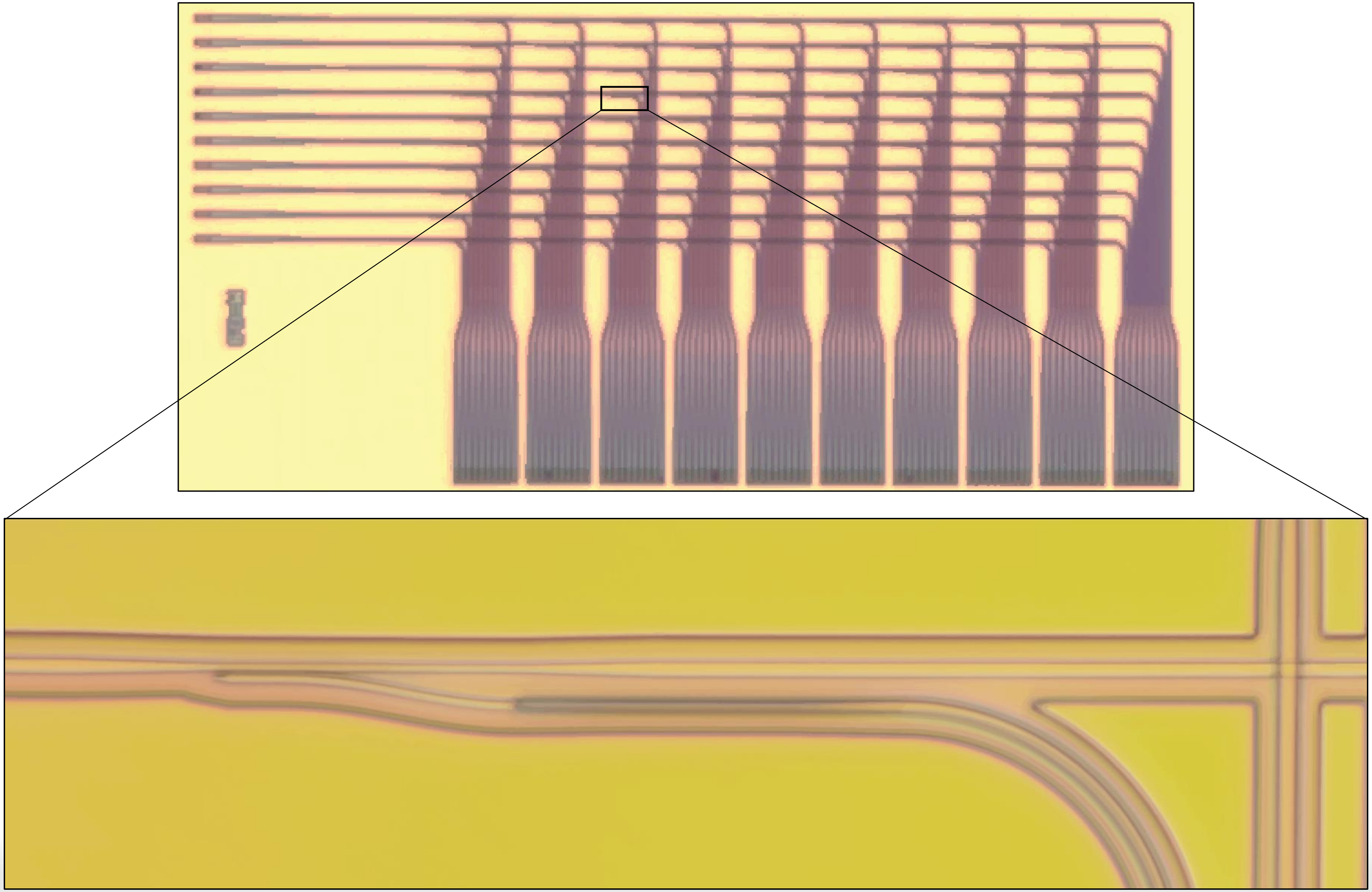


- 0.05 dB loss
- 36 μm length

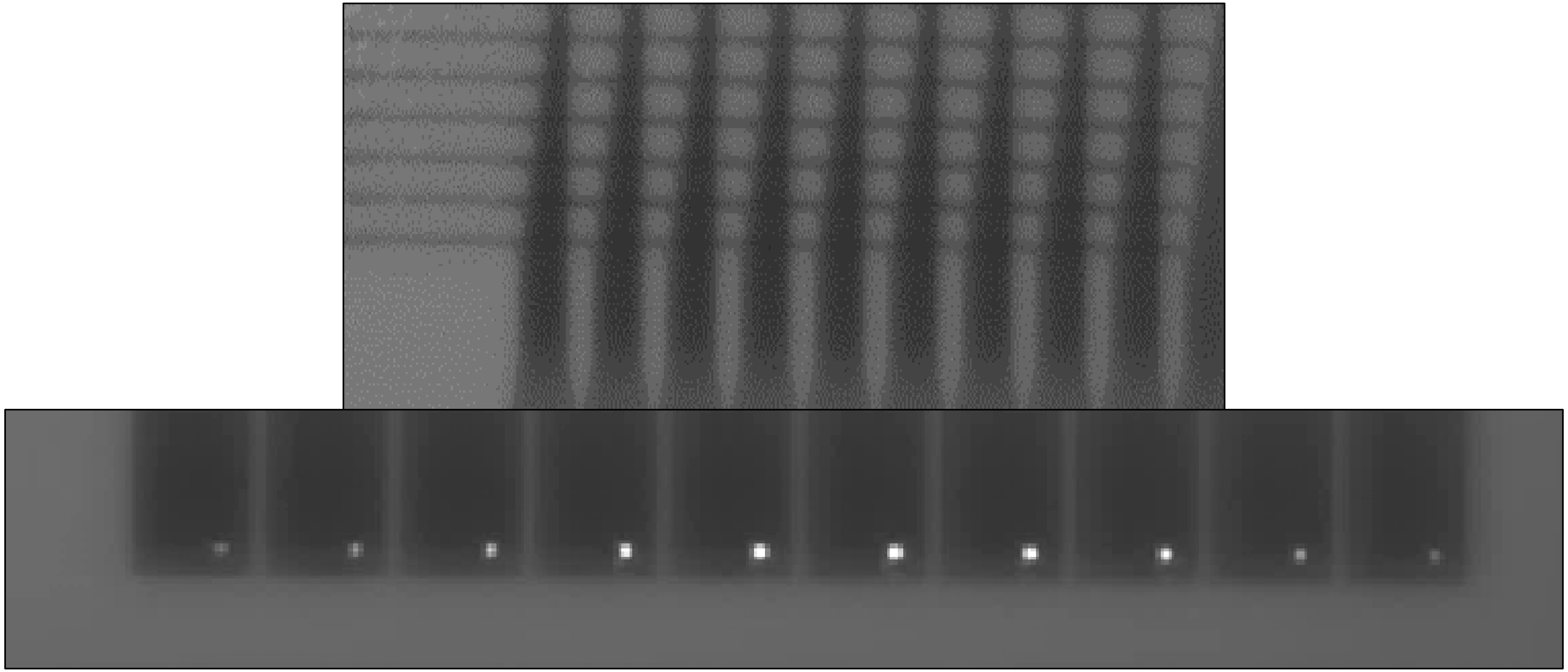
Chiles et al, APL Photonics **2** 116101 (2017).



