UNIVERSITY OF TRENTO - Italy
Department of Physics

1st Summer School of
Interdisciplinary Research on Brain Network Dynamics
June 24-28, 2019

Photonic Processing Unit - Acceleration of Neural Network Training Based on Analog Optical Crossbar Arrays

Roger Dangel, Ph.D
Neuromorphic Devices and Systems Group
IBM Zurich GmbH

June 24, 2019

IBM

Outlook

Quick Start :
- Recap from yesterday: Status of today's Deep Neural Network processing
- How to transfer the concept of analog crossbar arrays into optics (photonics)

Part 1:
- Holographic weight storage and signal processing
- Optical crossbar array: Weight processing operations
- Analogy: Optical crossbar array ⇔ electrical crossbar array
- Prior approach in free-space-optics
- Integrated-optical solution
- Short intermezzo: What is Si-photonics ?

Part 2:
- Building blocks of analog integrated-optical crossbar array
- Beam shaping and routing optics: simulation
- Optical coupling between Si-photonics and GaAs: Simulation
- Integration of photorefractive GaAs with Si-photonics

Part 3:
- First experimental results:
  - Chip for Si-Photonics functionality test
    - Beam shaping and routing optics
    - Input vector setup unit
  - Proof of single synapse function in photorefractive bulk GaAs
  - Photorefractive storage medium

Summary

People Involved in this PPU Project


Bert
Offrein


Folkert
Horst


me


Yannick
Baumgartner


Efe
Büyüközer

Recap from yesterday: Computational Challenge: Matrix-Vector Multiplications

Matrix-vector multiplications of the form

$$\boldsymbol{W}\boldsymbol{x} = \begin{bmatrix} w_{0,0} & w_{0,1} & w_{0,2} & & w_{0,N} \\ w_{1,0} & w_{1,1} & w_{1,2} & \cdots & w_{1,N} \\ w_{2,0} & w_{2,1} & w_{2,2} & & w_{2,N} \\ & \vdots & & \ddots & \vdots \\ w_{M,0} & w_{M,1} & w_{M,2} & \cdots & w_{M,N} \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{N} w_{0,i}\, x_i \\ \sum_{i=0}^{N} w_{1,i}\, x_i \\ \sum_{i=0}^{N} w_{2,i}\, x_i \\ \vdots \\ \sum_{i=0}^{N} w_{M,i}\, x_i \end{bmatrix}$$

are common to the mentioned workloads and dominate the computation time and energy consumption.

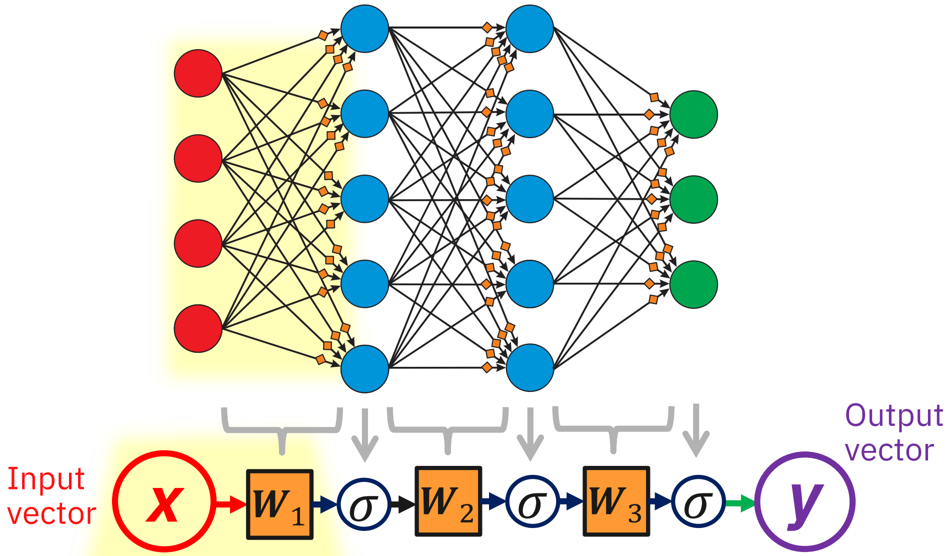
Matrix-vector multiplications are "computationally expensive" !

Our mission

Develop dedicated hardware (→ Analog Crossbar Arrays) which enables **efficient analog implementation of matrix-vector multiplications** and therefore acceleration of Deep Neural Network Learning

Recap from yesterday: DNN Training by Backpropagation Algorithm

Neural net as chain of vector operations:



Accumulate

Multiply

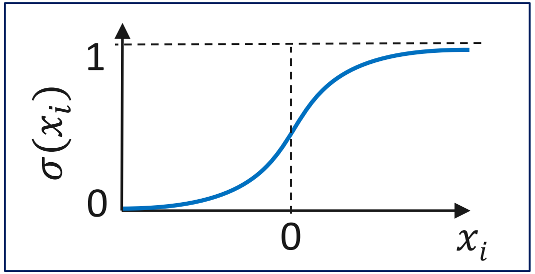
$$W_1 x = \begin{bmatrix} w1_{0,0} & w1_{0,1} & .. & w1_{0,N} \\ w1_{1,0} & w1_{1,1} & .. & w1_{1,N} \\ \vdots & \vdots & & \vdots \\ w1_{M,0} & w1_{M,1} & .. & w1_{M,N} \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{N} w1_{0,i}\, x_i \\ \sum_{i=0}^{N} w1_{1,i}\, x_i \\ \vdots \\ \sum_{i=0}^{N} w1_{M,i}\, x_i \end{bmatrix}$$

Components:

- Layers of neurons
- Synaptic interconnections

Mathematical operations:

- → : Signal vector
- W_n : Synaptic weight matrix $[W_n]$
- σ : Per-element neural (non-linear) activation function (sigmoid):



Recap from yesterday: Status of Today's Deep Neural Network Processing

Current situation

- Processing dominated by large matrix operations

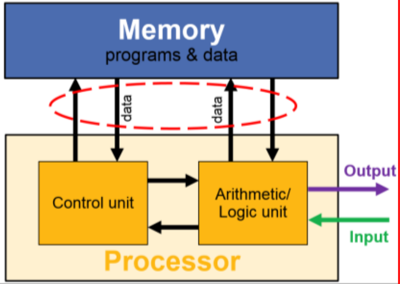
  Forward propagation: W
  Backward propagation: W^T
  Weight update: ΔW

  Scale $\propto N^2$ → Neurons/layer

- Large training datasets: Thousands of training cases

High performance computer

- Inefficient on standard Von-Neumann architecture systems:
  - (Mostly) Serial processing
  - Low computation to IO ratio
  - → Memory bottleneck



Today's standard computer architecture
(→ proposal by John Von-Neumann in 1945)

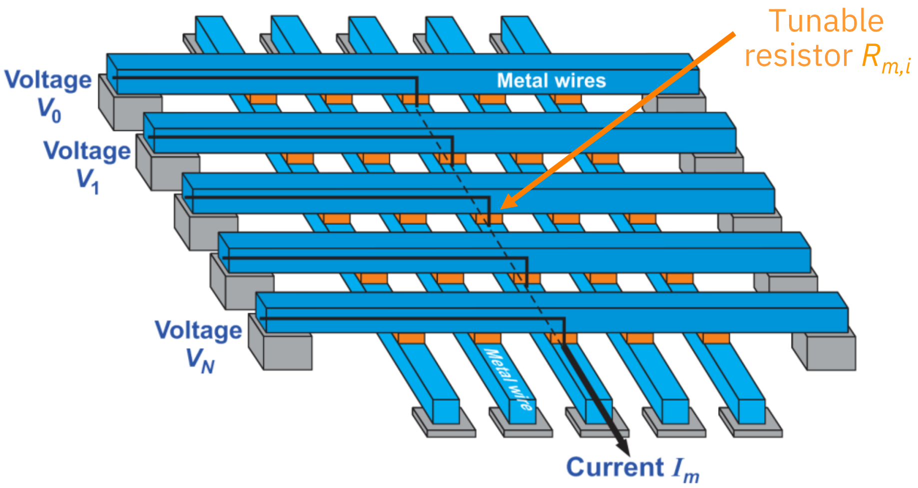
Need for faster and more efficient DNN processing

Borrow some concepts from the brain:
- Analog signal processing
- Fully parallel processing
- Tight integration of processing and memory

Analog Crossbar Arrays

How to Transfer Concept of Analog Crossbar Arrays into Optics

Analog **electrical** crossbar array



Tunable resistors (with "memory") ⇔ Synaptic weights

Resistive Processing Unit (RPU)
to accelerate processing of DNNs

Analog **optical** crossbar array
=
Analog **photonic** crossbar array

?

???? ⇔ Synaptic weights

Photonic Processing Unit (PPU)
to accelerate processing of DNNs

Outlook

Part 1:

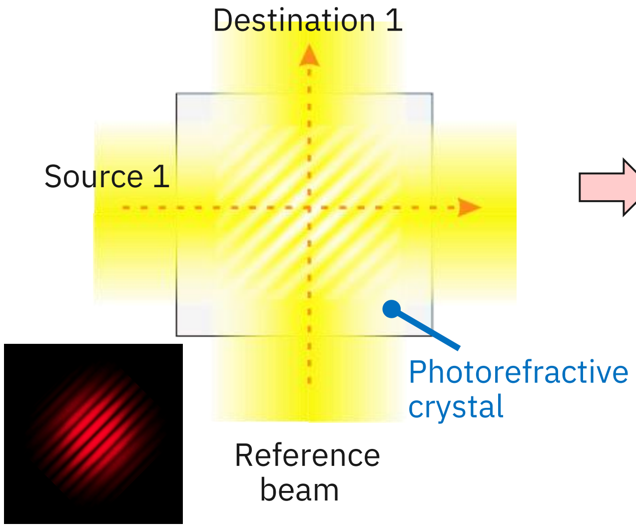
- Holographic weight storage and signal processing
- Optical crossbar array: Weight processing operations
- Analogy: Optical crossbar array ⇔ electrical crossbar array
- Prior approach in free-space-optics
- Integrated-optical solution
- Short intermezzo: What is Si-photonics ?

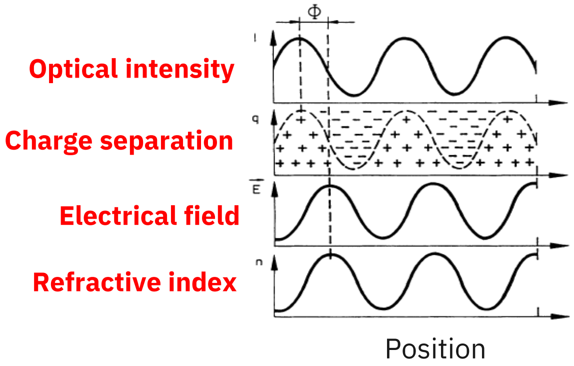
Holographic Weight Storage and Signal Processing

- **Concept:** Synaptic connections are created as refractive index gratings in a photorefractive material
- **Photorefractive material:** ① Optically active electron traps + ② Pockels effect

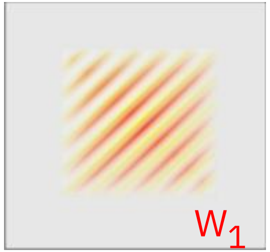
⇨ Two interfering optical beams in photorefractive material can write a refractive index grating:

Interference pattern → **Photorefractive effect** → **Stored diffraction grating**





Cornelia Denz, Optical Neural Networks, 1998.



