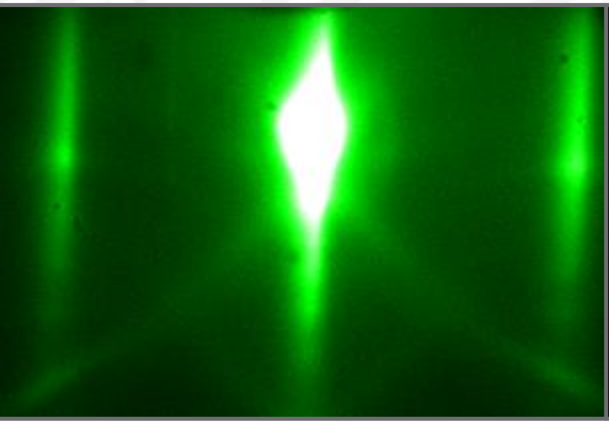
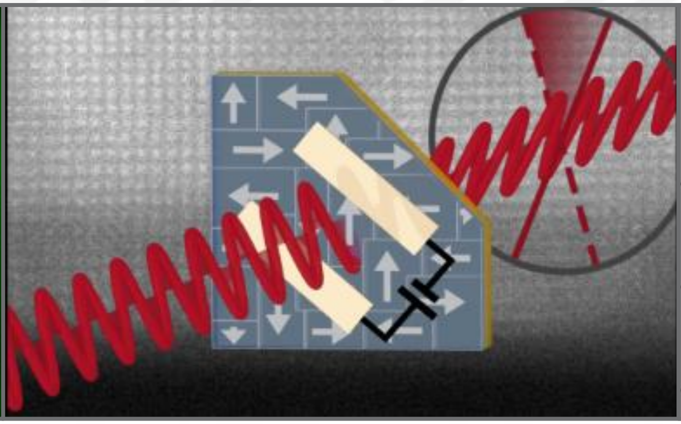
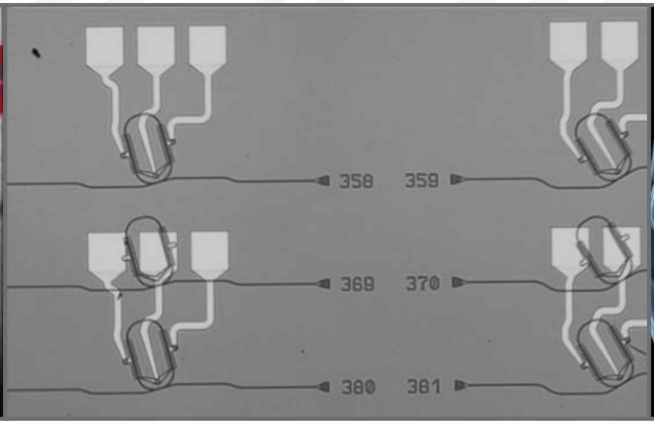
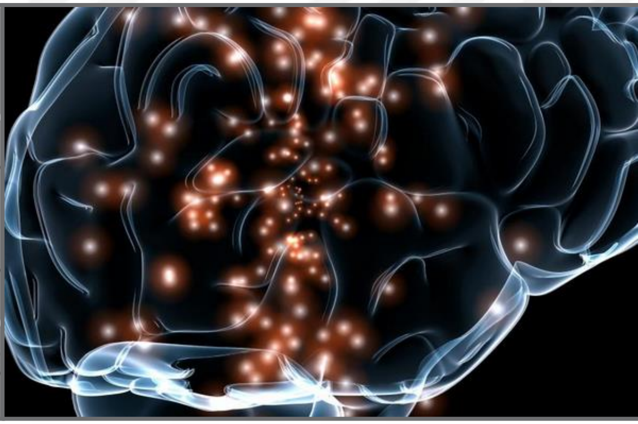
Analog photonic accelerators for neuromorphic computing

Bert Jan Offrein









Neuromorphic Devices and Systems Group

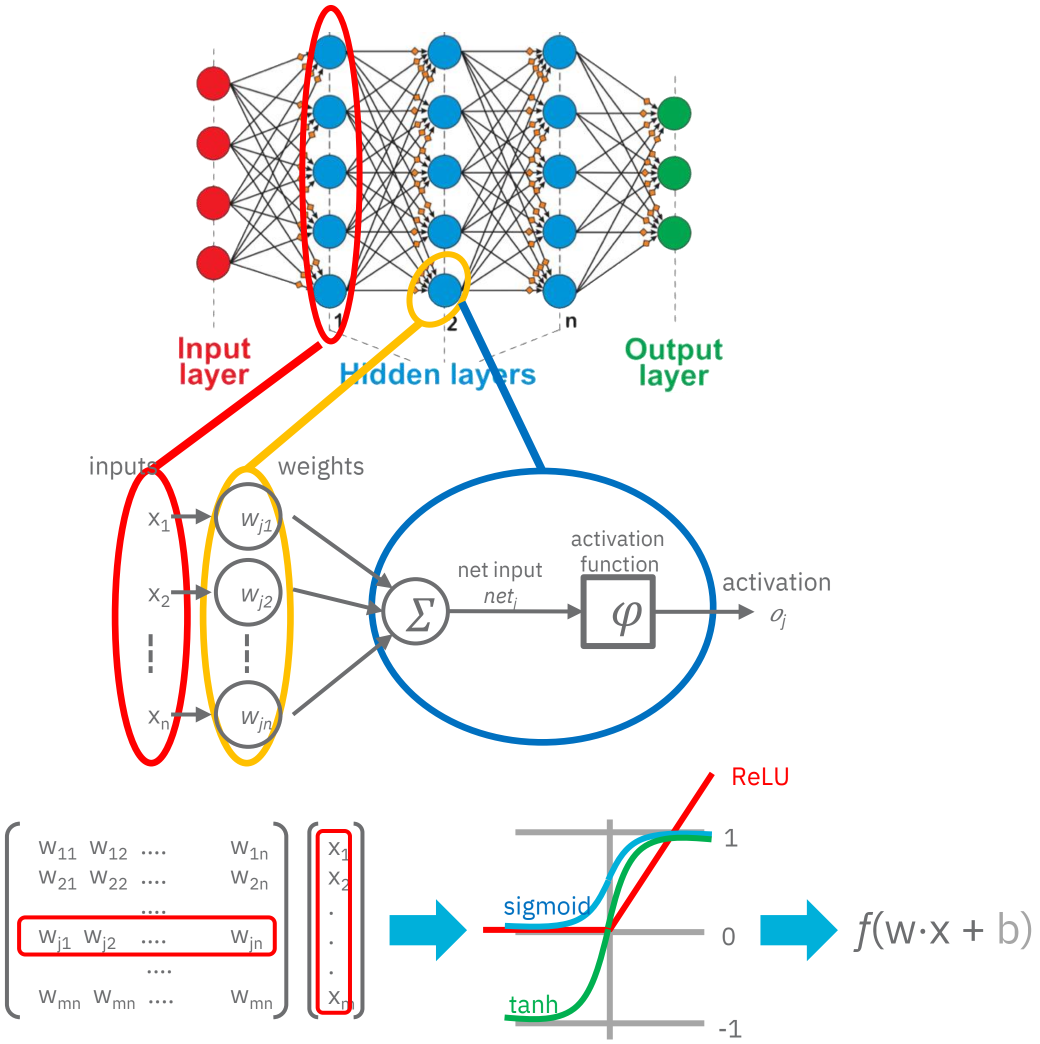
Outline

- Compute operations in Neural Networks
- Limitations of today's technology
- Analog signal processing & in-memory computing

