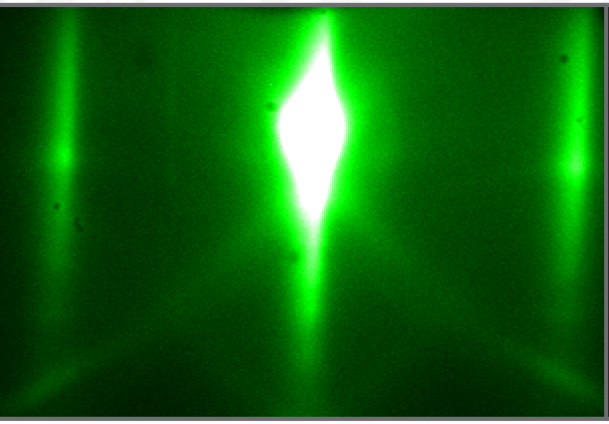
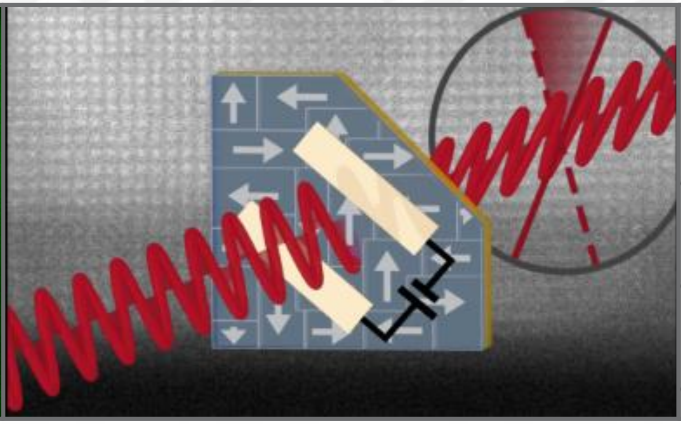
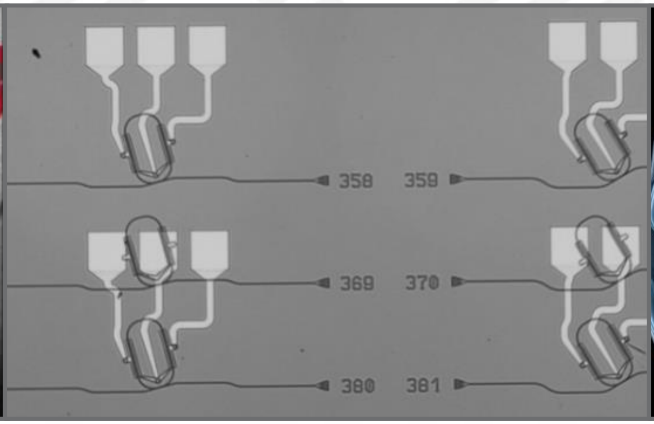
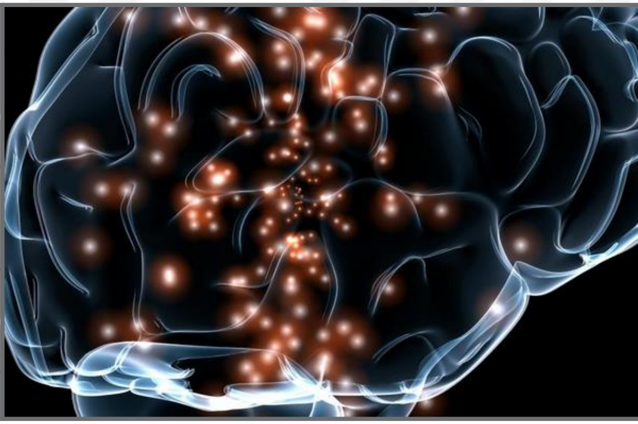
Advancing silicon photonics for traditional and novel computing paradigms