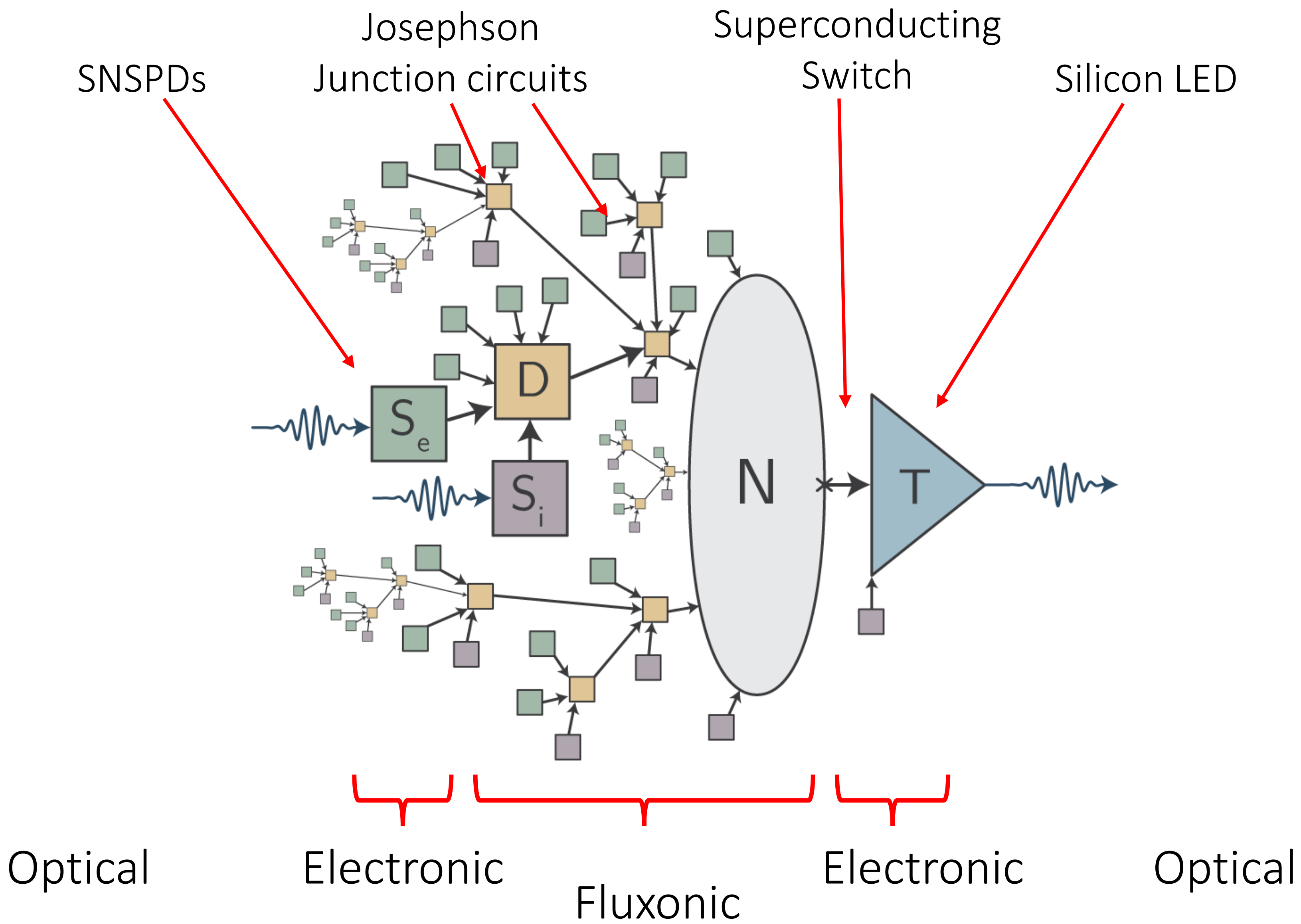
Chiles et al, APL Photonics **3** 106101 (2018).