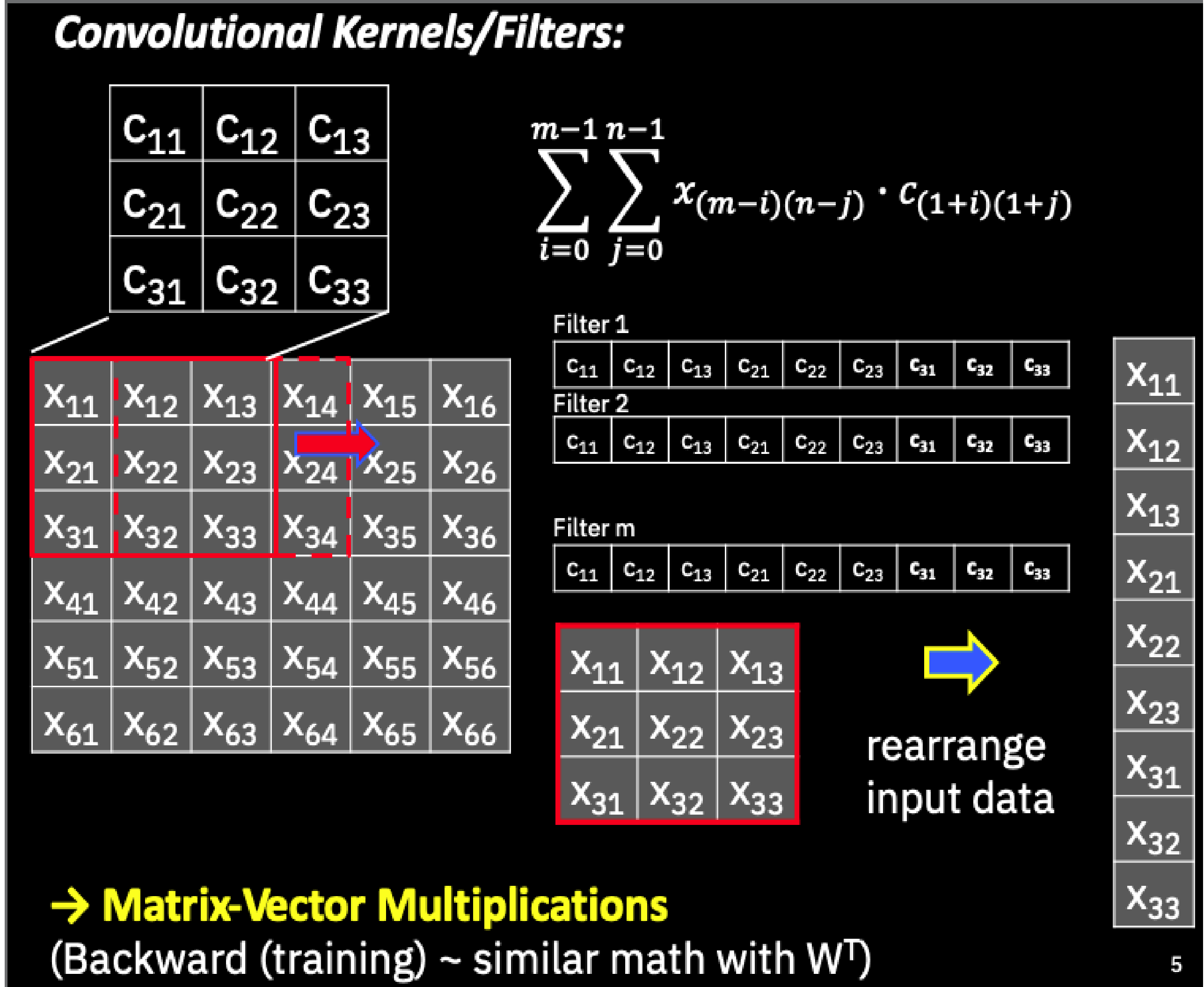
- Optical structures for neural network inference
- Non-linear optics for neural network training

- Conclusions

The math behind



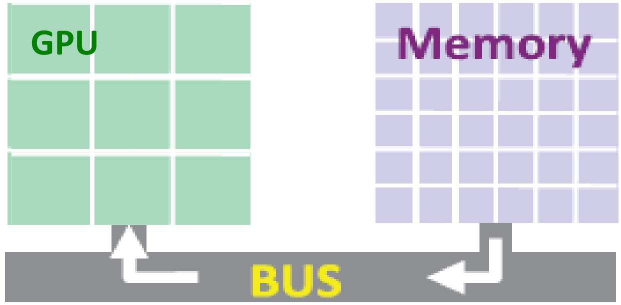
- ***Artificial Neural Networks***
  - Fully connected layer = m x n connections
    $$a = g^{[l]}(W_x x^{(i)} + b_l)$$
  - Activation function (sigm., tanh., ReLU, ...)



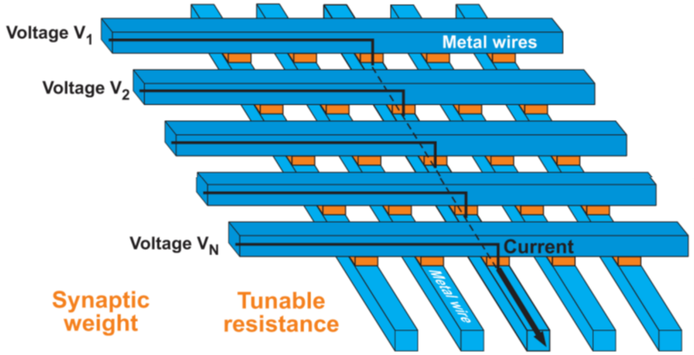
To enhanced performance and efficiency

- **Limiting factors**
  - Memory access
  - Sequential operations
  - Digital signal processing

- **Overcome by**
  - In-memory computing
  - Parallel operations
  - Analog signal processing



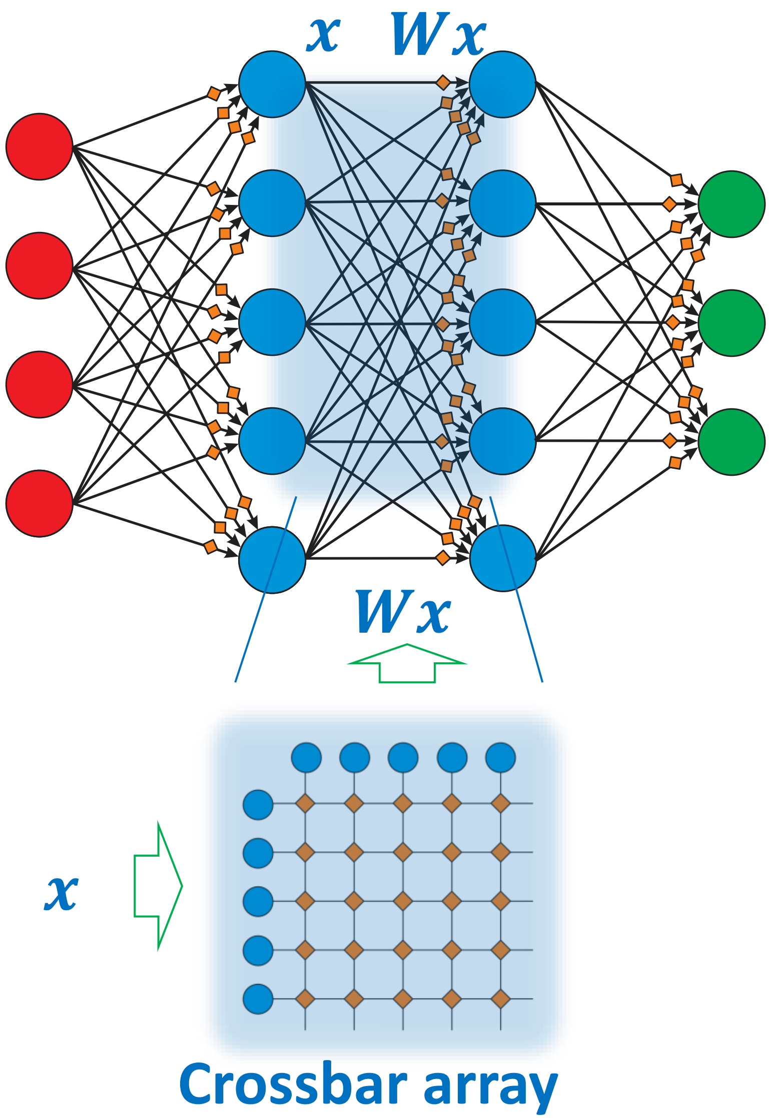
Compute effort ~O(#Neurons^2)