Bert Jan Offrein





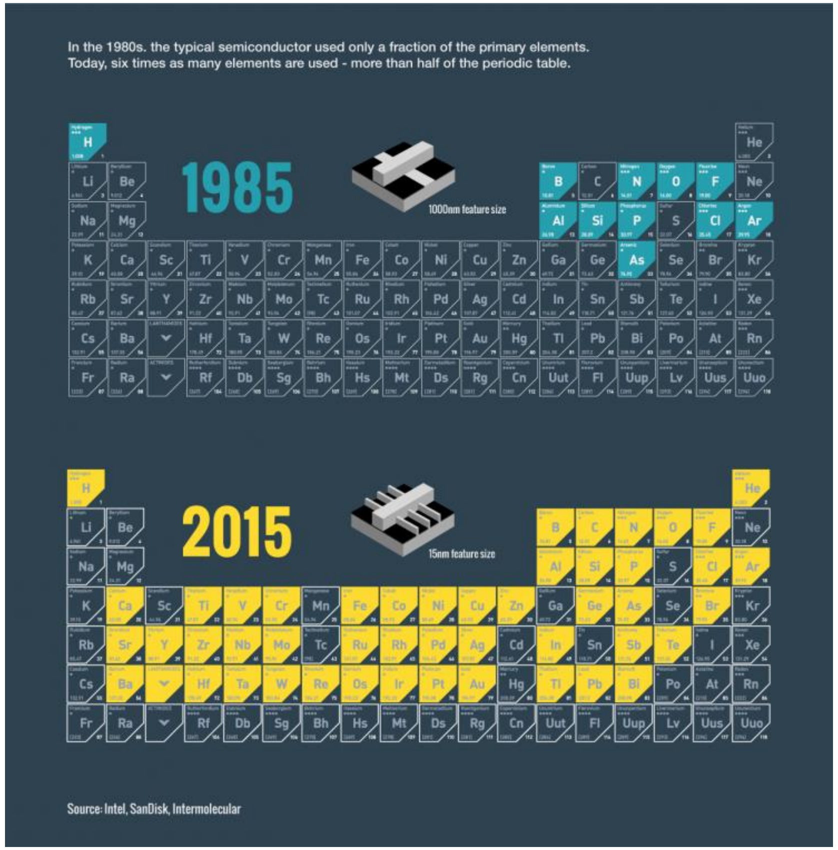




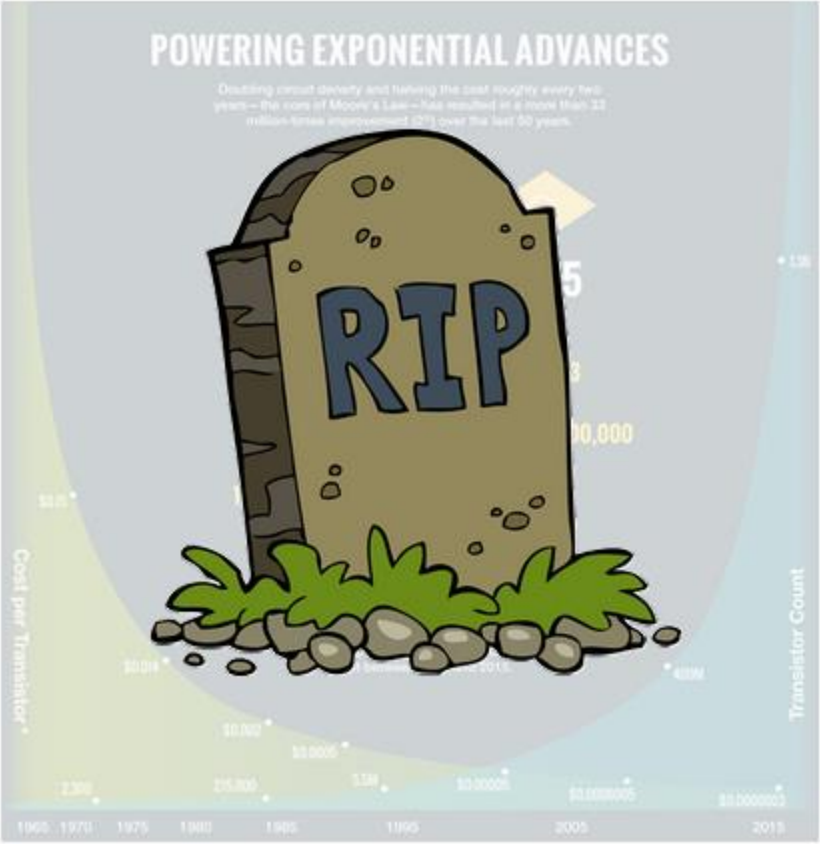
Neuromorphic Devices and Systems Group

Three pillars for Si technology

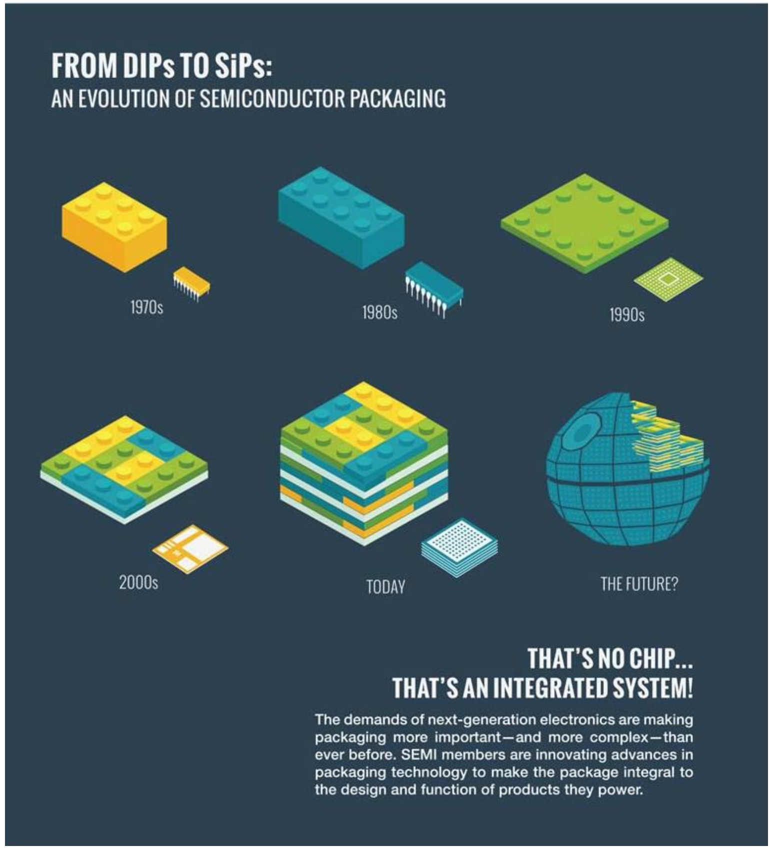
Materials



Scaling



Packaging



www.semi.org

Bert Jan Offrein – ofb@zurich.ibm.com

Experiment: “Human Brain vs. Computer”

Task 1: Mathematics

$$\sqrt{2} = ?$$

Task 2: Image recognition

Traditional silicon scaling ended
New types of problems gain interest

Explore new functionalities, More than Moore
Explore new computing paradigms

Bert Jan Offrein – ofb@zurich.ibm.com

Outline

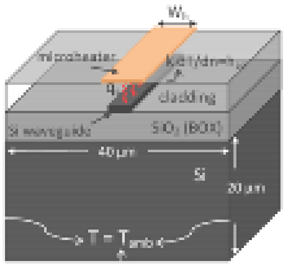
- Facilitate traditional scaling - Optical interconnects
  – CMOS Silicon Photonics
    - Pockels modulators, integrating $BaTiO_3$

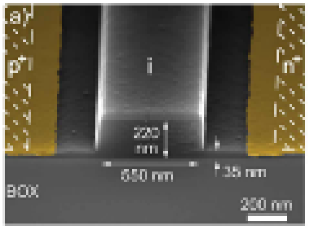
- New computing paradigms – Neuromorphic computing
  – Dedicated hardware for neural networks
    - Photonic Synaptic Processor