10 x 100 routing couplers

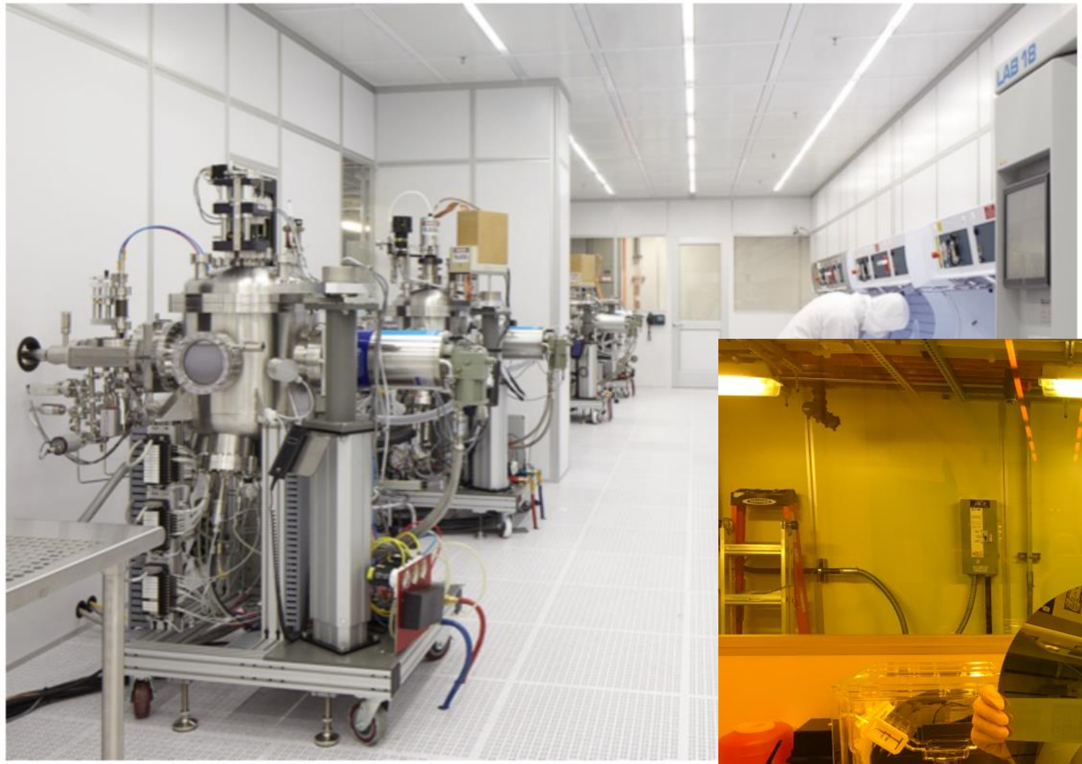


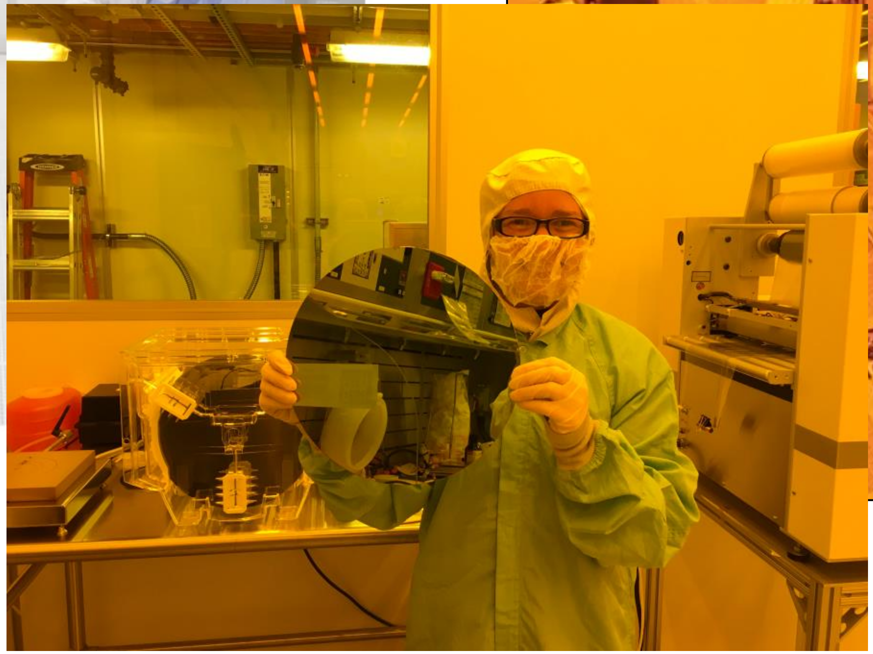
Full Superconducting Optoelectronic Neuron



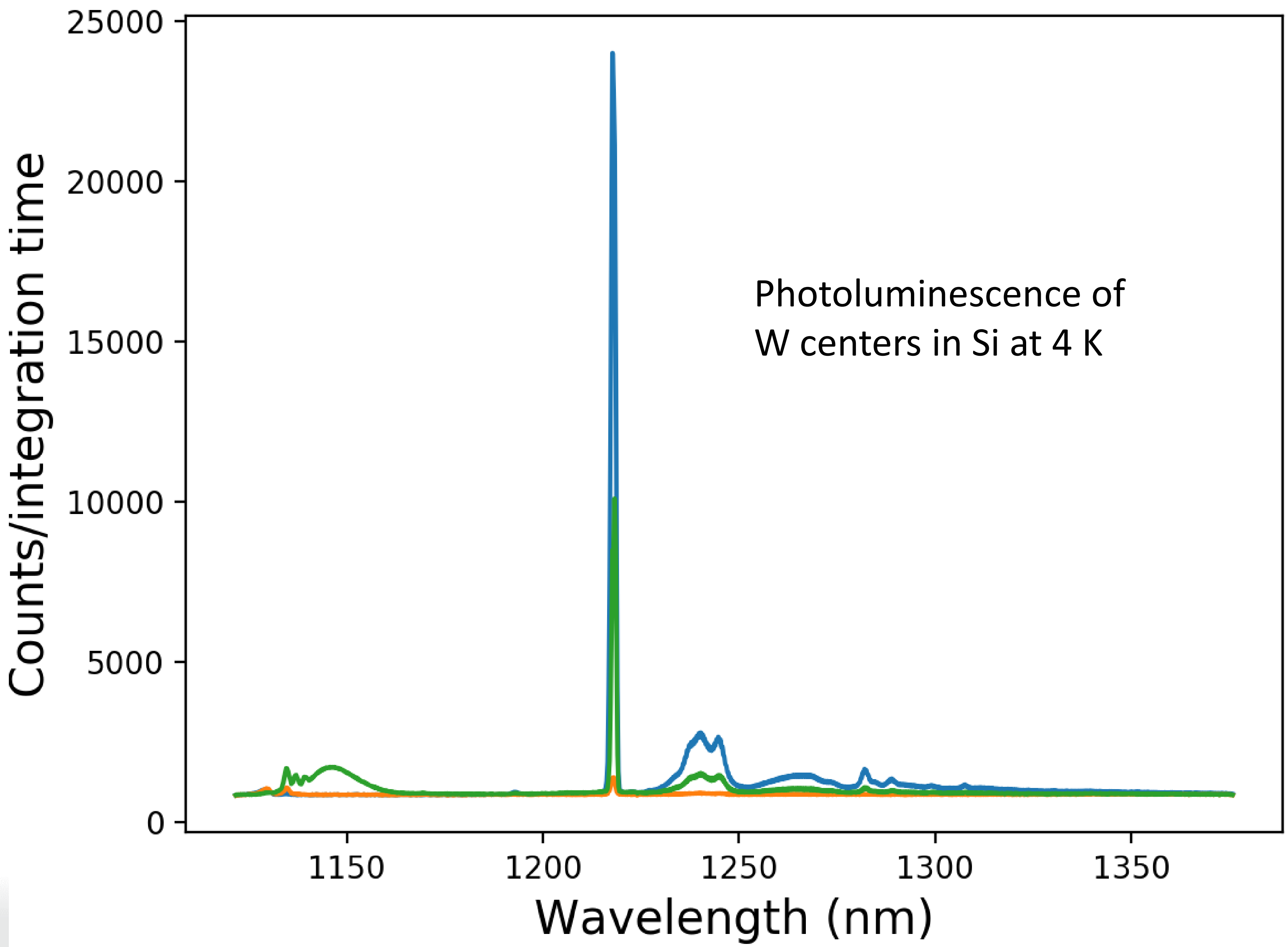
J. M. Shainline, “Fluxonic processing of photonic synapse events,” arXiv:1904.02807, 2019.