- Linear effect
- Symmetric: Write + π phase shift = Erase

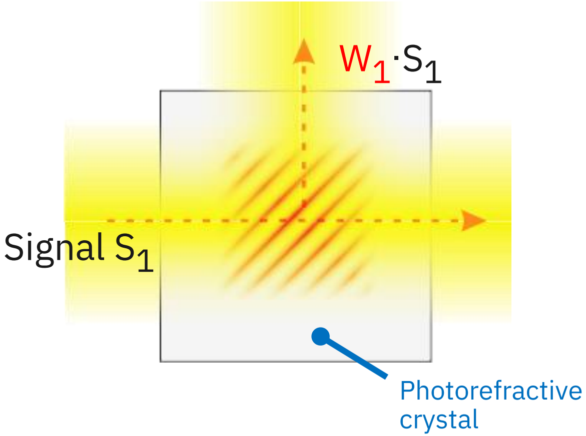
Holographic Weight Storage and Signal Processing

Synaptic weight processing:

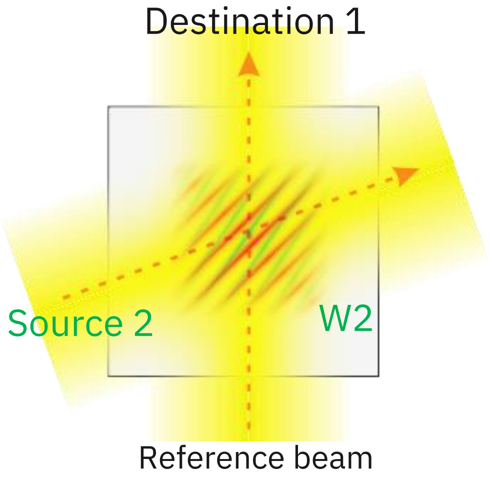
Readout on two diffraction gratings:

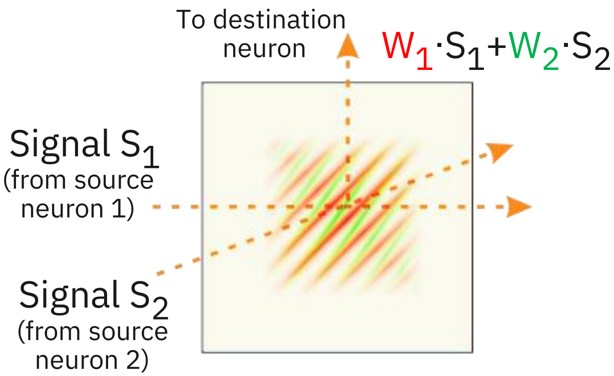
"Multiply & accumulate"

Diffraction grating **readout**:



Write a **2nd** grating:





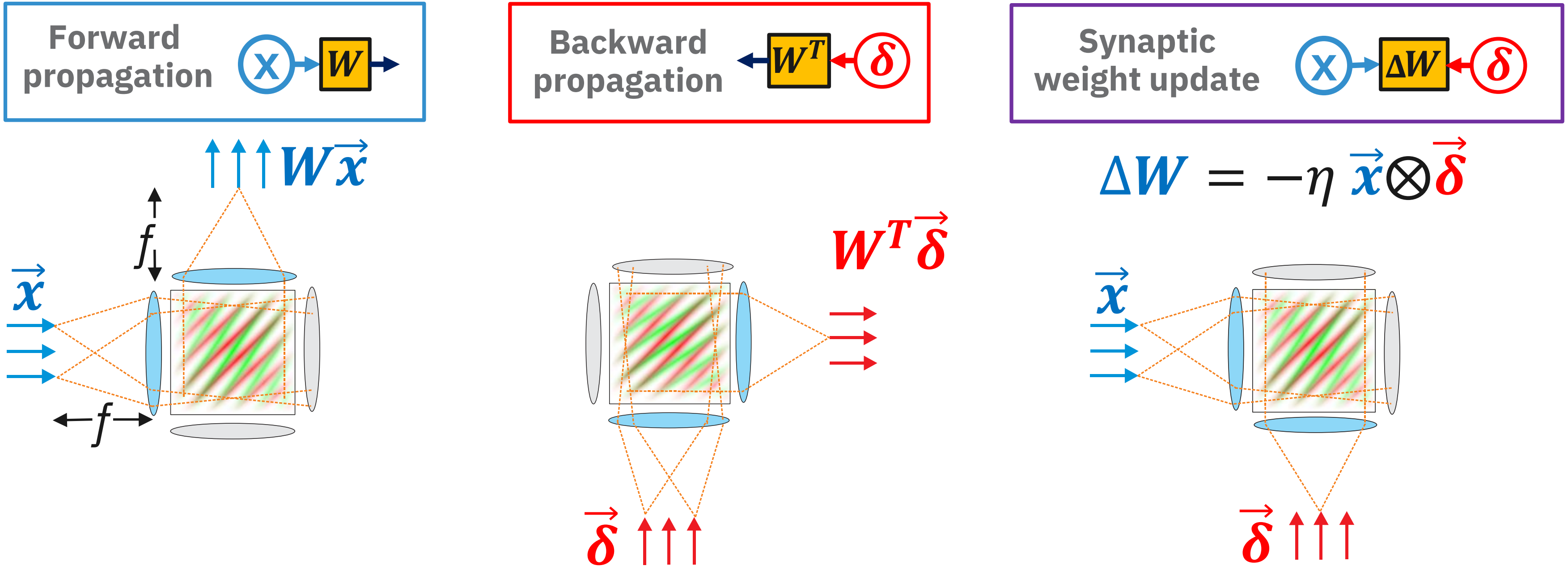
Synaptic weight gratings diffract light from optical input beams to optical output beams

- Source and destination neurons are encoded by different beam angles in the crystal
- There is a unique grating for every source - destination combination
- Optical signaling: Amplitude & Phase ➔ Bipolar signals and weights

Optical Crossbar Array: Weight Processing Operations

Interaction in the photorefractive medium uses collimated beams under different angles

- Add lenses around the medium for conversion to/from arrays of point sources or images:



All weight processing operations for neural network processing are supported

Analogy: Optical Crossbar Array ⇔ Electrical Crossbar Array

Analog optical crossbar array

Optical waveguides

Distributed weights

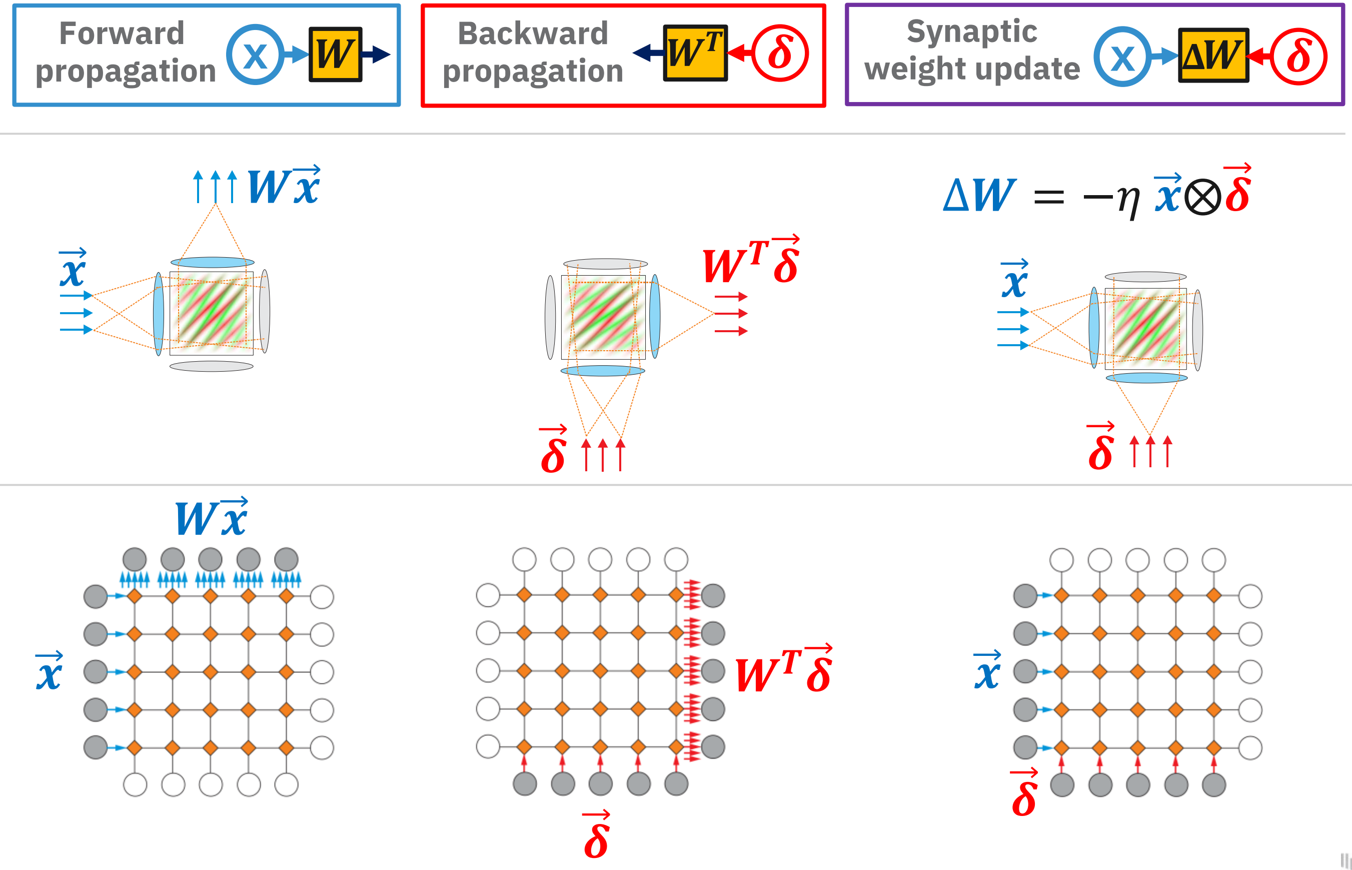
Refractive index tuning

Resistance tuning

Local weights

Electrical wires

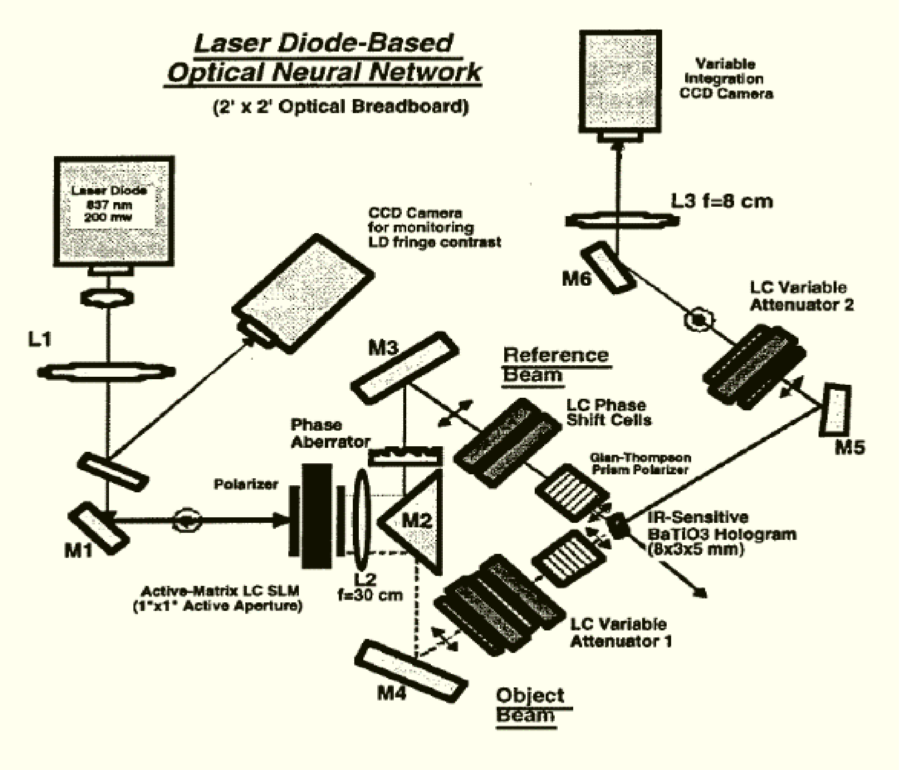
Analog electrical crossbar array



Optical Crossbar Array: Prior Approach in Free-Space-Optics

Concept was demonstrated in the 90s using free-space optics:

- Example: Hughes Research Laboratories
- Backpropagation training of Artificial Neural Networks shown
- Large optical breadboard setup, slow electro-optics



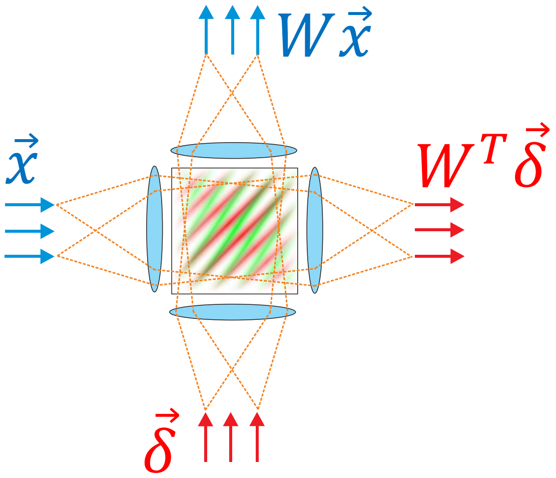
Yuri Owechko and Bernard H. Soffer, "Holographic neurocomputer utilizing laser diode light source“, 1995

Optical Crossbar Array: Integrated Solution

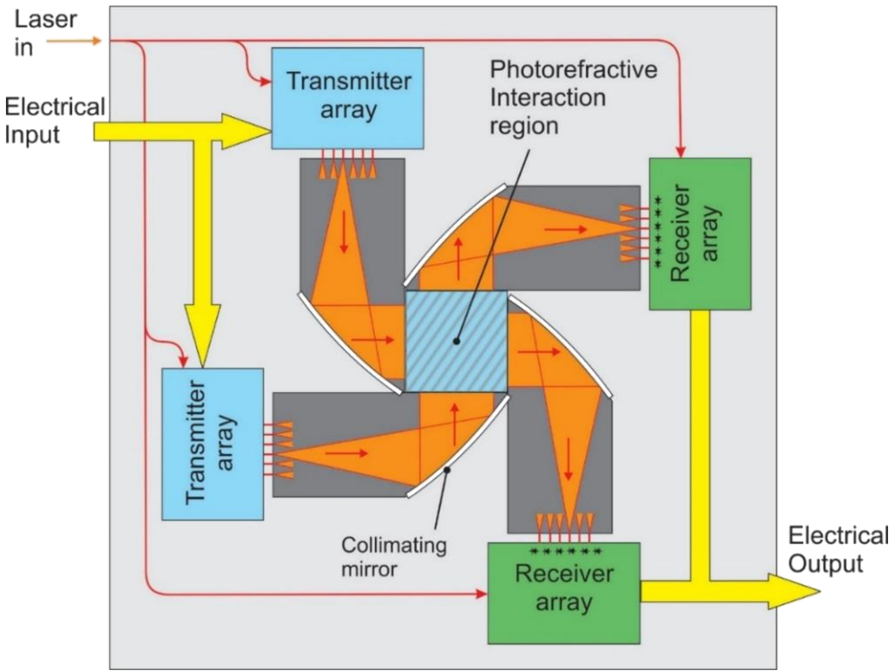
Our novel approach

Miniaturize using Integrated Optics

- Electro-optic conversion and beam shaping optics **on a Silicon-Photonics (Si-Pho) chip**
- Memory: Photorefractive **thin film** on silicon



Schematic layout

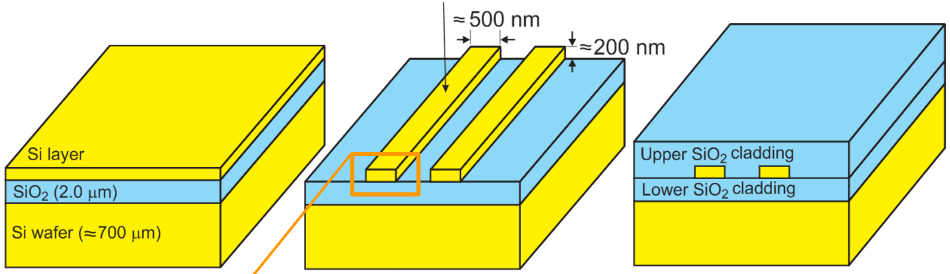


Short Intermezzo: What is Si-Photonics ?

- **Silicon photonics (Si-Pho)** is the technology of photonic integrated systems which use silicon (Si) as light carrier medium
- Si-Pho technology exploits semiconductor fabrication techniques (CMOS-technology) established for electrical integrated circuits which also use Si as technology platform

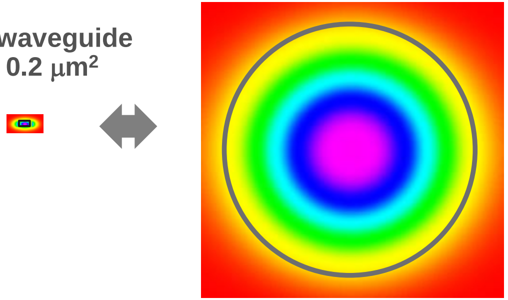
Silicon-On-Insulator (SOI) wafer | **Si-Pho waveguide**

Refractive indices:
$n_{Si} = 3.5$
$n_{SiO_2} = 1.45$



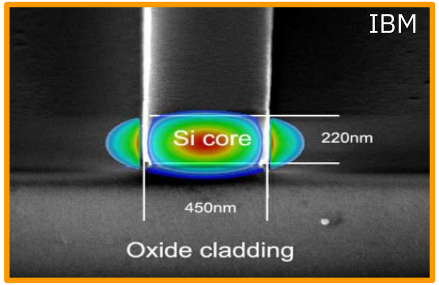
Si-Pho waveguide
0.5 x 0.2 μm²

Single-mode fiber core
∅ = 9 μm

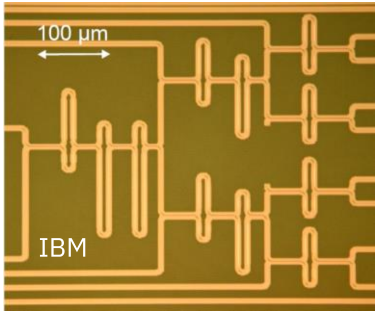


Refractive index difference:
$n_{core} - n_{cladding} \approx 0.01$

SEM image:
Si-Pho waveguide facet



Micrograph of complex Si-Pho device:
Cascaded Mach-Zehnder interferometers enabling wavelength filtering for WDM



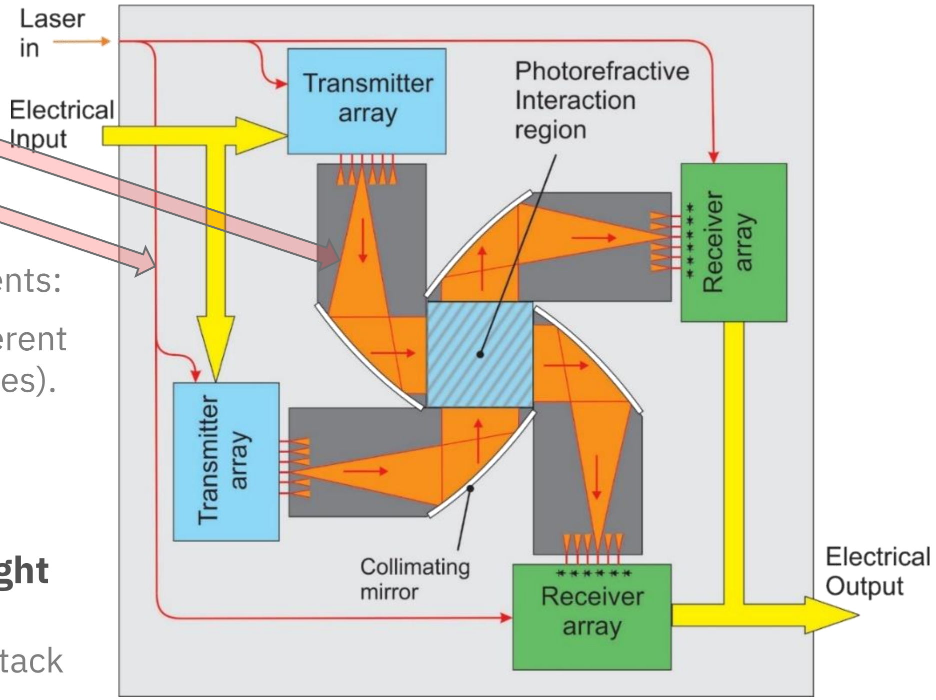
Outlook

Part 2:

- Building blocks of analog integrated-optical crossbar array
- Beam shaping and routing optics: simulation
- Optical coupling between Si-photonics and GaAs: Simulation
- Integration of photorefractive GaAs with Si-photonics

Building Blocks of Analog Integrated-Optical Crossbar Array

- **2-dimensional and planar Si-photonics waveguides**
- **Beam shaping optics:**
  - Parabolic-shaped collimating mirrors
  - Curved/tilted focal planes for aberration correction
- **Transmitter & receiver array**
  Standard Si-photonics modulators and detector components:
  - Transmitter: Sets up the input vector as an array of coherent point light sources with adjustable intensities (and phases).
  - Detector: Detects intensity (and phase) of the refocused output signals
- **Photorefractive interaction region for synaptic weight storage**
  - III/V photorefractive material bonding to Si-photonics stack
  - Overlap area for vertical directional coupling