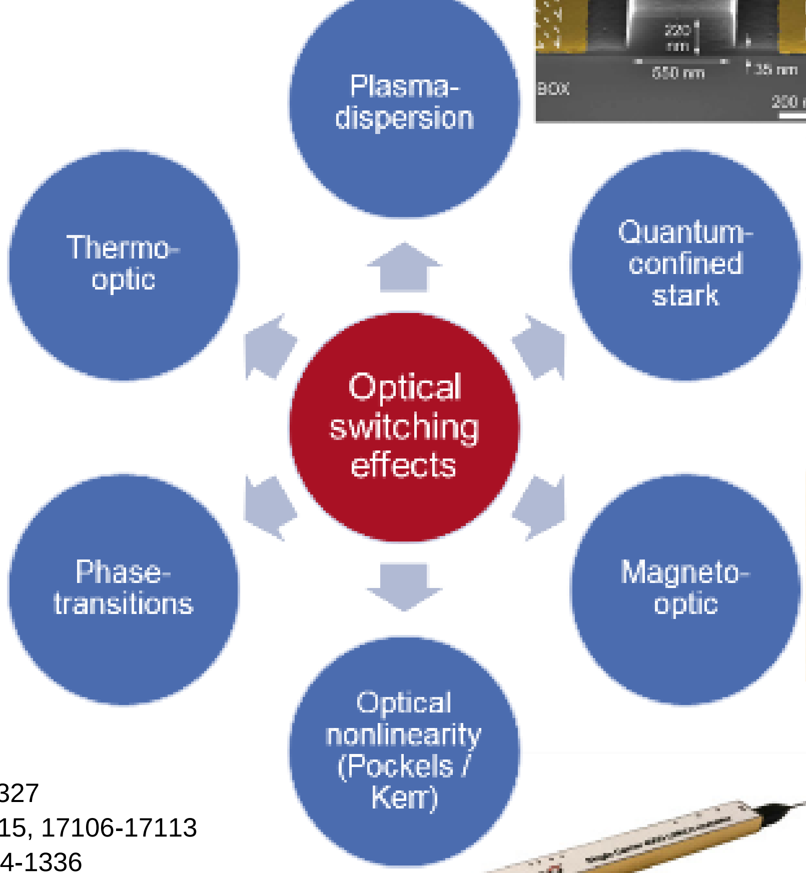
- Conclusions

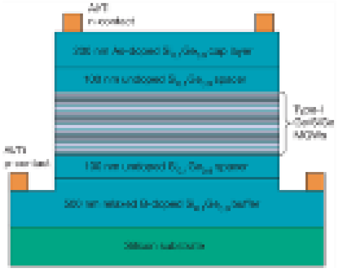
Bert Jan Offrein – ofb@zurich.ibm.com

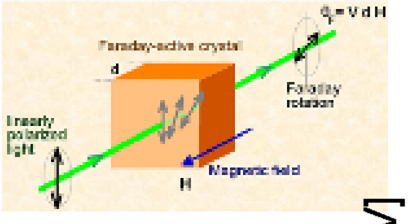
Integrated silicon photonics: modulators