Compute effort ~O(1)

Electrical and optical solutions under investigation

Crossbar arrays – impact on performance and efficiency



Analog signal processing in crossbar arrays improves the neural network performance & efficiency by several orders of magnitude

G. W. Burr et al., “Tech. Dig. - Int. Electron Devices Meet. IEDM, vol. 2016–Febru, no. 408, p. 4.4.1-4.4.4, 2016.
T. Gokmen and Y. Vlasov, Front. Neurosci., vol. 10, no. JUL, pp. 1–13, 2016.

How can optics help?

Basic neural network operations

1. Multiply accumulate
2. Nonlinear function

Assembled in a massively connected architecture

Photonic properties of interest

- The ability to process signals with low latency, real time signal processing.
- Large bandwidth, processing of high-speed data.
- Massive parallelism, i.e. wavelength multiplexing.

Fast signals & Low latency

Light-matter interactions provide additional functionality

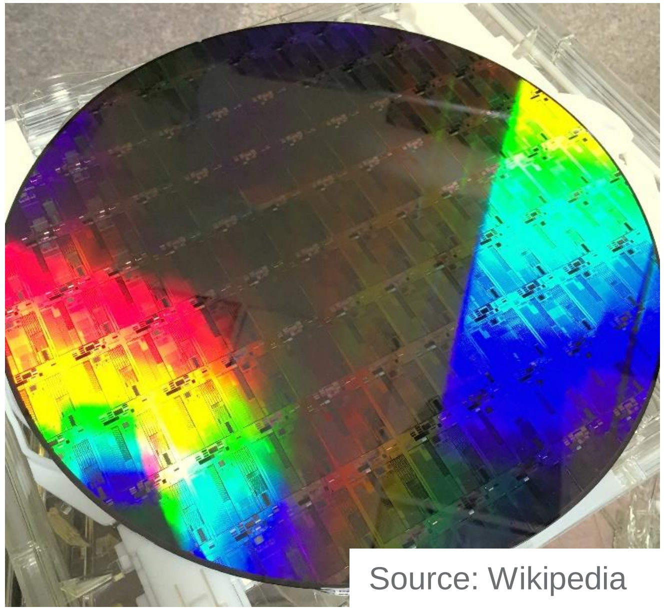
- Electro-optic effect, photonic signal phase control by applying a current or electrical field
- Electro-absorption, impact optical signal transmission by an electrical signal.
- Trimmable refractive index or absorption, persistent change of the material optical properties
- Photorefractive effect, local change of the refractive index through exposure by light

Fast & controlled change

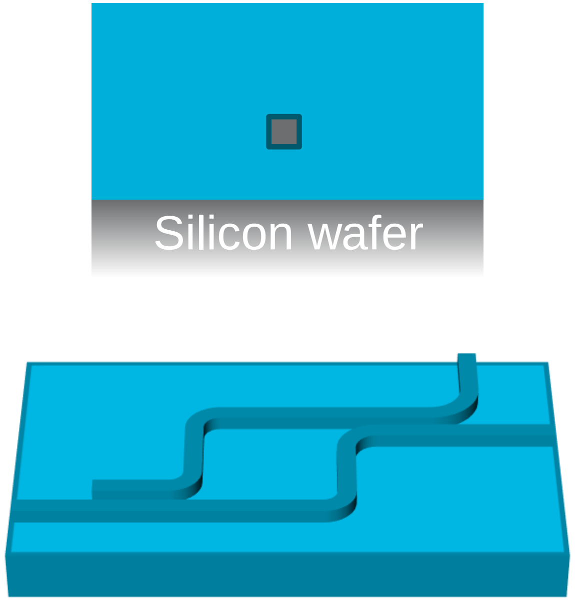
Real time, large bandwidth signal processing & fast 'processor' updating

Silicon photonics

- Emerged as a versatile, scalable and cost-effective platform


Source: Wikipedia

Silicon photonics 300 mm wafer



Silicon on insulator structures

Wide range of integrated devices:
Passive devices
Modulators
Detectors
Lasers

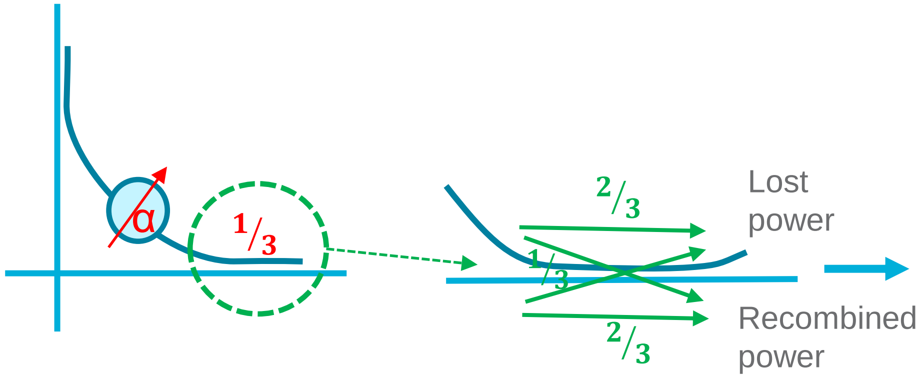
Co-assembly with other functions:
Drivers
Amplifiers
Digital signal processors