Eventual transfer to 300 mm foundry



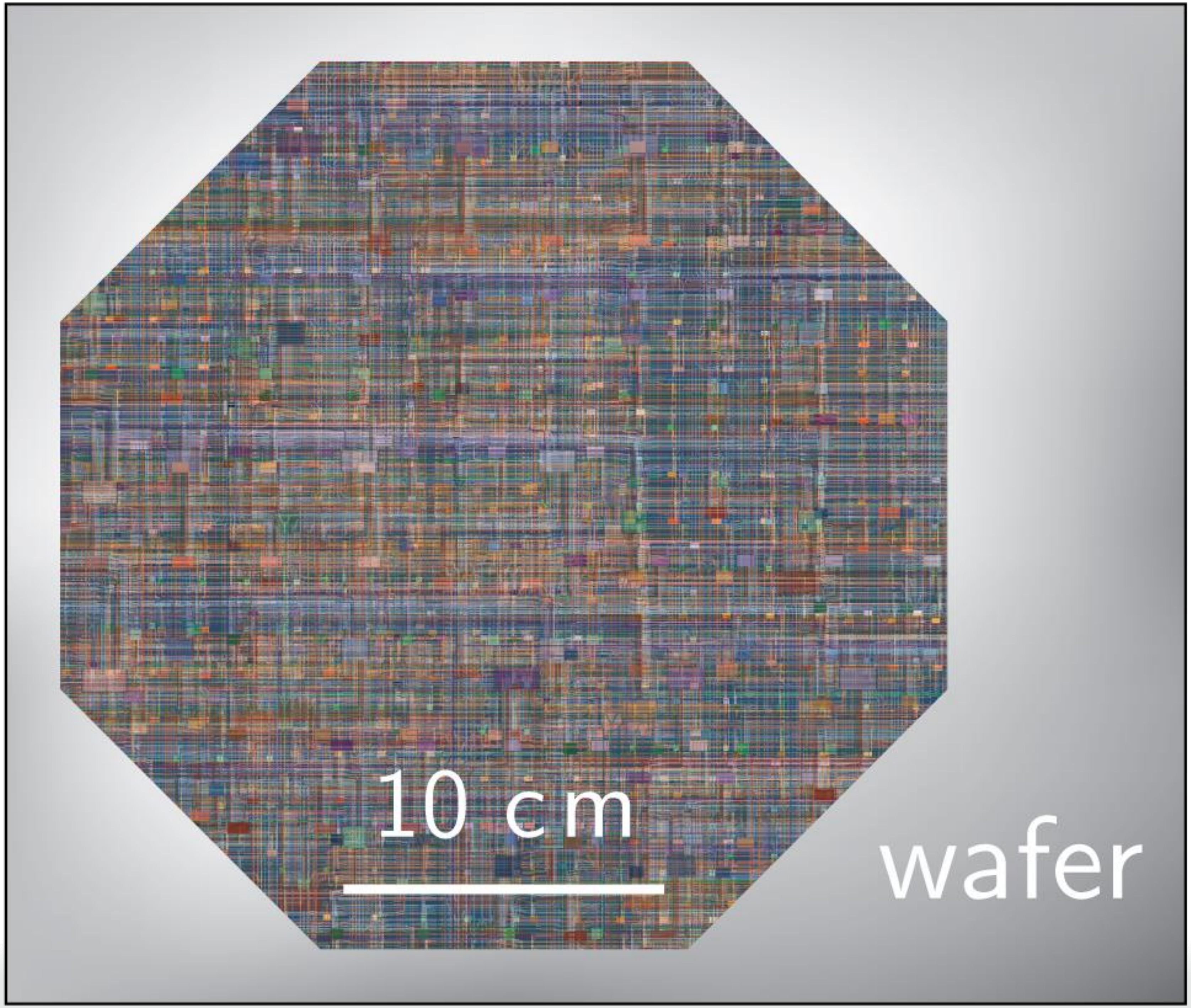


Light sources on 300-mm wafers

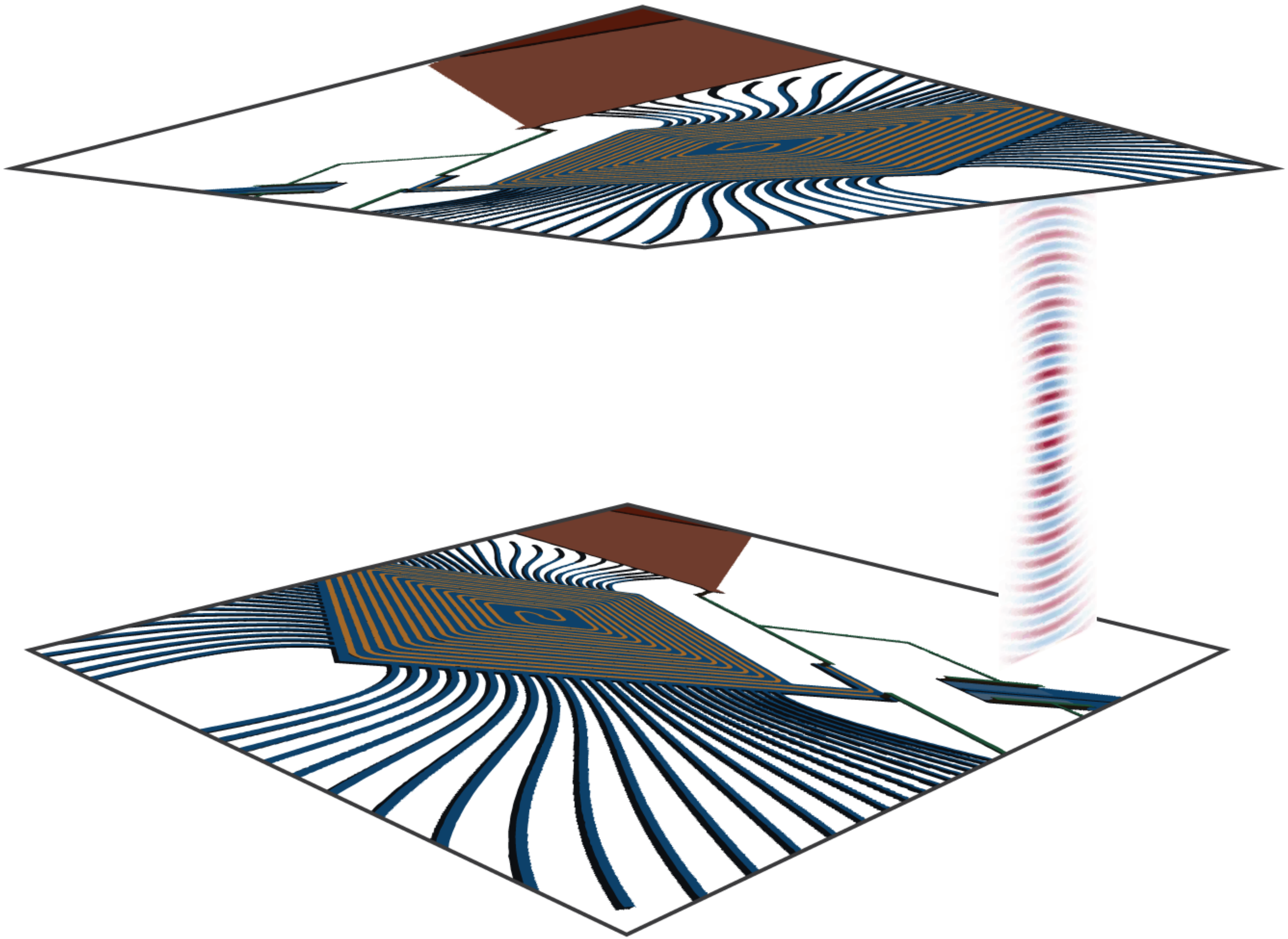


- Fabricated at SUNY Poly by Pops Papa Rao and team
- Measured at NIST by our team