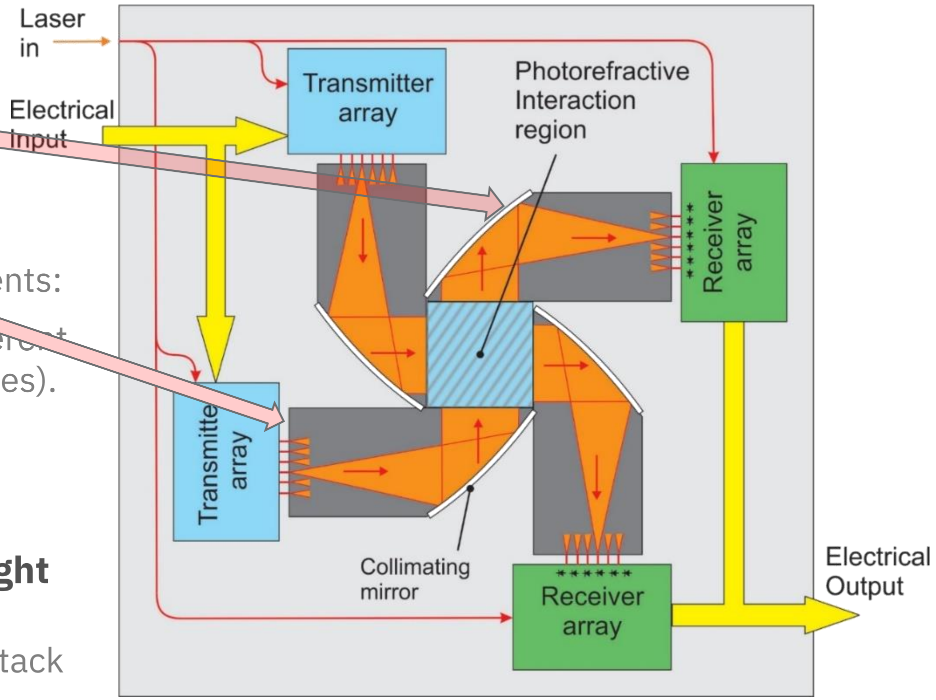
Building Blocks of Analog Integrated-Optical Crossbar Array

- **2-dimensional and planar Si-photonics waveguides**

- **Beam shaping optics:**
  - Parabolic-shaped collimating mirrors
  - Curved/tilted focal planes for aberration correction

- **Transmitter & receiver array**
  Standard Si-photonics modulators and detector components:
  - Transmitter: Sets up the input vector as an array of coherent point light sources with adjustable intensities (and phases).
  - Detector: Detects intensity (and phase) of the refocused output signals

- **Photorefractive interaction region for synaptic weight storage**
  - III/V photorefractive material bonding to Si-photonics stack
  - Overlap area for vertical directional coupling



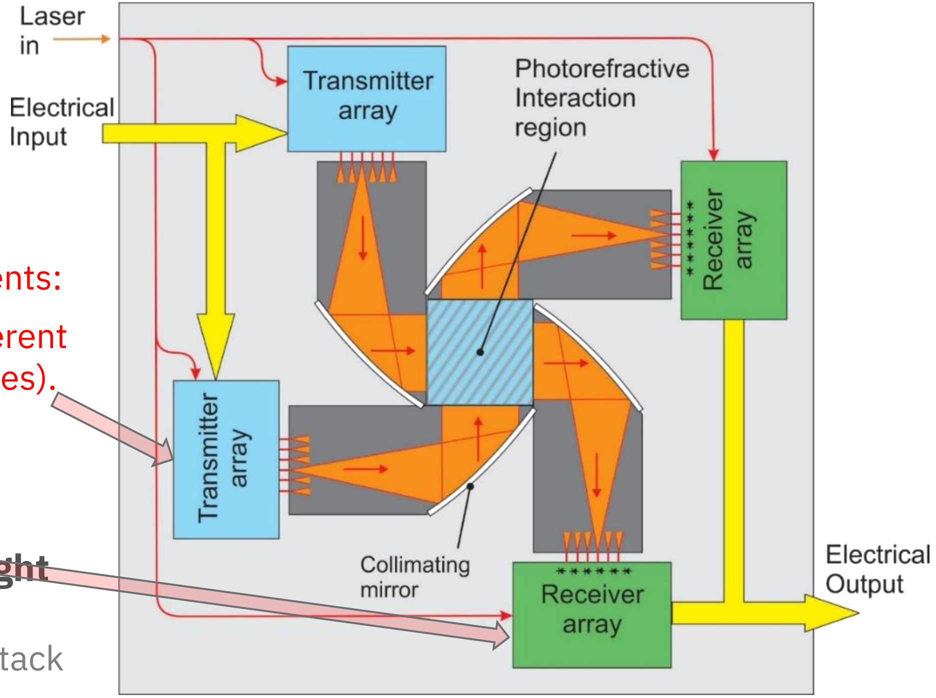
Building Blocks of Analog Integrated-Optical Crossbar Array

- **2-dimensional and planar Si-photonics waveguides**

- **Beam shaping optics:**
  - Parabolic-shaped collimating mirrors
  - Curved/tilted focal planes for aberration correction

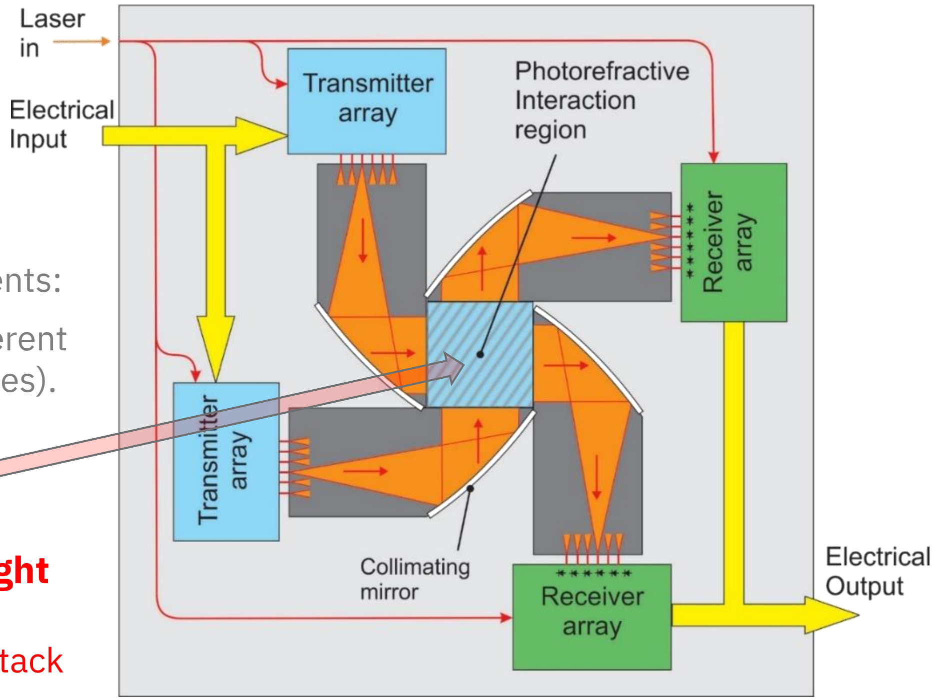
- **Transmitter & receiver array**
  Standard Si-photonics modulators and detector components:
  - Transmitter: Sets up the input vector as an array of coherent point light sources with adjustable intensities (and phases).
  - Detector: Detects intensity (and phase) of the refocused output signals

- **Photorefractive interaction region for synaptic weight storage**
  - III/V photorefractive material bonding to Si-photonics stack
  - Overlap area for vertical directional coupling

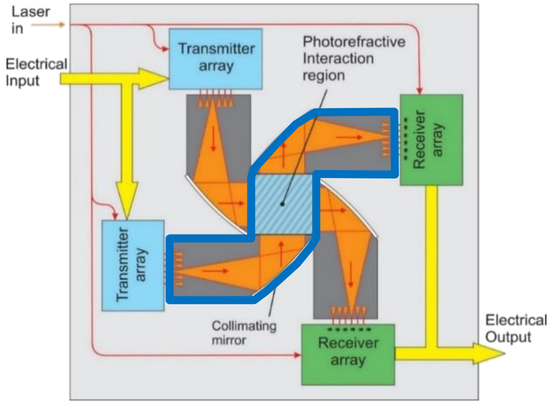


Building Blocks of Analog Integrated-Optical Crossbar Array

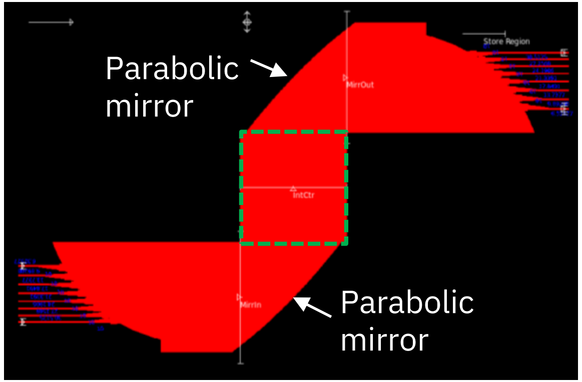
- **2-dimensional and planar Si-photonics waveguides**
- **Beam shaping optics:**
  - Parabolic-shaped collimating mirrors
  - Curved/tilted focal planes for aberration correction
- **Transmitter & receiver array**
  Standard Si-photonics modulators and detector components:
  - Transmitter: Sets up the input vector as an array of coherent point light sources with adjustable intensities (and phases).
  - Detector: Detects intensity (and phase) of the refocused output signals
- **Photorefractive interaction region for synaptic weight storage**
  - III/V photorefractive material bonding to Si-photonics stack
  - Overlap area for vertical directional coupling



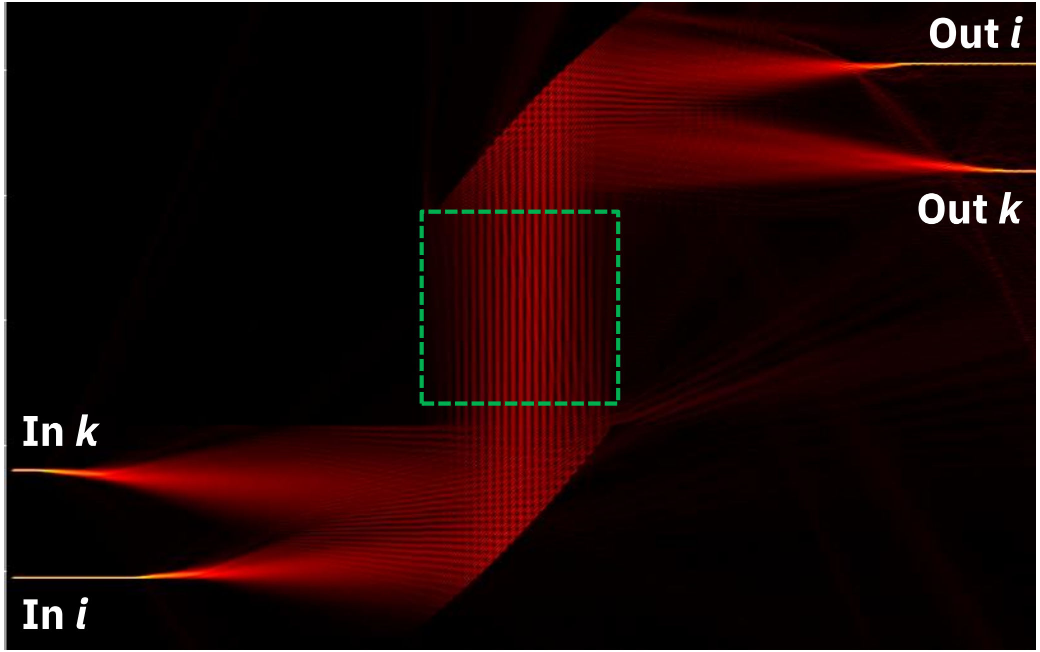
Beam Shaping and Routing Optics: Simulation