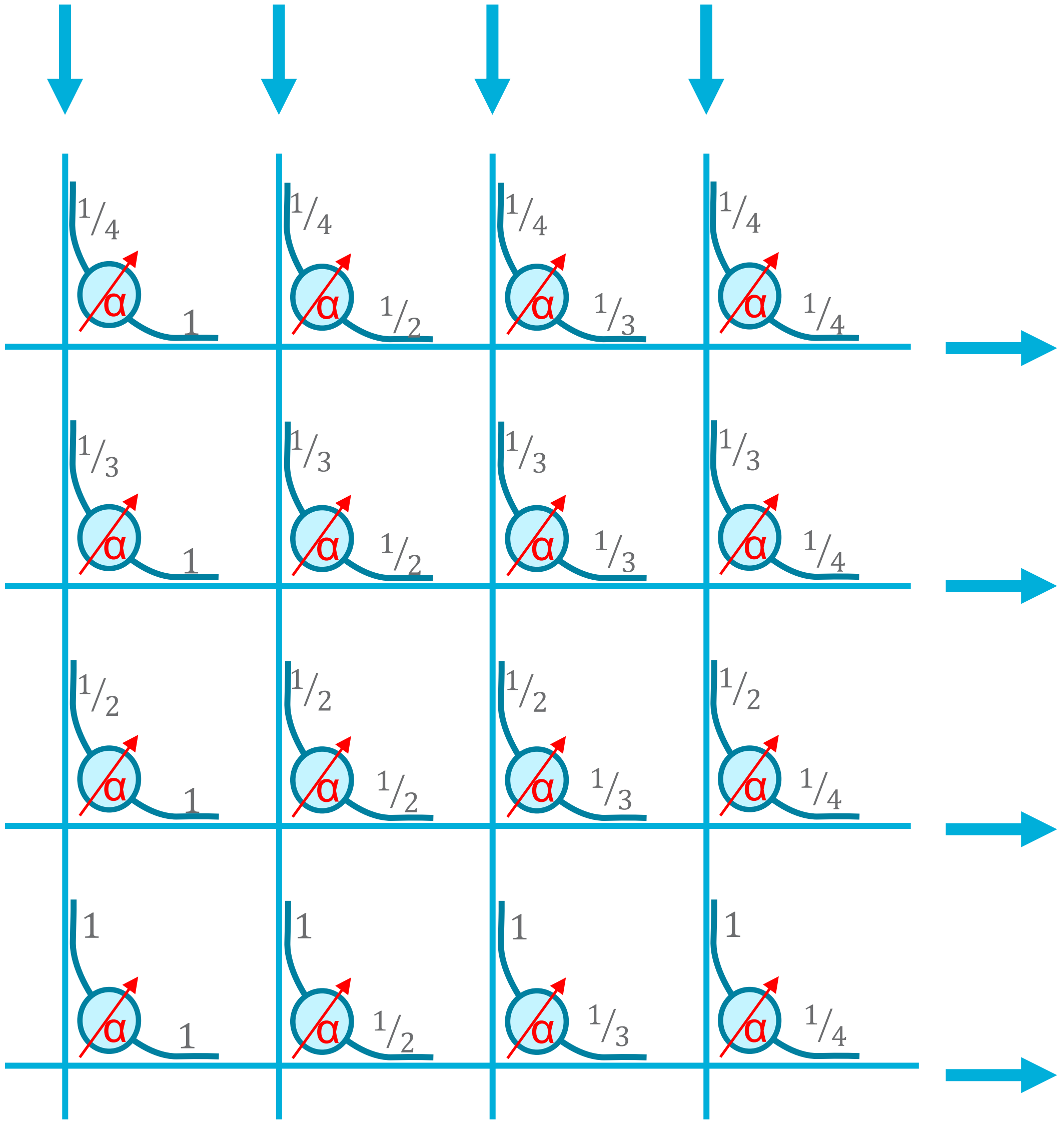
PDK's available at several foundries

- Co-integration of new materials widens the functionality

Optical crossbars

- Fixed signal distribution along columns
- Fixed signal accumulation along rows
- One tunable attenuator per intersection/coefficient:
  - N^2 heaters
  - Simple control
- But: Power loss (factor 1/N) in the directional couplers for signal accumulation along the output rows:

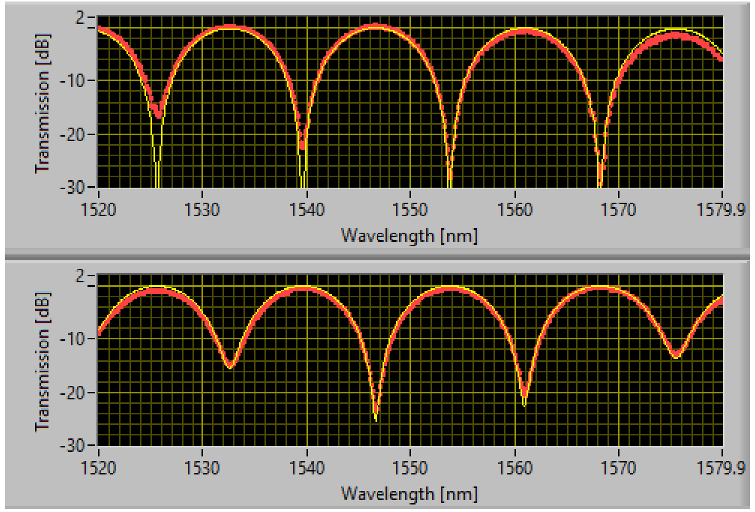




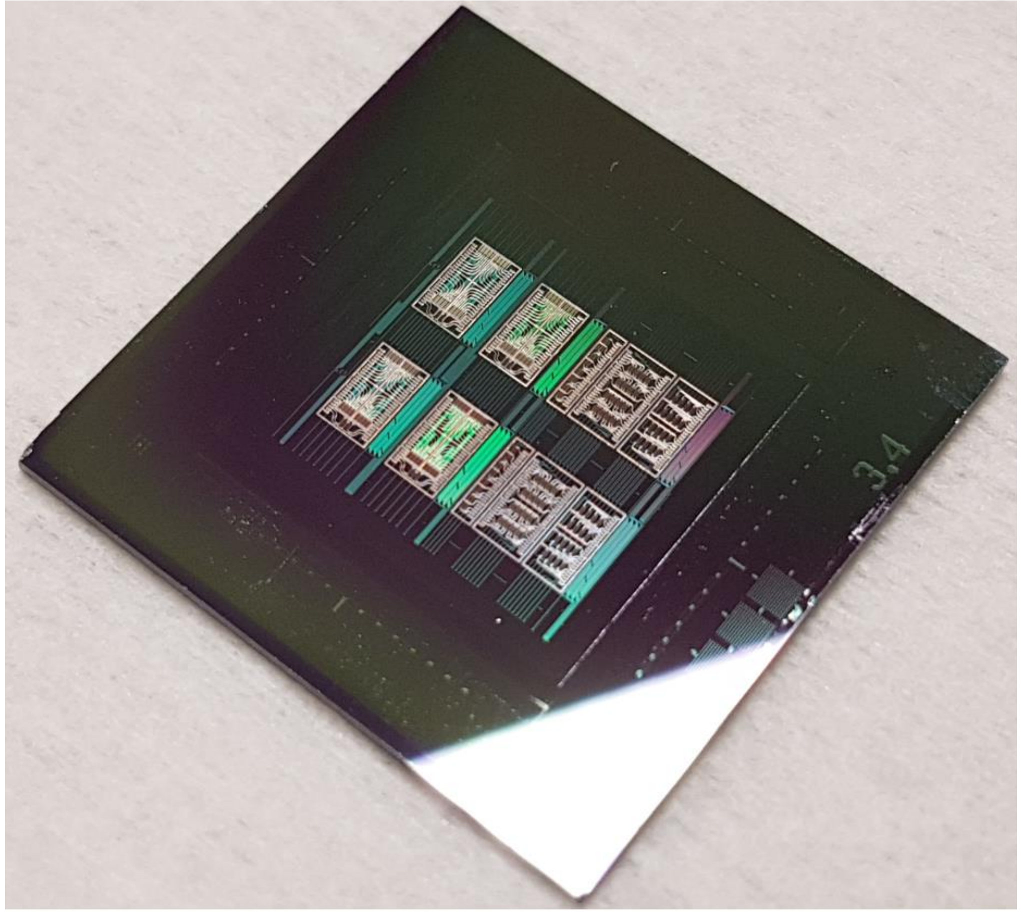
Silicon photonics synaptic chip

- Characterization of test structures:
  - Waveguide loss: ~2.8 dB/cm
  - Directional couplers: match design specs