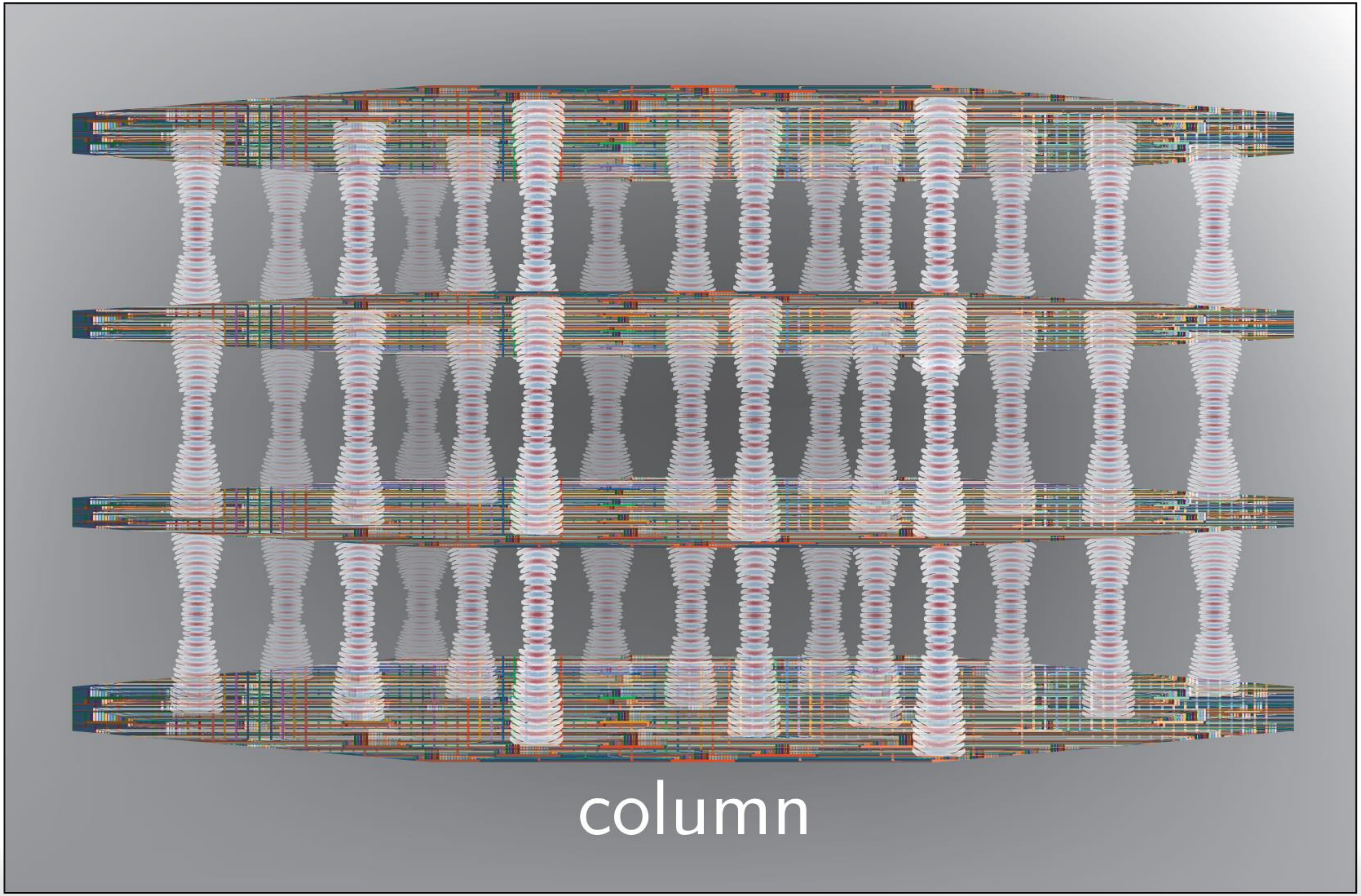
Wafer-scale modules



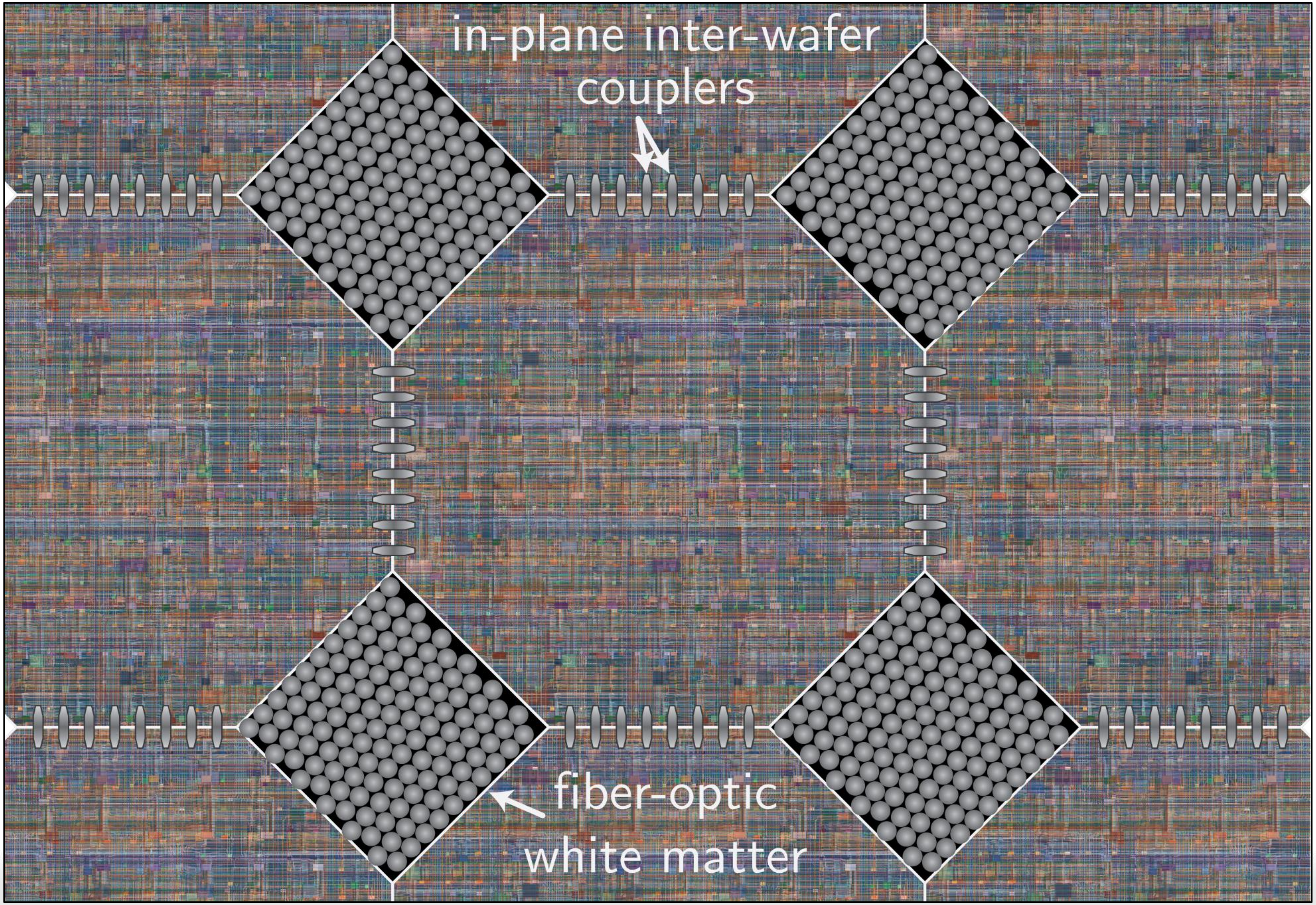
Free-space interconnects for 3D