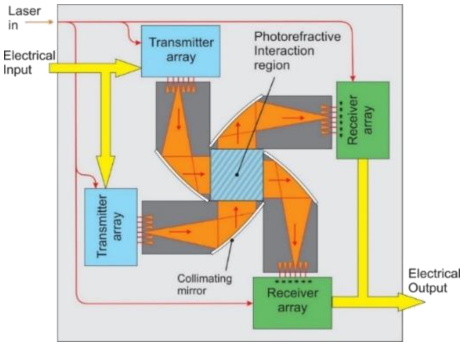
Layout

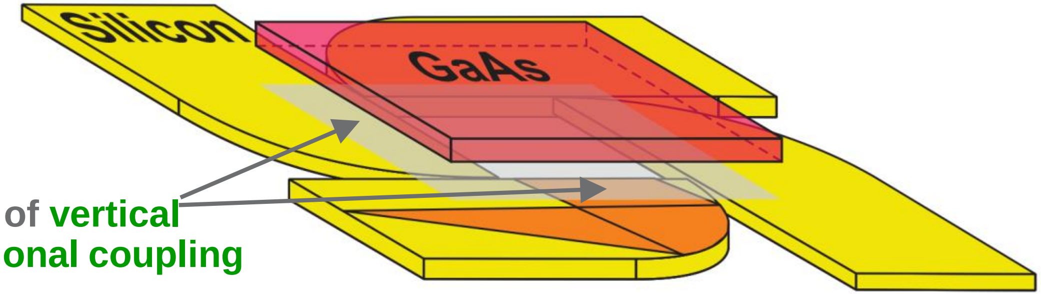


Simulated transmission



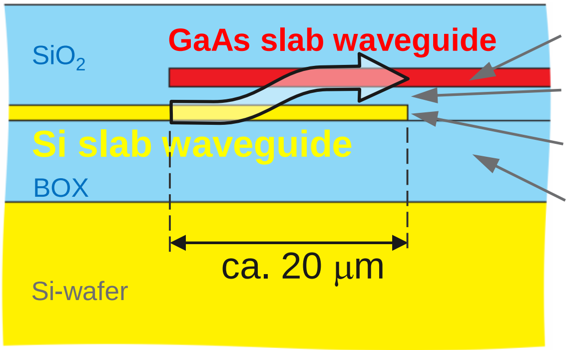
Optical Coupling Between Si-Photonics and GaAs: Simulation





Areas of **vertical directional coupling**

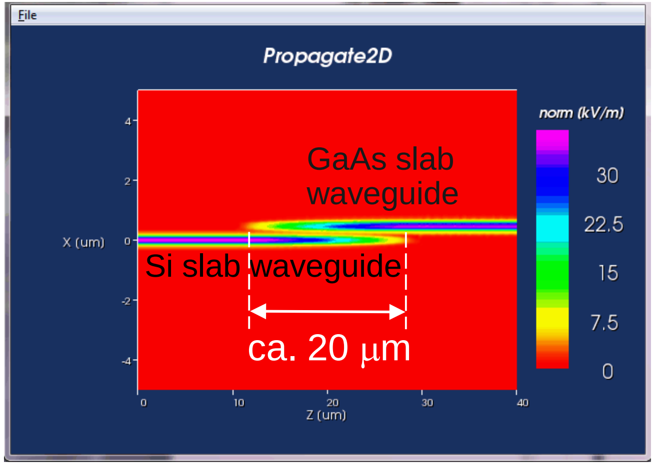
Design



Layer thicknesses:

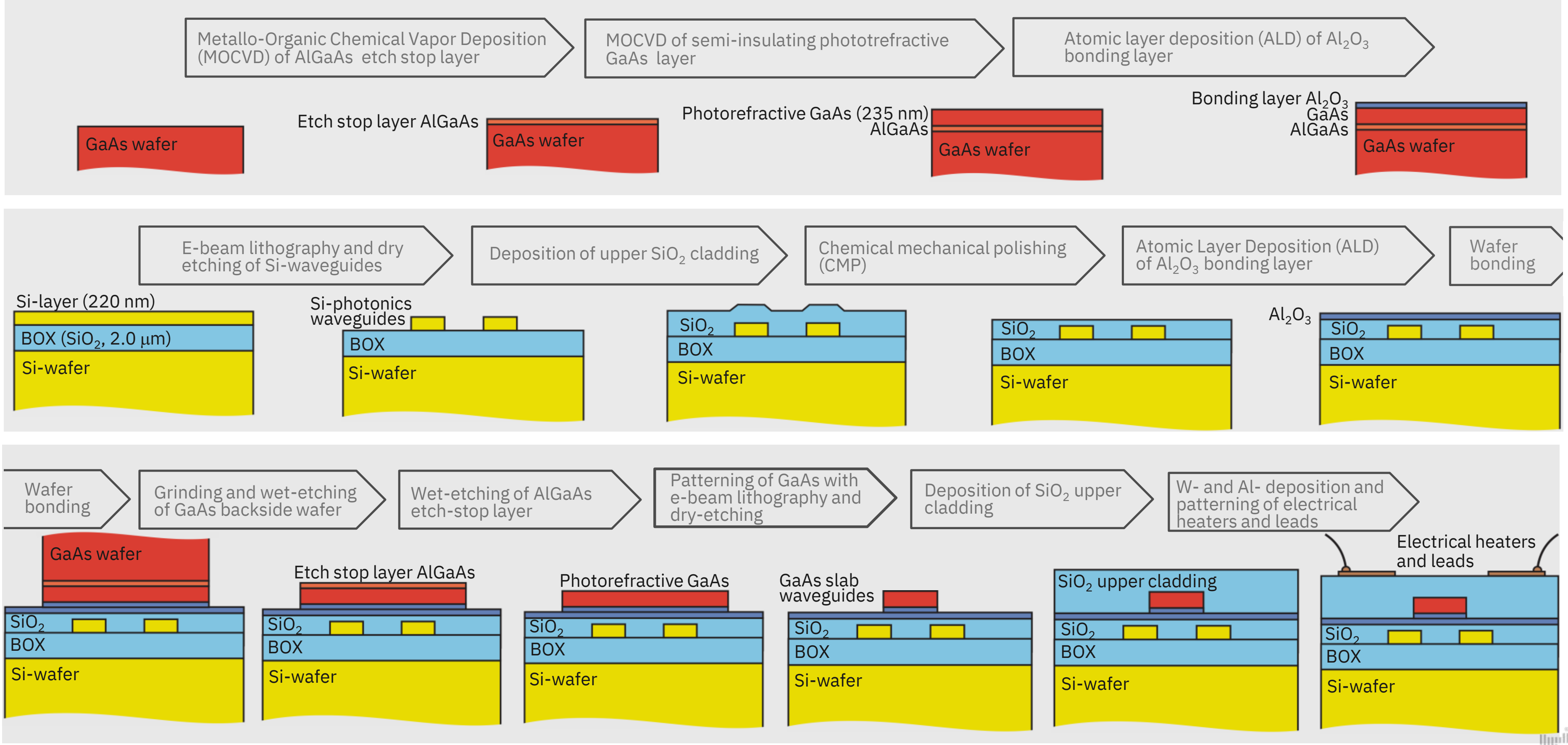
$t_{GaAs\text{-}WG}$ = 260 nm
t_{Gap} = 180 nm
$t_{Si\text{-}WG}$ = 220 nm
t_{BOX} = 2.0 µm

Simulation result



Mode overlap calculation: **99.7%**

Integration of Photorefractive GaAs with Si-Photonics



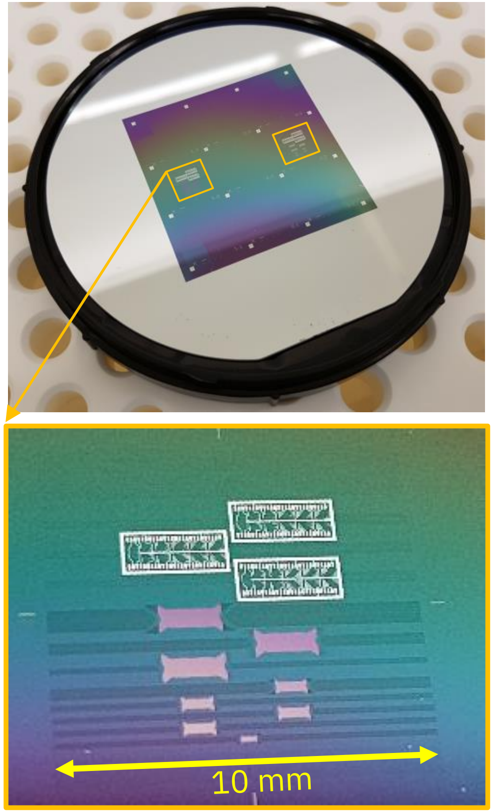
Outlook

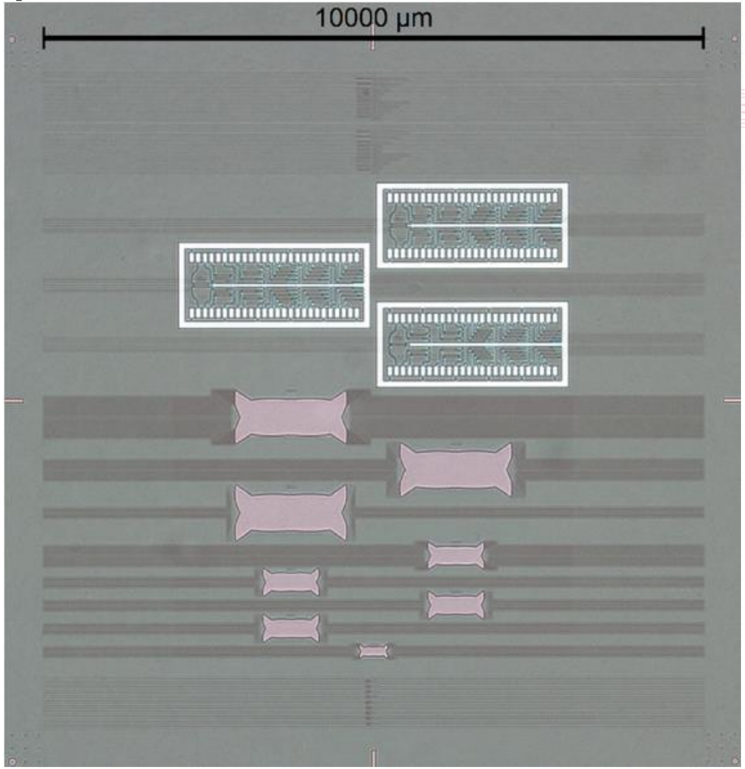
Part 3:

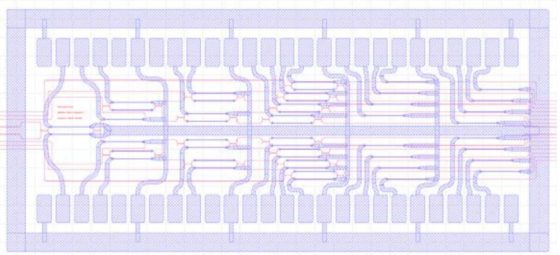
- First Experimental Results:
  - Chip for Si-Photonics functionality test
    - Beam shaping and routing optics
    - Input vector setup unit
  - Proof of single synapse function in photorefractive bulk GaAs
  - Photorefractive storage medium