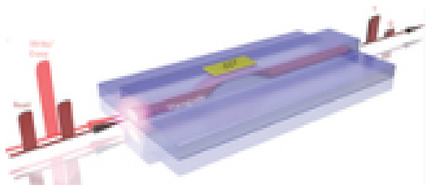




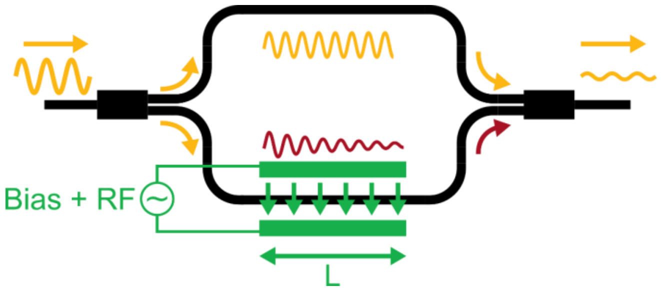




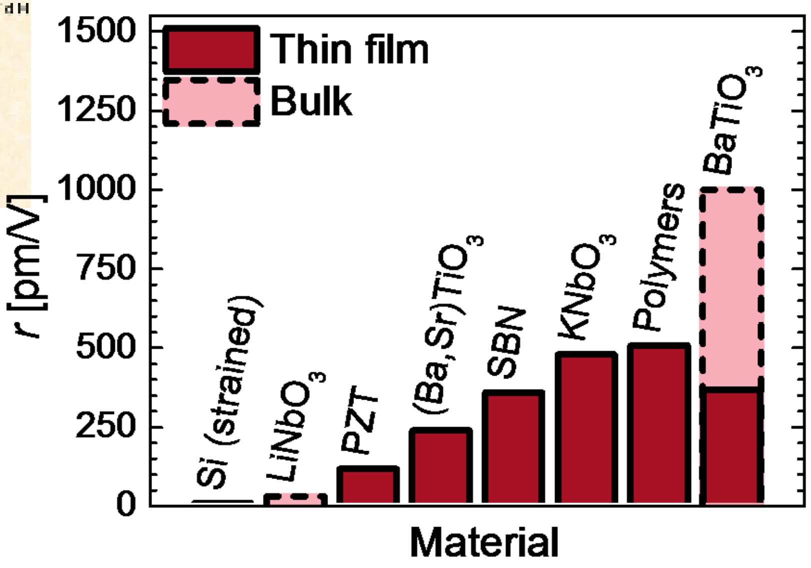




- Xu et al., Nature, 2005, 435, 325-327
- Green et al., Opt. Express, 2007, 15, 17106-17113
- Kuo et al., Nature, 2005, 437, 1334-1336
- Atabaki et al., Opt. Express, 2010, 18, 18312-18323
- Wall et al., Nature Comm., 2012, 3, 721-724
- Rios et al., Nature Phot. 2015, 9, 725-732



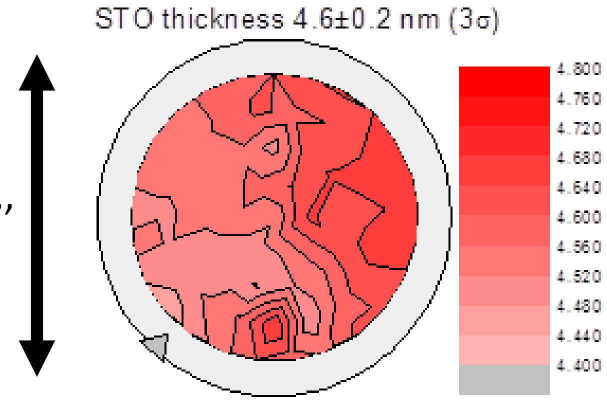
- Wessels, Annu. Rev. Mater. Res., 37, 659-679 (2007)
- Jacobsen, Nature, 441, 199-202 (2006)
- Chmielak et al., Opt. Express, 19, 17212-17721 (2011)
- Bernasconi et al., J. Appl. Phys., 78, 2651–2658 (1995)
- Warusawithana, Science, 324 (2009)
- Petraru et al., Appl. Phys. Lett., 81 (2002)



Bert Jan Offrein – ofb@zurich.ibm.com

Barium titanate on silicon

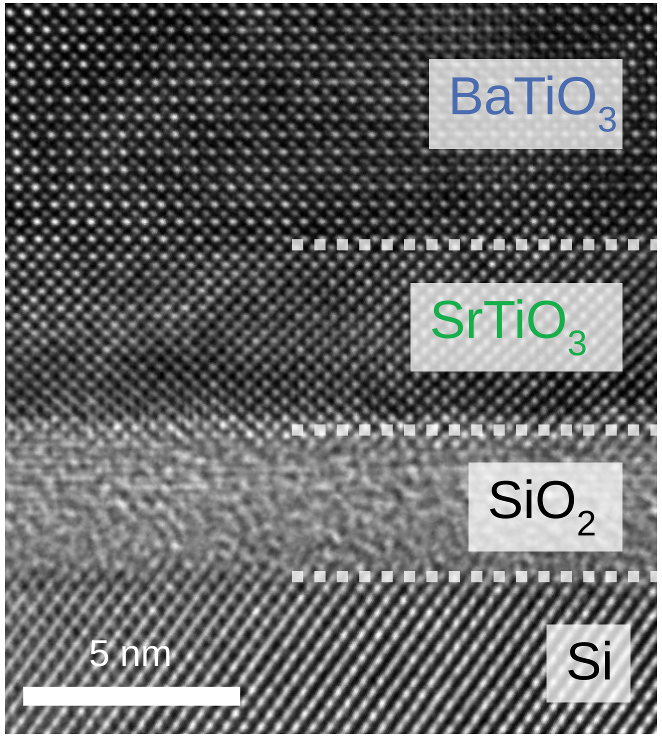
- Molecular beam epitaxy (MBE) or other (PVD, PLD, CVD)
- $BaTiO_3$ deposition on $SrTiO_3$ / Si templates
  - Tetragonal crystal structure
  - Thickness dependant polarization
- Availability on 200mm