Free-space inter-wafer interconnects

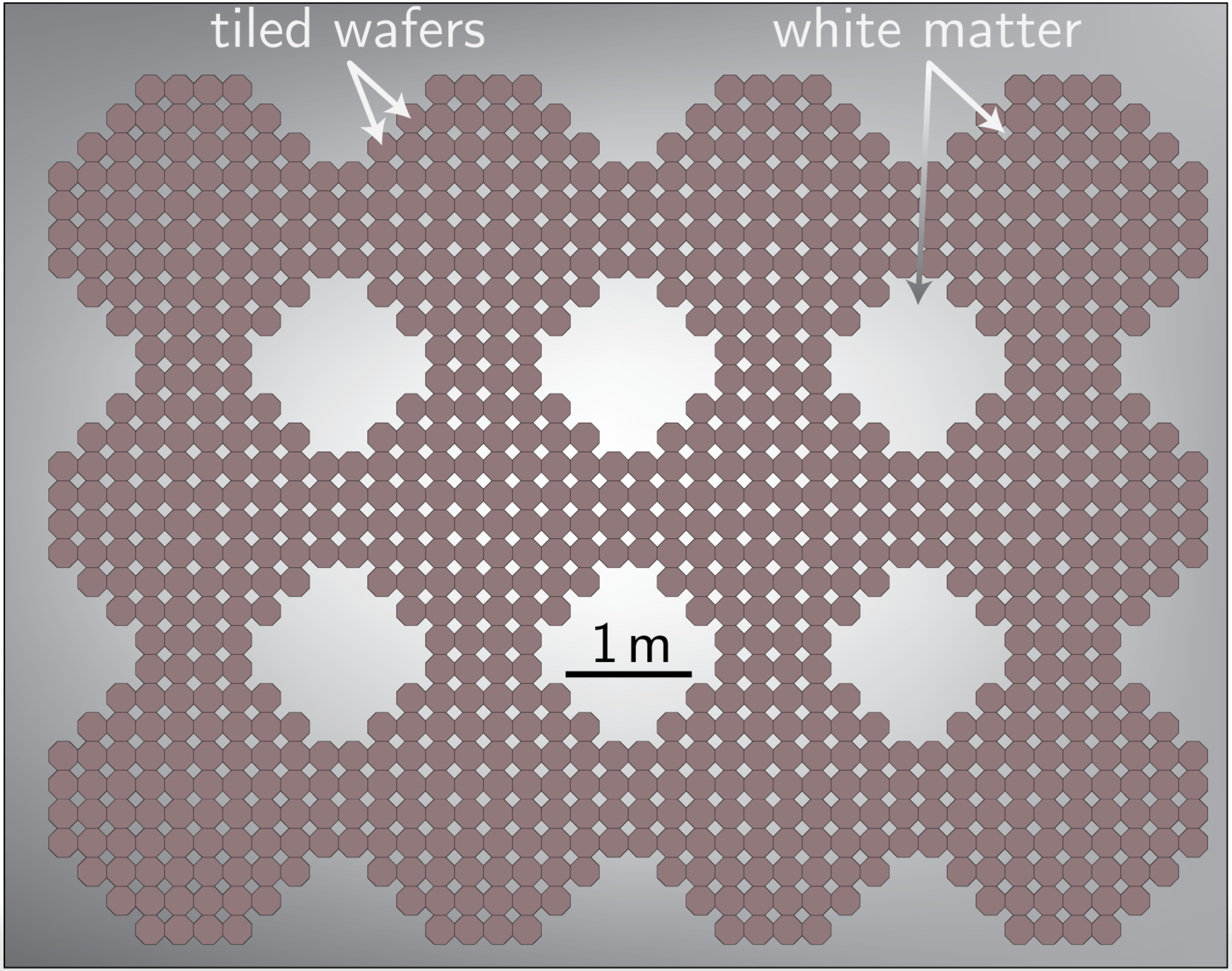


Multi-wafer modules

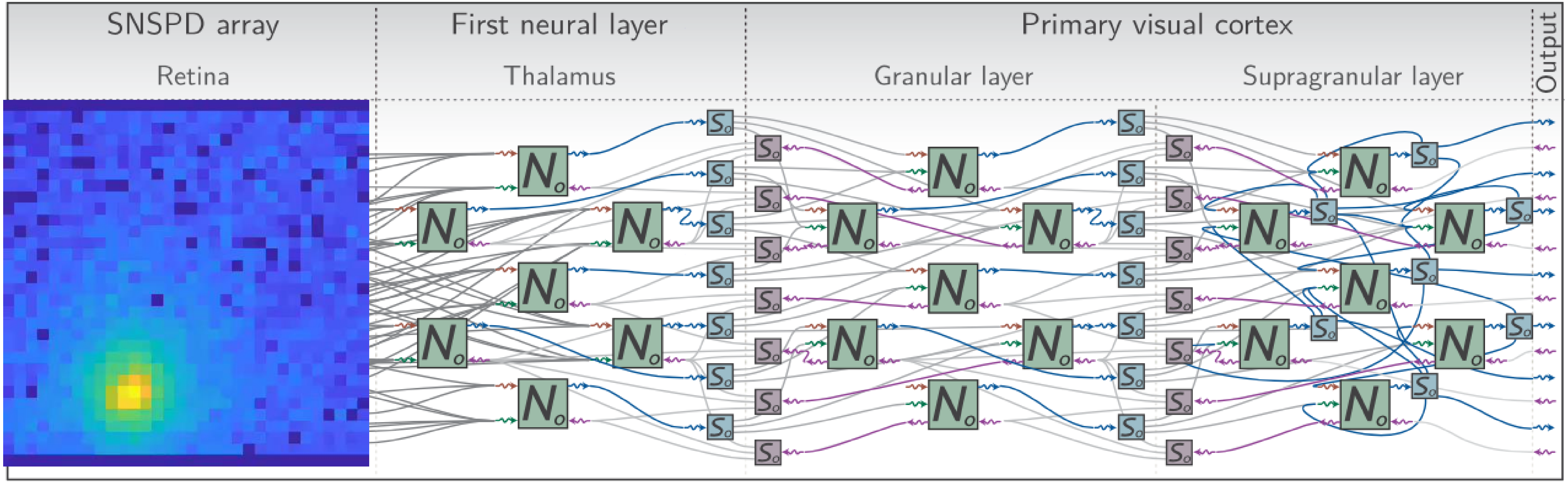


Brain-scale systems?