Test MZI transmission

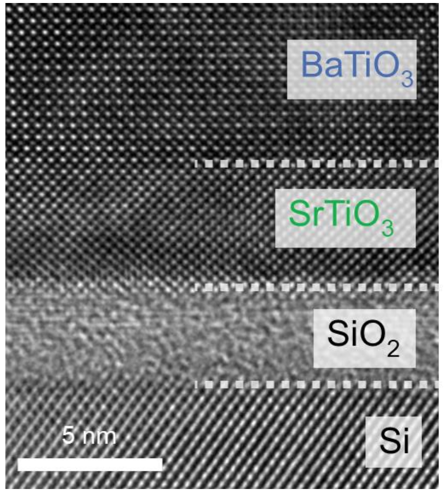


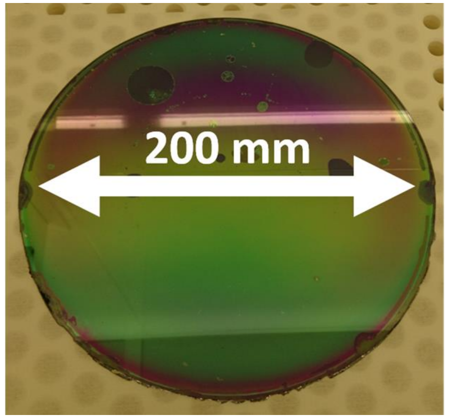
ChipAI die

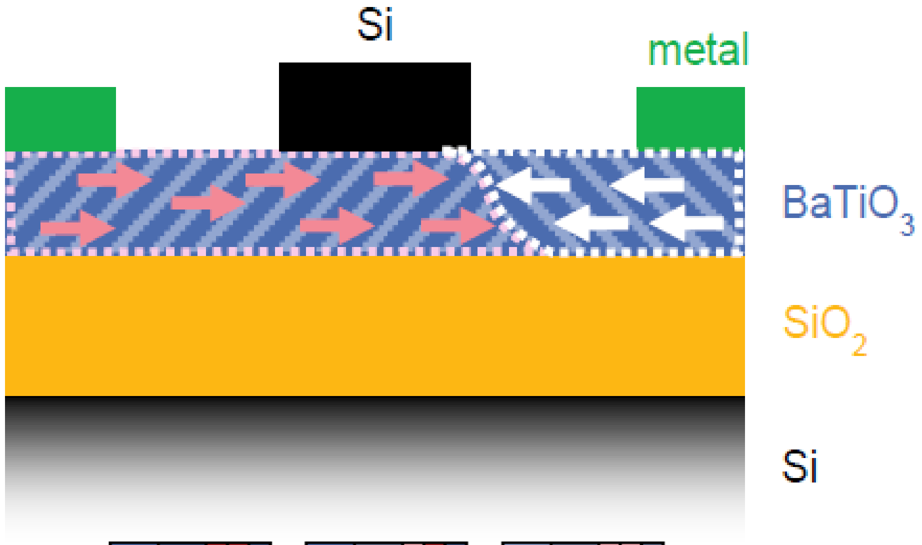


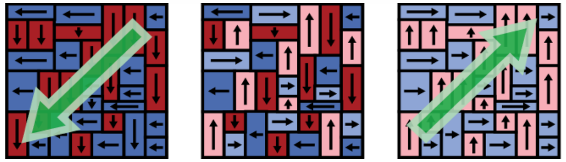
Non-volatile optical weights

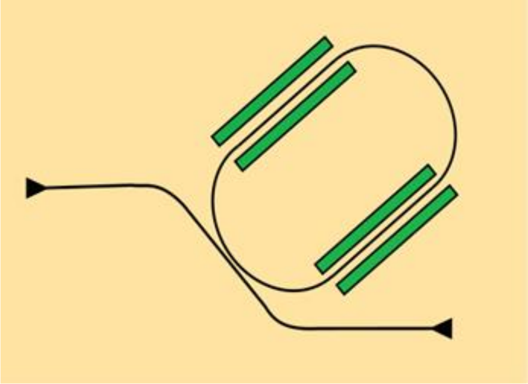
- Silicon photonics emerged as a versatile integrated optic platform