First Experimental Results: Chip for Si-Photonics Functionality Test

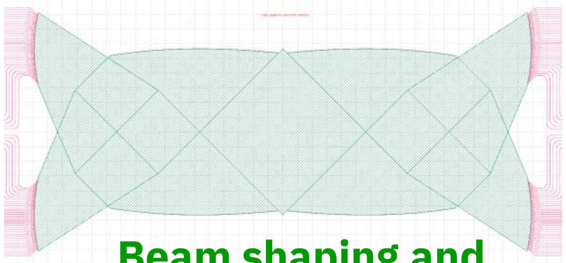
Photograph of realized test chip





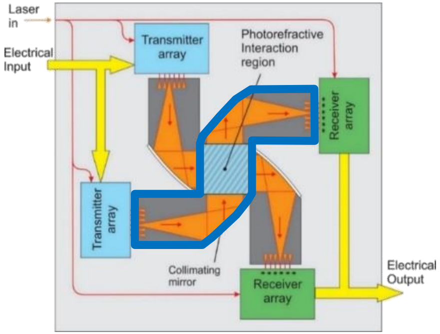


Input vector setup unit

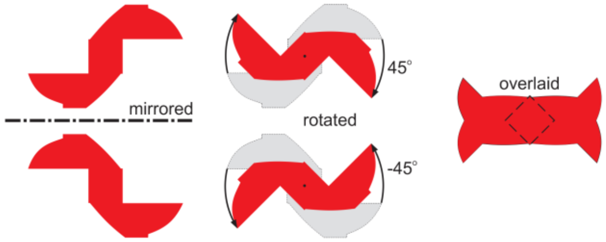


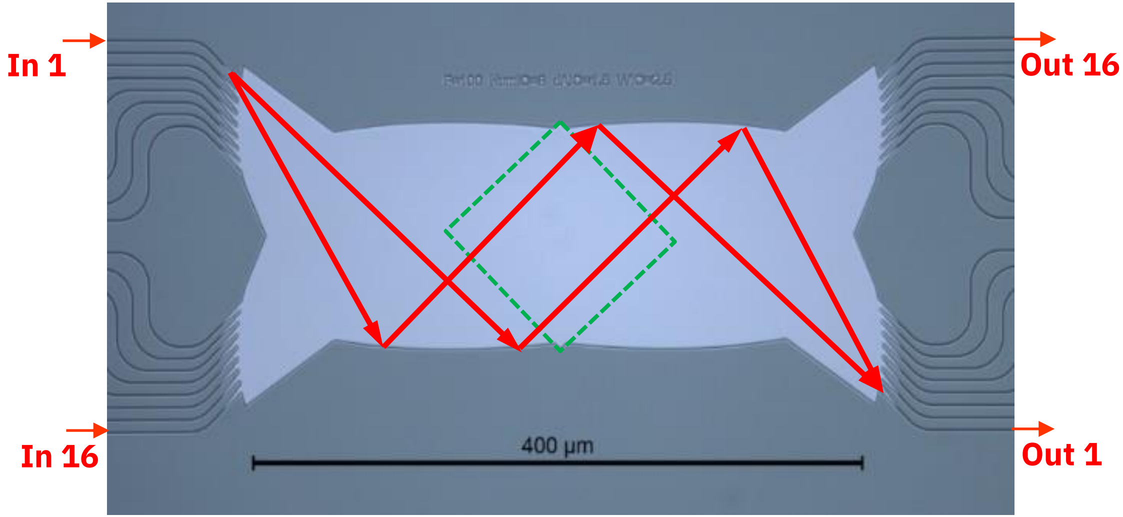
Beam shaping and routing optics

Beam Shaping and Routing Optics

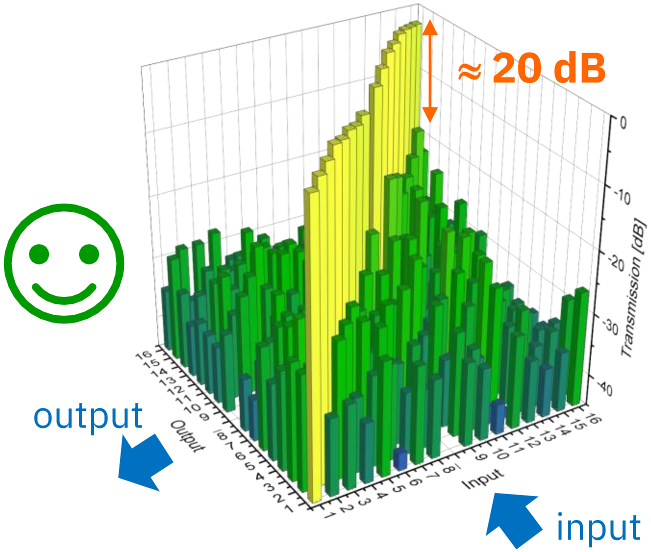


Slightly modified structure to test Si-photonics functionality

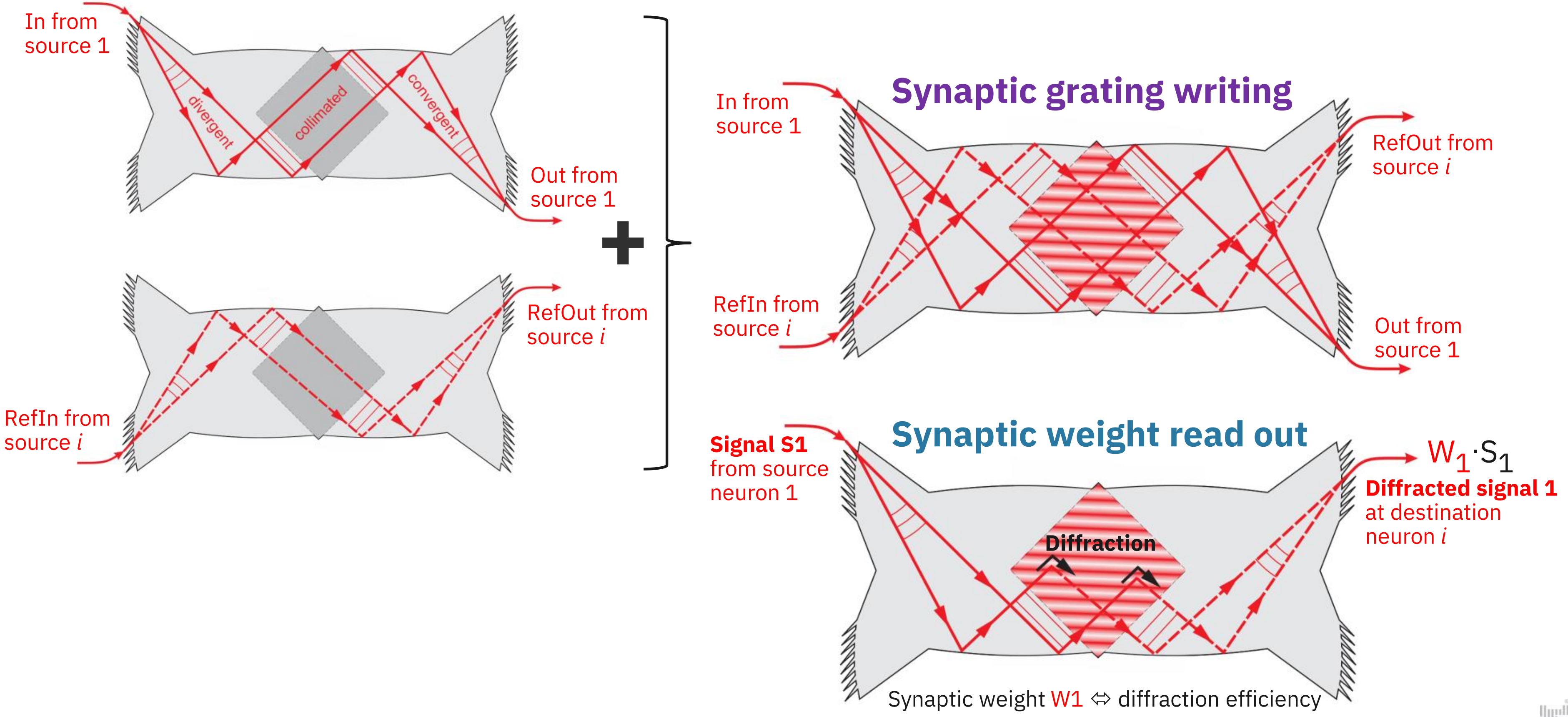




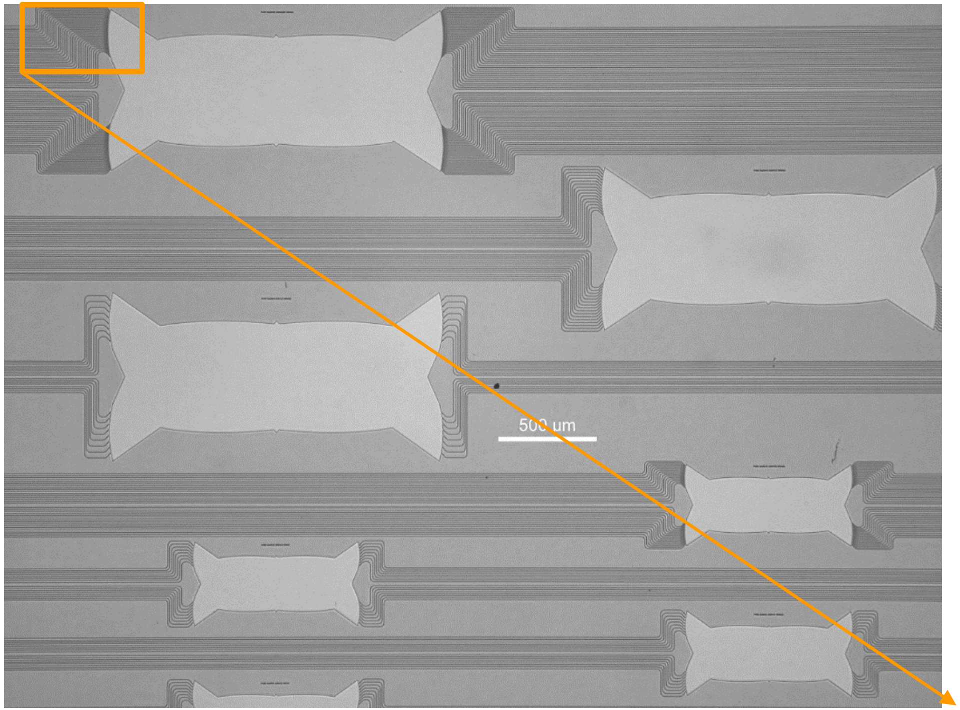
Experimental result of transmission matrix:



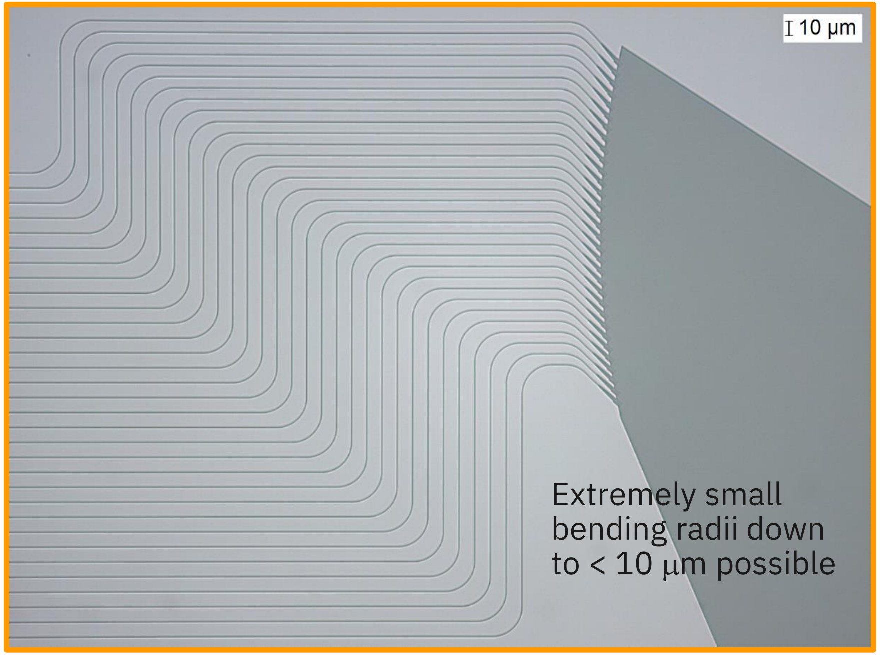
Beam Shaping and Routing Optics



Beam Shaping and Routing Optics



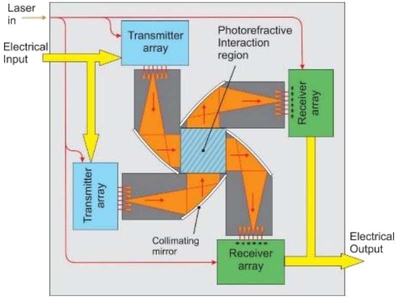
2 x 32 Inputs → 2 x 32 Outputs



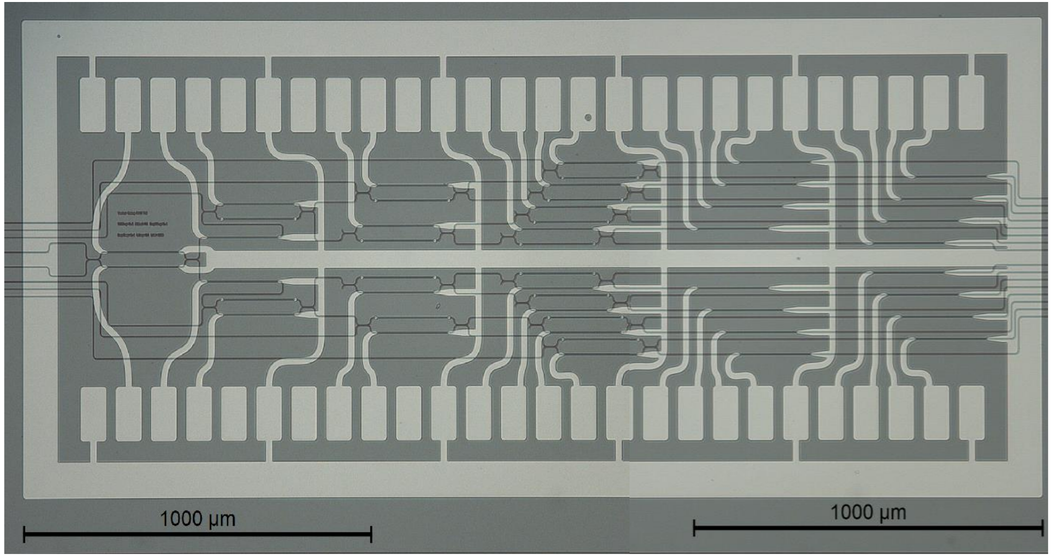
Input Vector Setup Unit

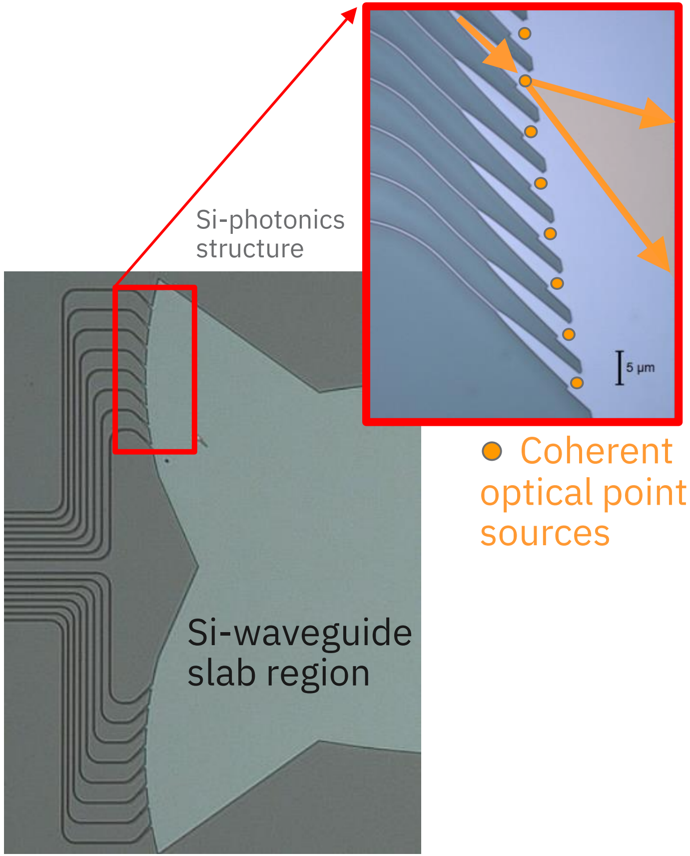
Task of input vector setup unit:

Encoding electrical signals onto the transmitter array of coherent optical point sources.



Micrograph of input vector setup unit

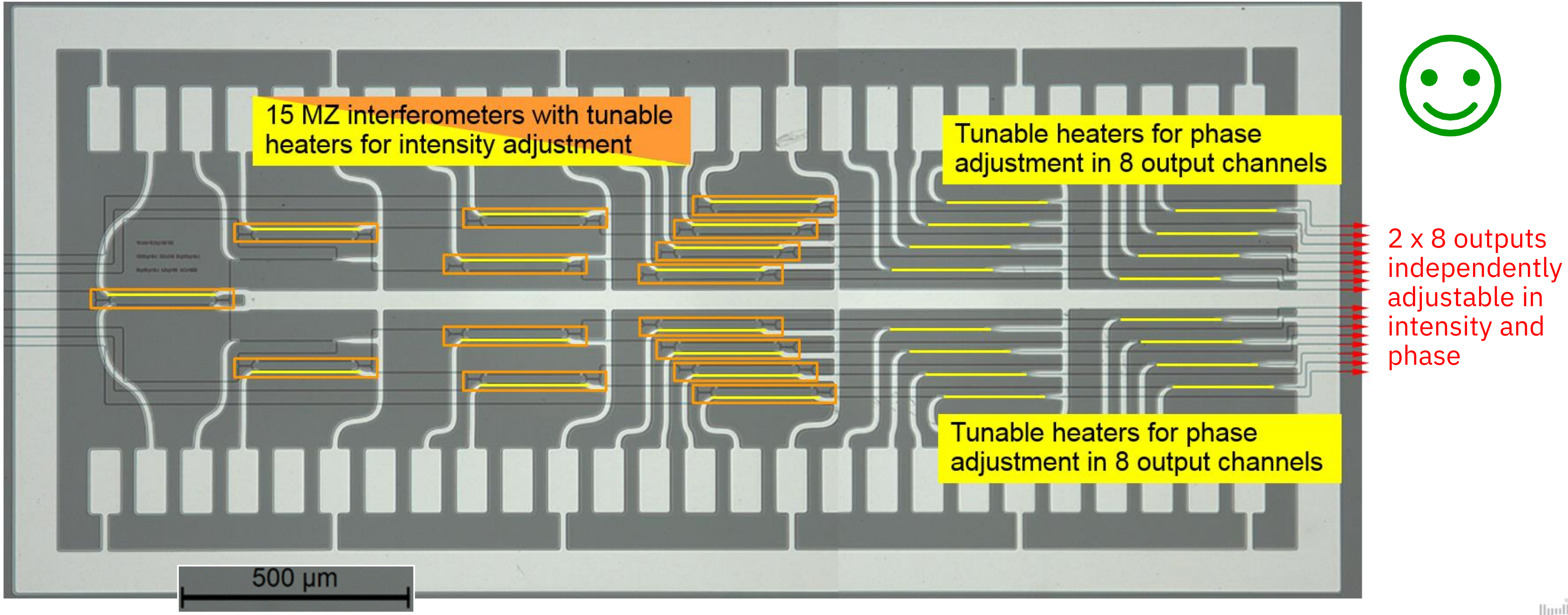




Input Vector Setup Unit

Micrograph of complete unit comprising:

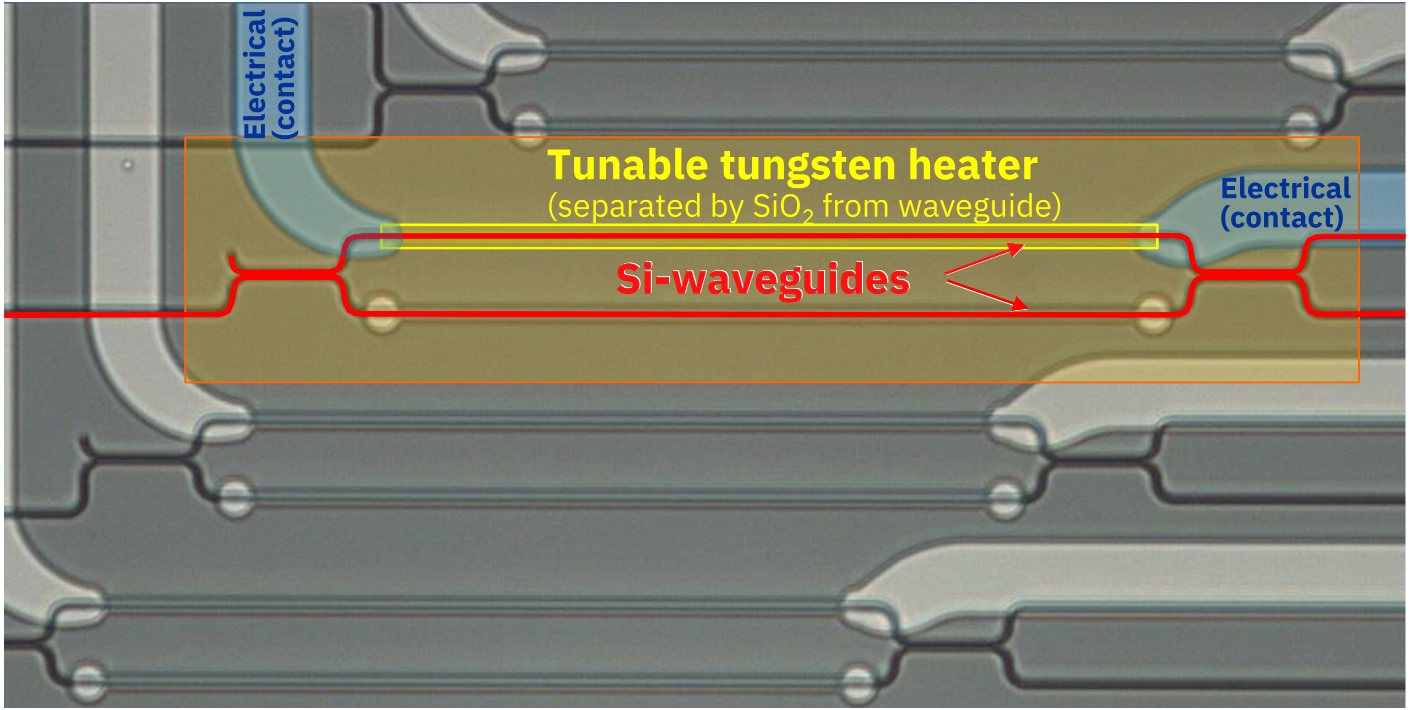
- 15 (= 1 + 2 + 4 + 8) Mach-Zehnder interferometers with tunable heaters for individual intensity adjustment
- 2 x 8 tunable heaters for individual phase adjustment