> 100 Billion neurons



What can we do in five years?

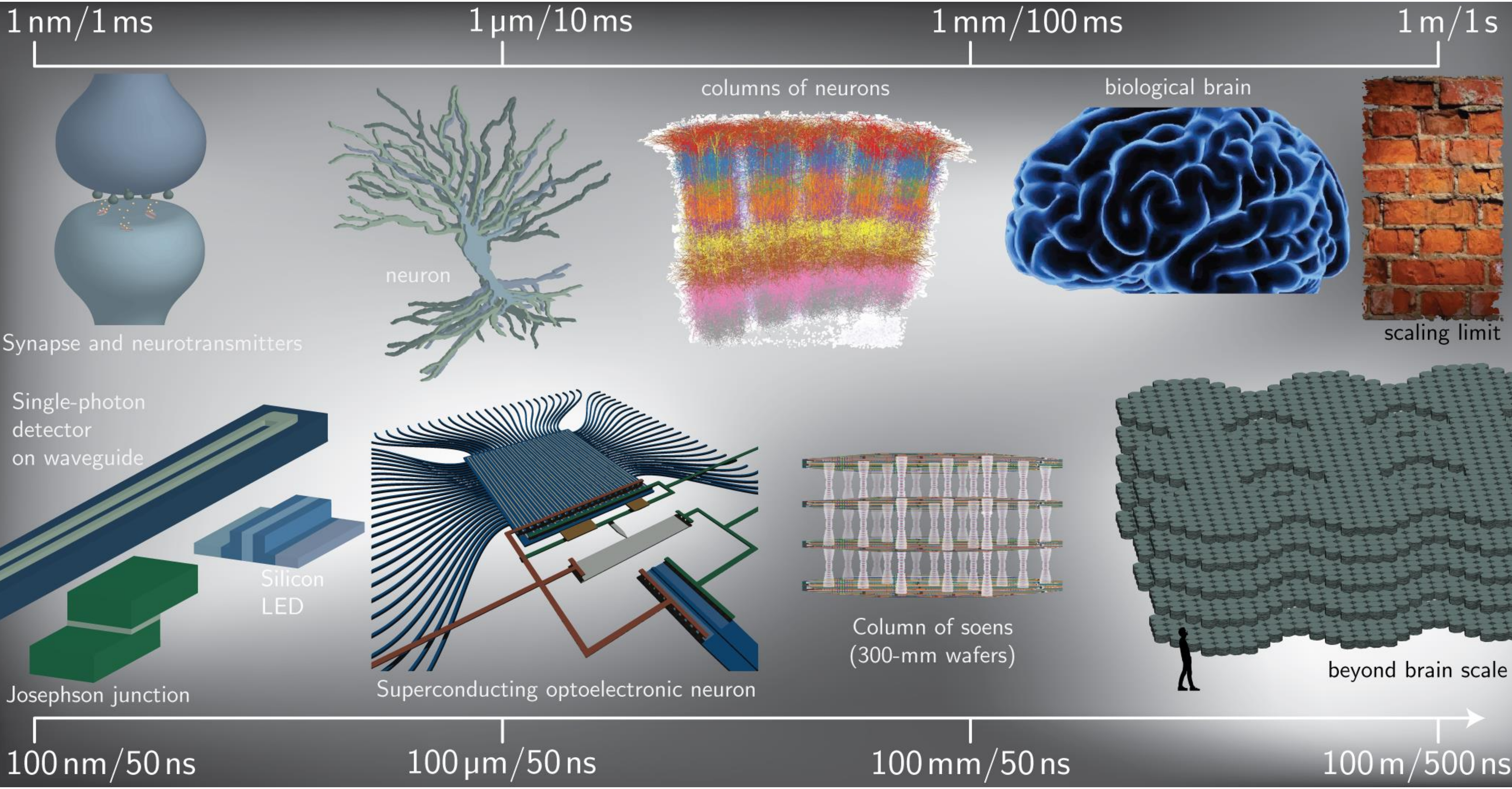


Artificial visual cortex

Superconducting optoelectronic networks

- Dense local fan-out with photonics
- Long-range communication at light speed
- Computing and memory with superconducting electronics

Would this reach the limits of cognition?


1 nm/1 ms
1 μm/10 ms
1 mm/100 ms
1 m/1 s
columns of neurons
biological brain
neuron
Synapse and neurotransmitters
scaling limit
Single-photon
detector
on waveguide
Silicon
LED
Josephson junction
Superconducting optoelectronic neuron
Column of soens
(300-mm wafers)
beyond brain scale
100 nm/50 ns
100 μm/50 ns
100 mm/50 ns
100 m/500 ns

Summary

- Low Temperature Devices have the potential to be important in **super-neuromorphic computing**
  - Light for communication
  - Superconductors for computation
  - Complex neuro circuits

Can we get to a size scale for "Cognitive" Systems?

- Rich Mirin
- Jeff Shainline
- Sonia Buckley
- Adam McCaughan
- Alex Tait
- Jeff Chiles