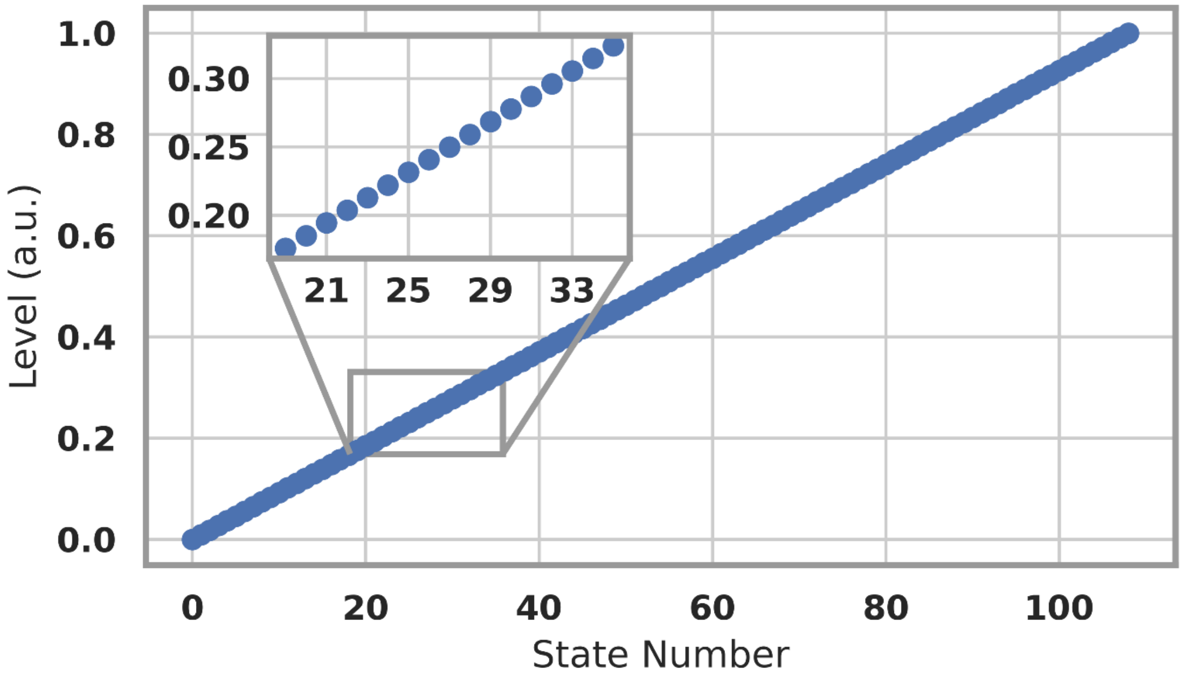


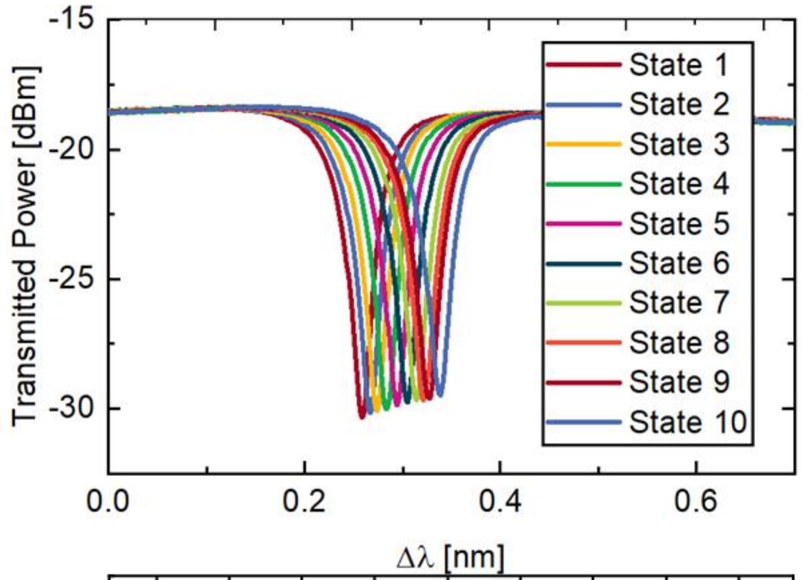
Store photonic weight in the polarization state of a ferro-electric material integrated on silicon photonics

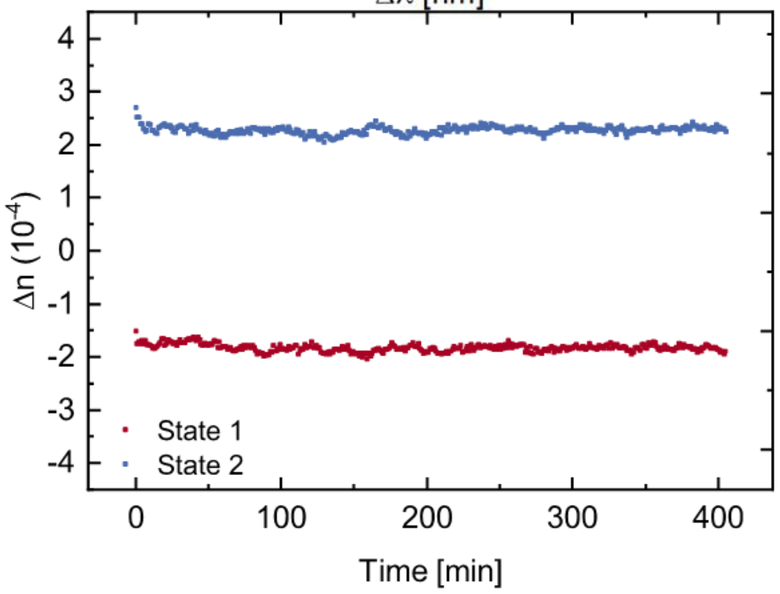
Abel et al., Nanotechnology 2013, 24, 285701

Non-volatile optical weights

- Tuning the effective Pockels coefficient
  - 100 discrete levels can be addressed
  - Retention time larger than 400 min



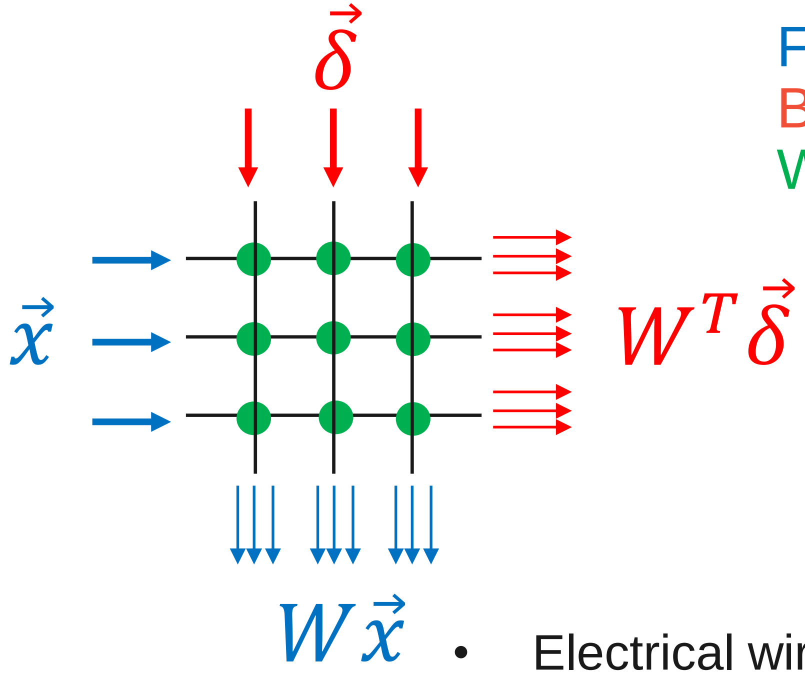




Can address 100 levels in feedback loop operation, 10 through direct access

Photonic crossbar unit - operating principle

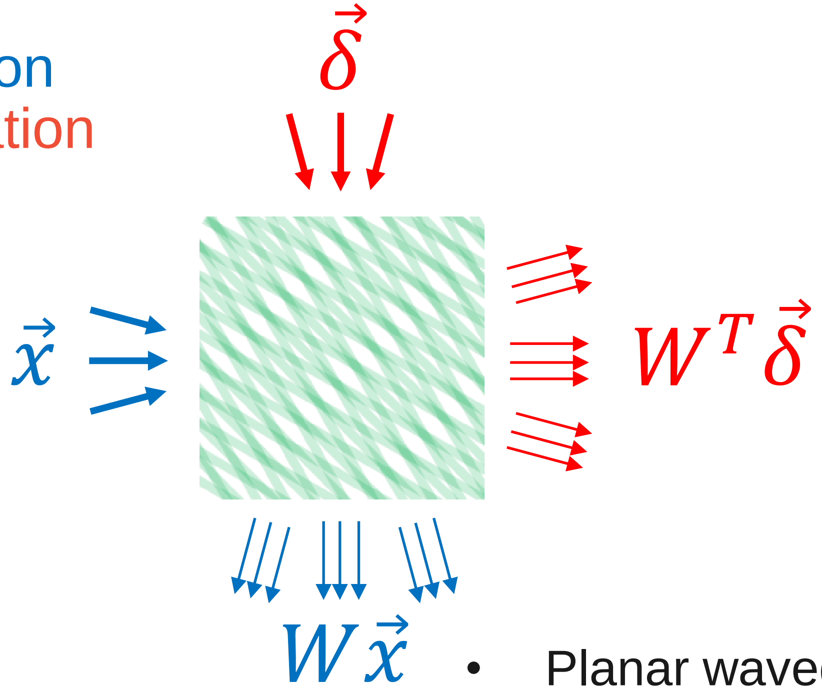
Electrical crossbar



- Electrical wires
- Local weights
- Resistance tuning

Forward propagation
Backward propagation
Weights

Photonic crossbar



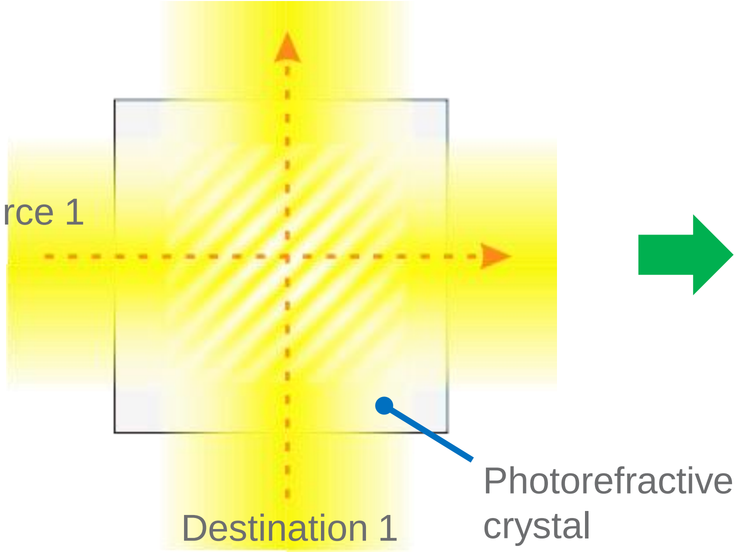
- Planar waveguiding
- Distributed weights
- Refractive index tuning