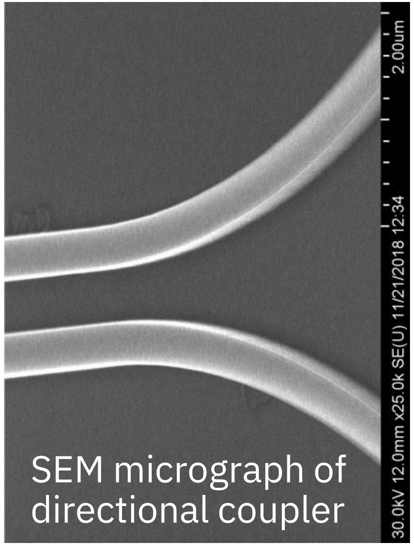


Input Vector Setup Unit

Micrograph of Mach-Zehnder interferometers (each) comprising:

- 2 interferometer waveguide arms (one arm with tungsten heater for tuning phase changes)
- 2 directional couplers for (50:50 splitting and combining)

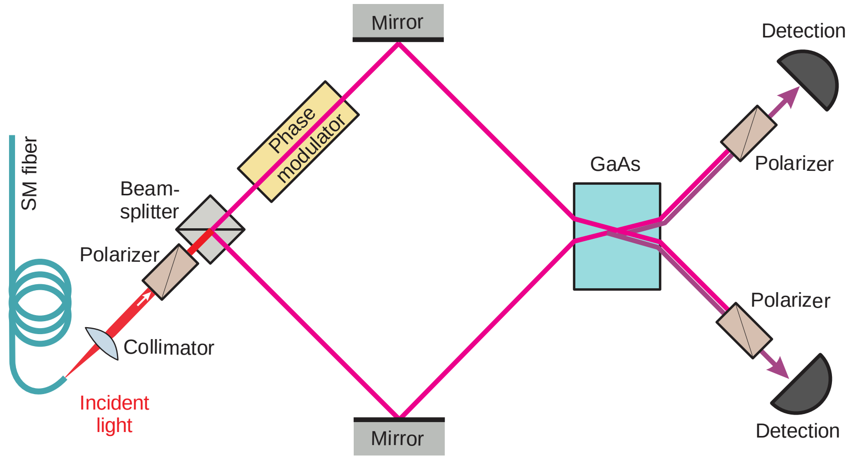




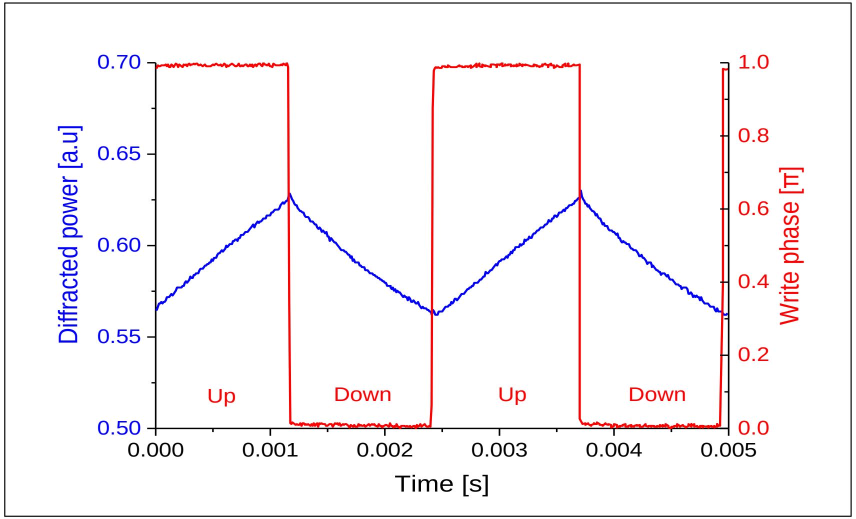
SEM micrograph of directional coupler

Proof of Single Synapse Function in Photorefractive Bulk GaAs

Two-wave mixing in bulk GaAs crystal ≈ single synapse:



📖 P. Yeh, "Two-Wave Mixing in Nonlinear Media"
doi: 10.1109/3.18564



Control of photorefractive weight:

- **Smooth slopes: analog behavior**
- **Symmetric**

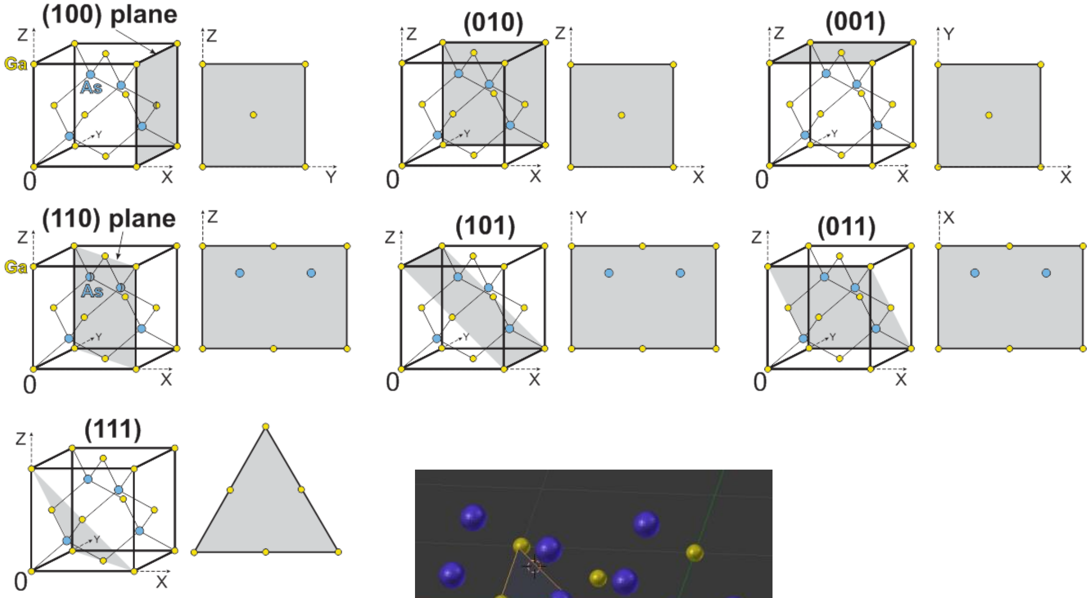
Next step: Reproduce in thin GaAs film

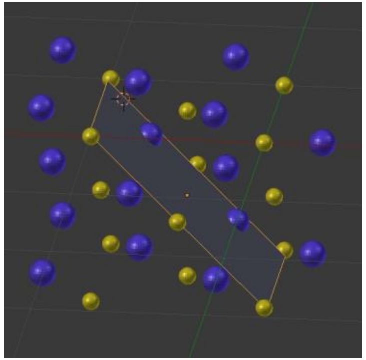
Photorefractive Storage Medium: First Results

Thin film of semi-insulating GaAs

- MOCVD growth, optimized for:
  - 110 substrate:
    - ⇨ Correct orientation of Pockels tensor
  - As-rich growth → "EL2" Deep traps

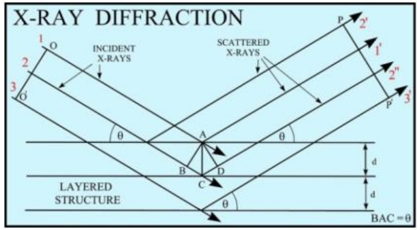
- Material alternatives would be:
  - GaAs:Cr
  - InP:Fe
  - $BaTiO_3$

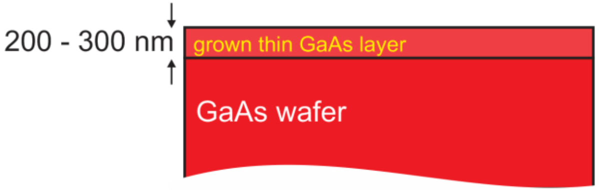


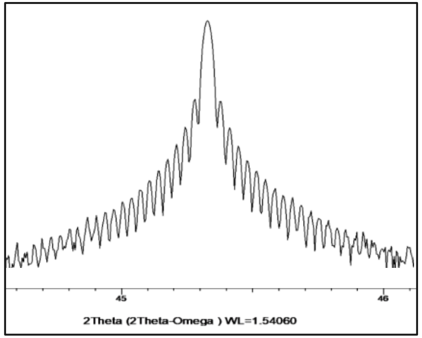


Photorefractive Storage Medium: First Results

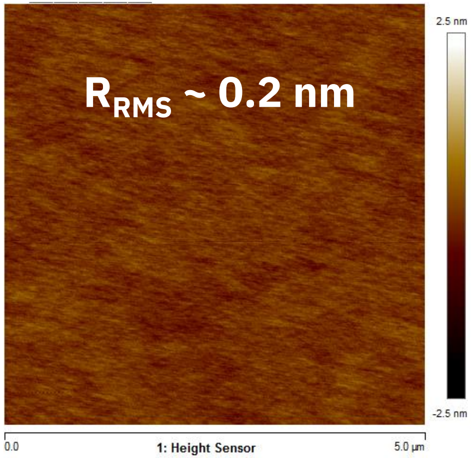
- Analysis of grown GaAs layers my means of **XRD** (X-Ray Diffraction):



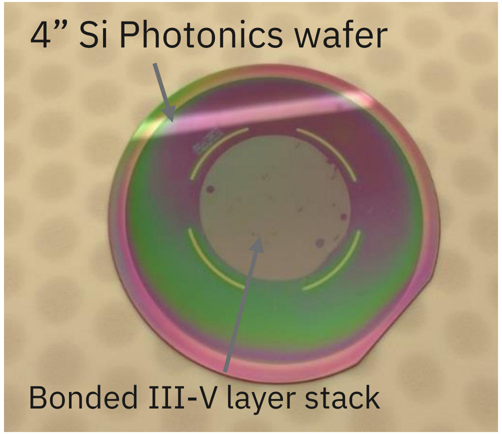




- Roughness measurement result obtained with **AFM**:



- **Wafer bonding** of 2” GaAs wafer onto 4” Si photonics wafer



Photorefractive Storage Medium: Current Status

Results:

- **Epitaxial layer growth**
  - Good crystalline quality
  - Smooth surface
  - Bonds well to SiO_2 on Si
- **Semi-insulating layer**
- **But: Strong optical absorption**
  - Because concentration too high

Next step:
Fine-tuning of EL2 trap concentration

Summary

- **Our group mission:**
  - Research for Artificial Intelligence (AI)

- **Our strategy in the Photonic Processing Unit project:**
  - Exploitation of our **Si-photonics** expertise for neuromorphic computing hardware development

- **Our project:**
  - Realization of integrated analog optical crossbar arrays for speed-up of computationally "expensive" matrix-vector multiplications in deep neural network processing

- **Our main tasks:**
  - Demonstration of analog synaptic weight storage in photorefractive semi-insulating thin GaAs layers
  - Realization of Si-photonics chip containing all building blocks to show NxN crossbar array operation

- **Our current status:**
  - Correct functioning of major Si-photonics building blocks demonstrated
  - Photorefractive effect in semi-insulating GaAs bulk material demonstrated
  - Wafer bonding of GaAs to Si-photonics demonstrated
  - MOCVD based epitaxial growth of semi-insulating thin GaAs layers demonstrated
  - Tuning of GaAs growth process with correct EL2 trap concentration for achieving suitable photorefractivity and low optical propagation loss is ongoing