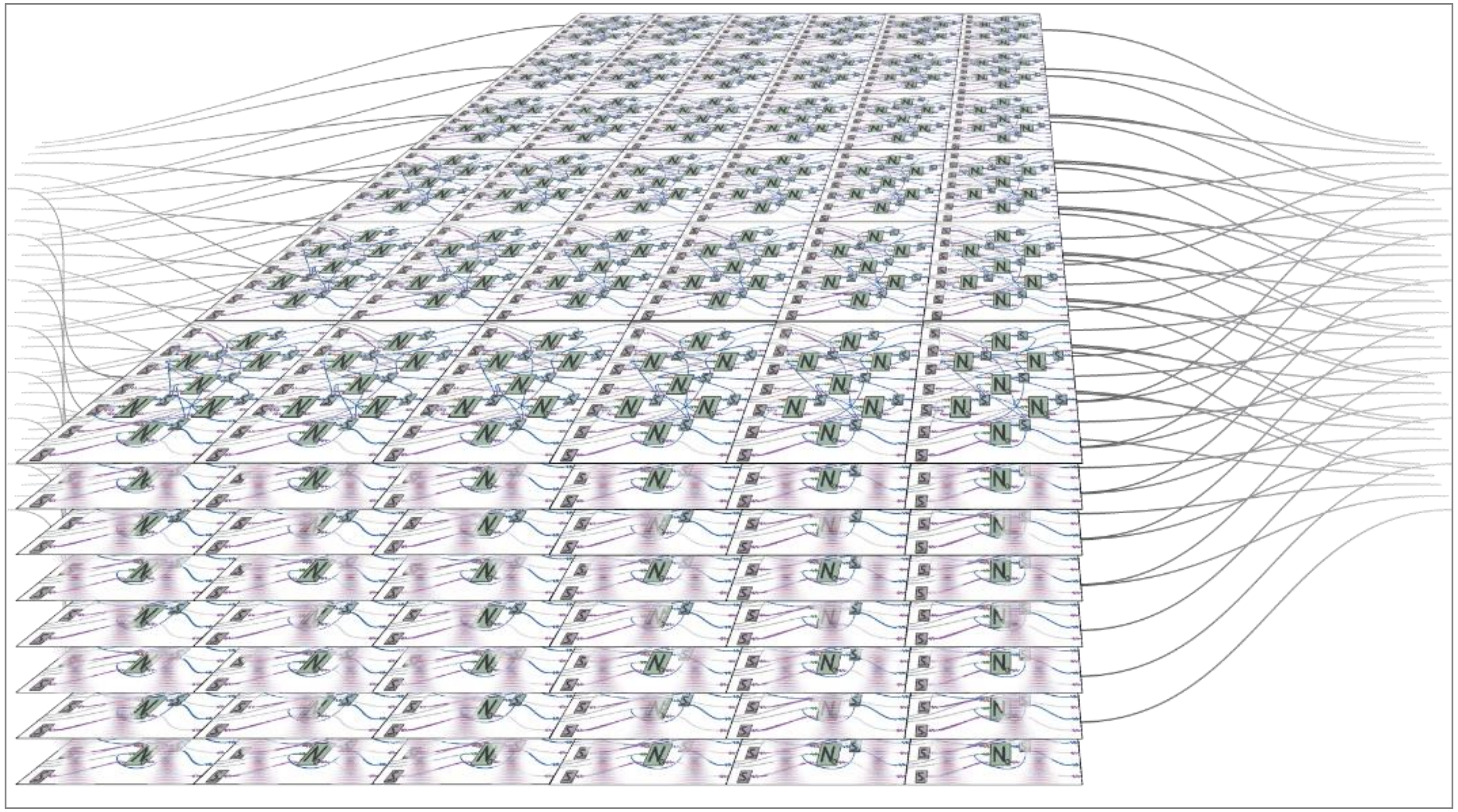
- Saeed Khan
- Krister Shalm
- Marty Stevens
- Adriana Lita
- Varun Verma
- Nima Nader

- Mike Mazurek
- Dileep Reddy
- Eric Stanton
- Galen Moody
- Kevin Silverman
- Thomas Gerrits

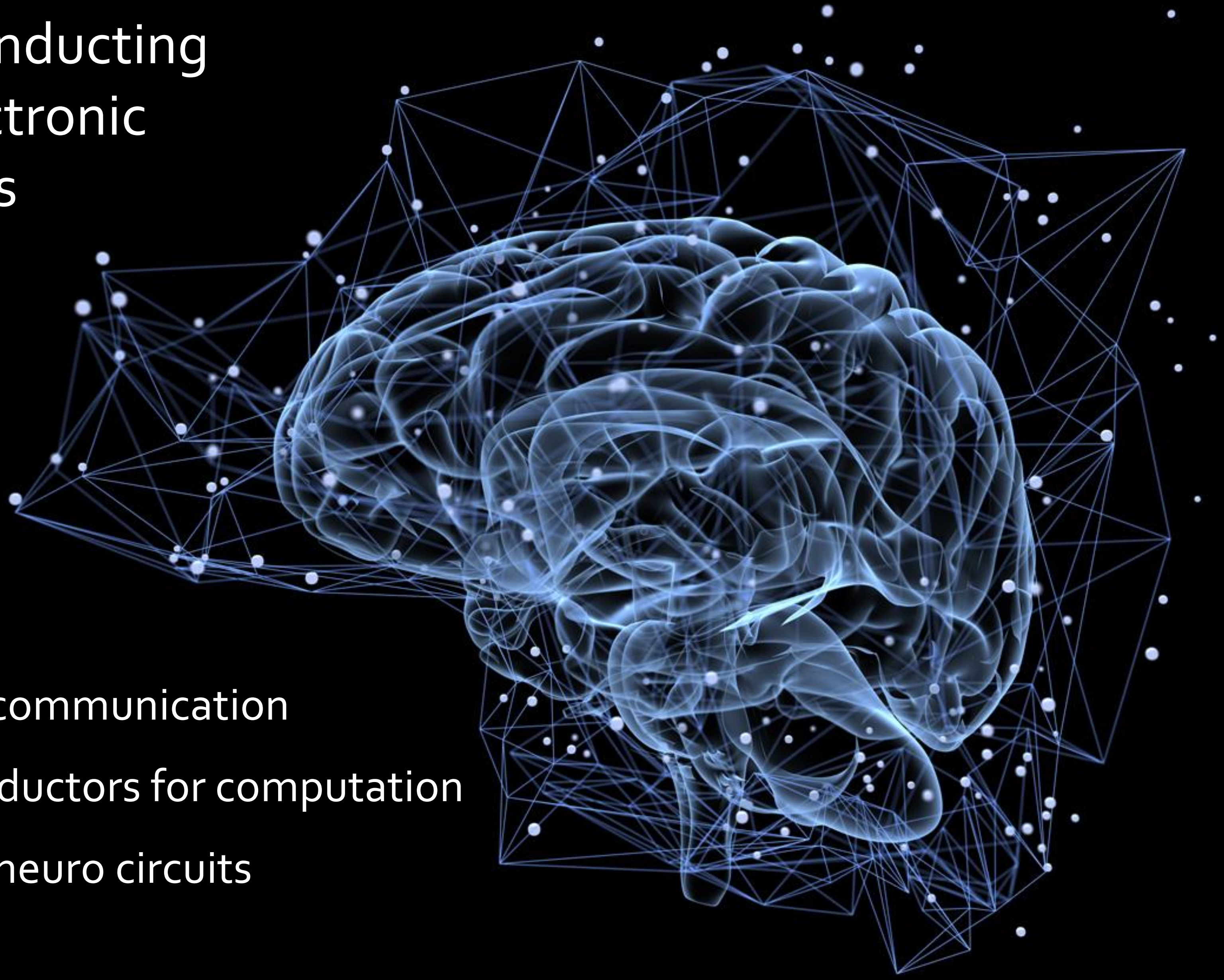
saewoo.nam@nist.gov

MIT, JPL, Suny-Poly

- 10^{10} times the number of synapses as the brain
- Operating 30,000 times faster



Superconducting optoelectronic networks



- Light for communication
- Superconductors for computation
- Complex neuro circuits

Quantum-neural hybrid computing