Writable photorefractive gratings provide the same functionality as the tunable resistive elements in a crossbar unit

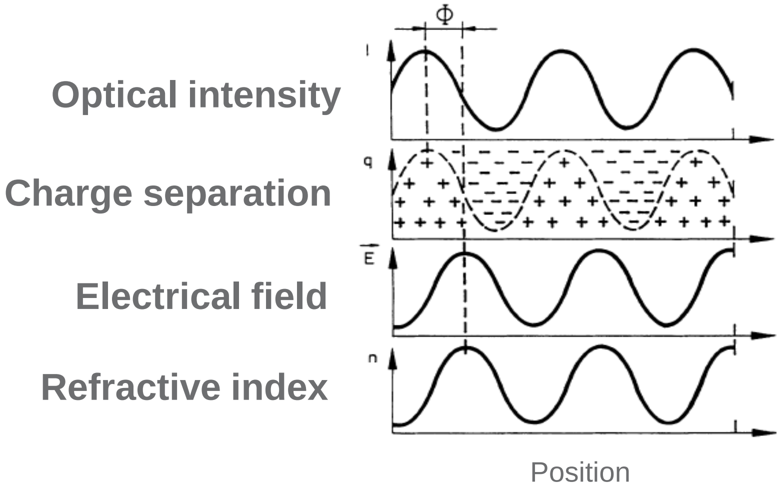
Holographic storage and signal processing

Weight Storage:

Interference pattern:

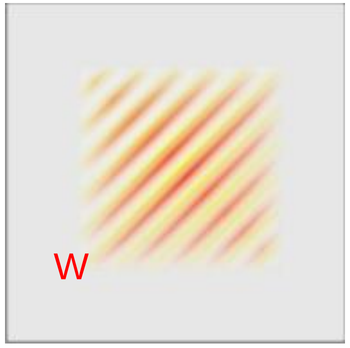


Photorefractive effect:



Cornelia Denz, Optical Neural Networks, 1998.

Stored diffraction grating:



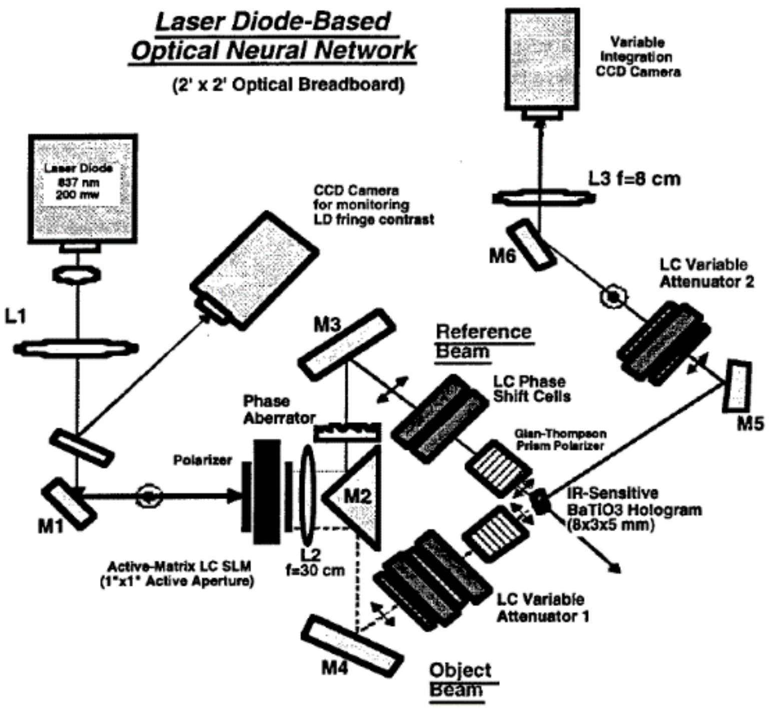
Synaptic weights are stored as refractive index gratings in a photorefractive material:

- Grating are written by two interfering optical beams
- Photorefractive effect: Optically active electron traps + Pockels effect → refractive index grating
- Linear and symmetric process

Optical crossbar arrays: Integrated Solution

Concept demonstrated in bulk optics

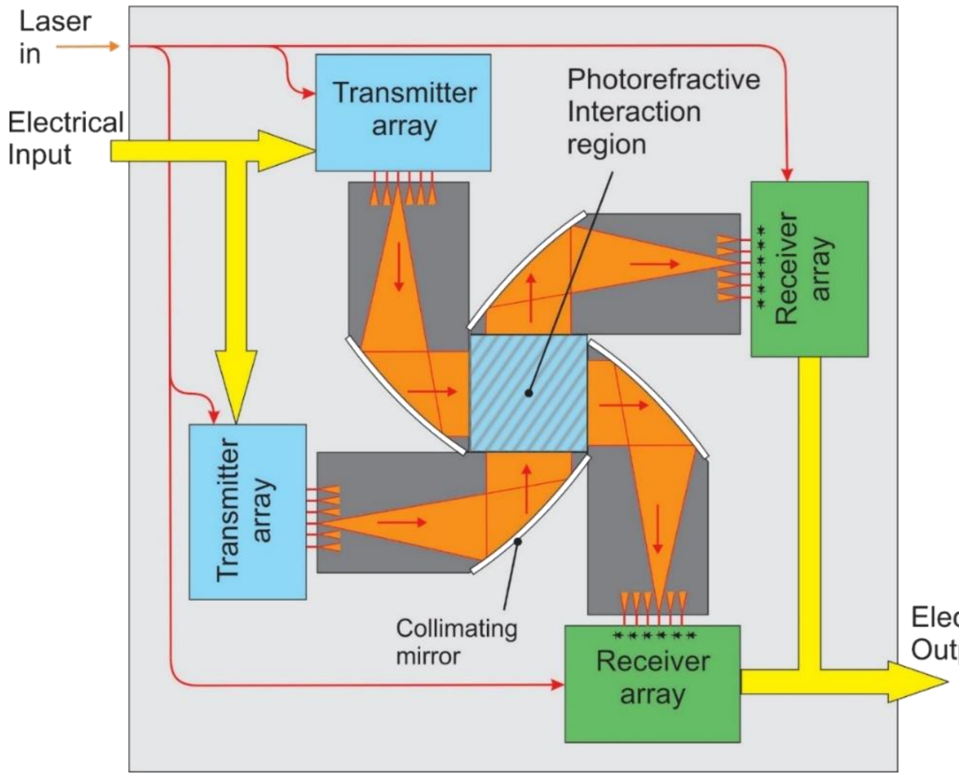
- Backpropagation training of neural networks with hidden layers
- Large setup, slow electro-optics, stability issues



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

Our approach: Integrated Optics

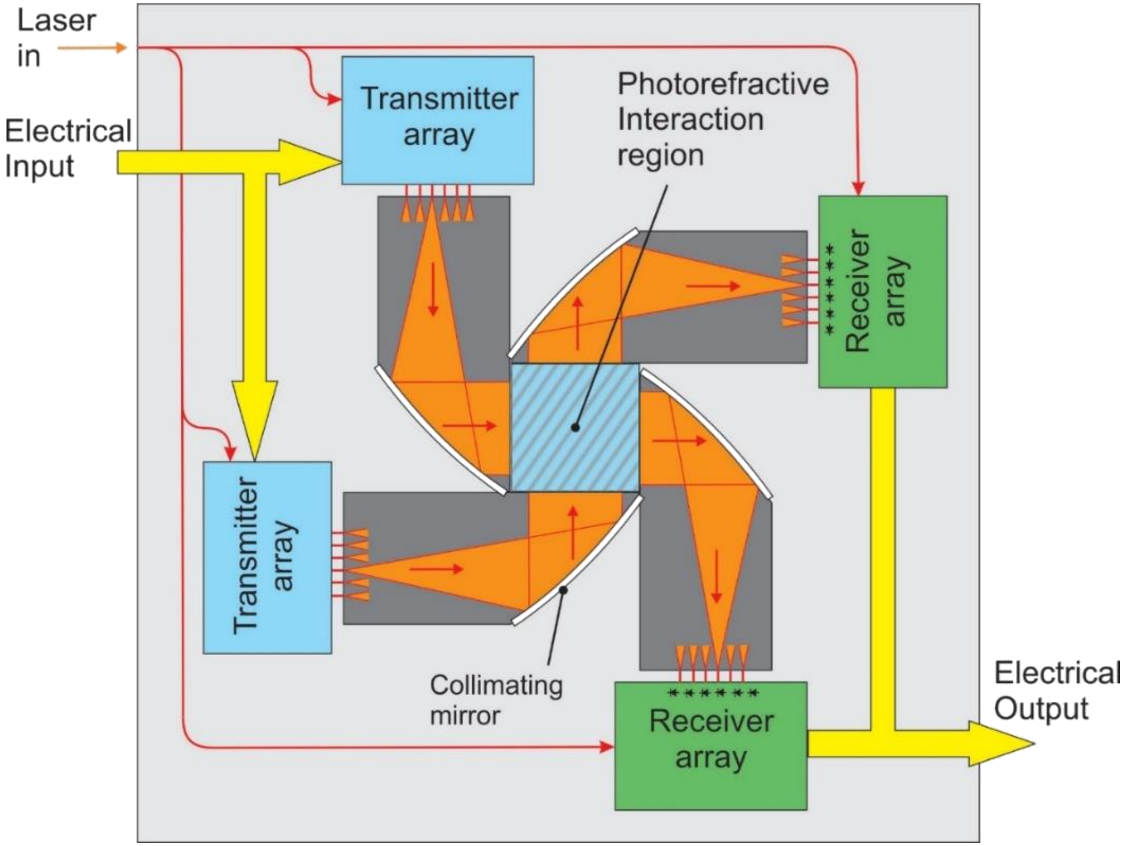
- Electro-optic conversion and beam shaping optics on a silicon photonics chip
- Memory: Photorefractive thin film on silicon



Next steps

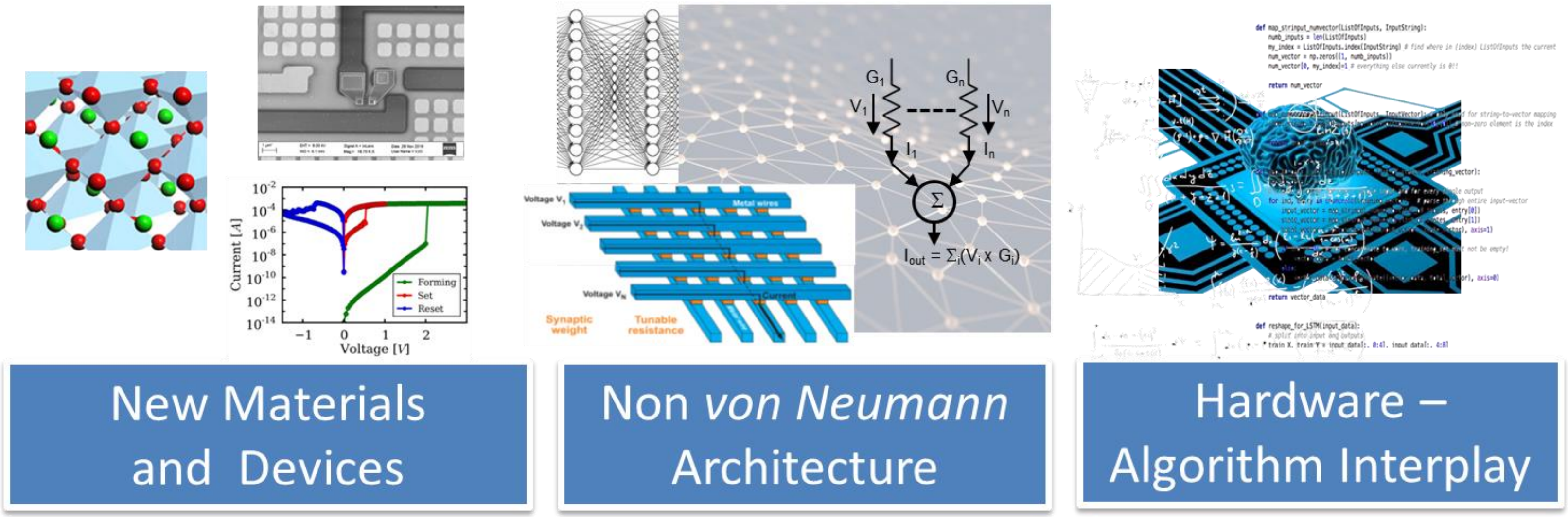
- **Demonstrate small size integrated optical crossbar**
  - 8 x 8 array
  - Holographic weight storage
  - Inference, backpropagation, weight update

- **Implement optical crossbar in neural network**



Summary and conclusions

- To overcome the bottlenecks in today's systems
  - In-memory computing, parallelism, analog signal processing
- Photonics offers exciting prospects for neuromorphic computing
  - High bandwidth signal processing
  - Parallelism
  - Accurate and fast light-matter interactions
- Comparison to other neuromorphic computing platforms
  - Resistive crossbar arrays
- Innovation is required at all-levels



Acknowledgments

IBM Research – Zurich, Switzerland
R. Dangel, F. Horst, P. Stark, J. Geler-Kremer, D. Caimi, K. Moselund, M. Sousa, H. Siegwart, D. Jubin, A. La Porta, J. Weiss, U. Drechsler

Co-funded by
The European Union Horizon 2020 Programme &
The Swiss National Science Foundation

EU: NEUROTECH, ChipAI, PLASMONIAC, NEBULA, NEOTERIC, Post-Digital, PHOENICS
SNF: NAPRECO

Thank you
for your attention

Bert Jan Offrein – ofb@zurich.ibm.com