Thickness variations of the 4 nm thick film across the wafer

$SrTiO_3$

Si

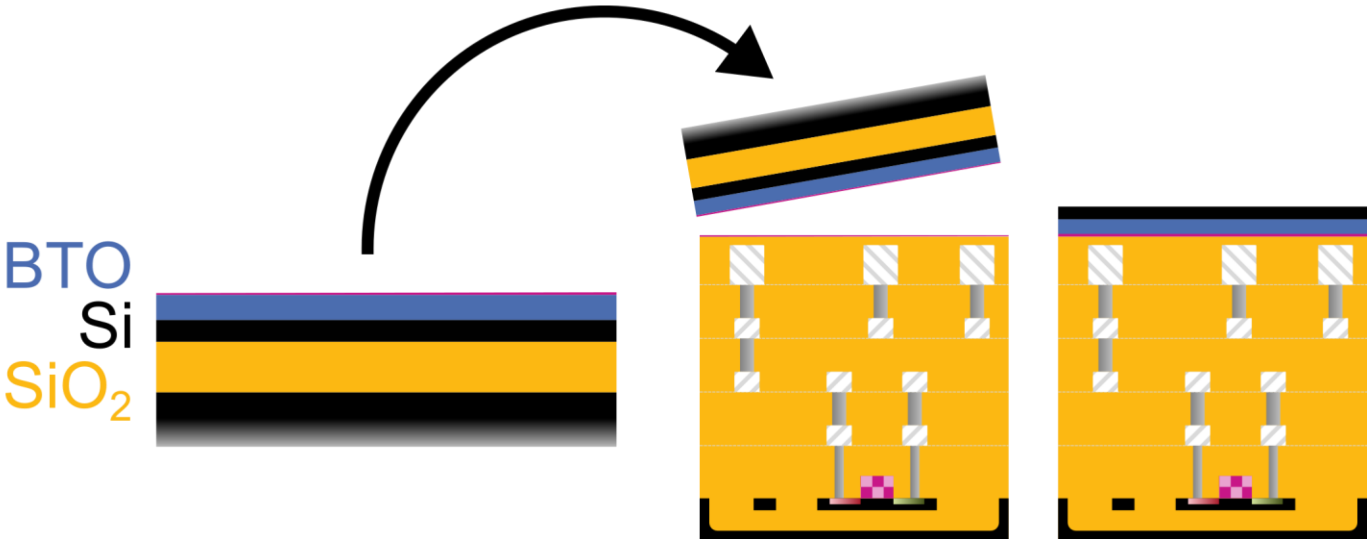


Abel et al., Nanotechnology 2013, 24, 285701

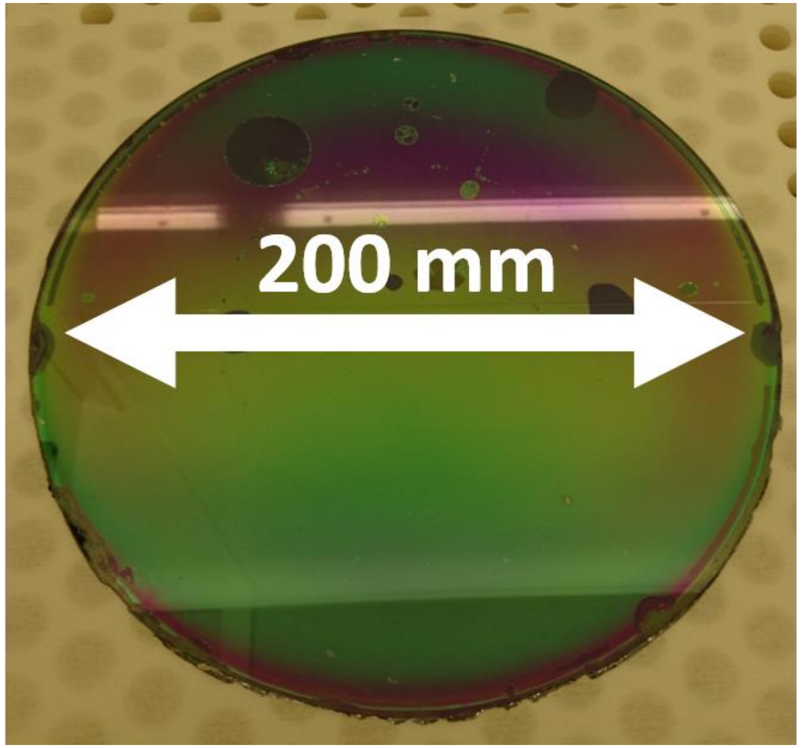
Bert Jan Offrein – ofb@zurich.ibm.com

BaTiO3 integration

- Back-end-of-line integration
  - Direct epitaxy not possible on interlayer dielectric
- Transfer by wafer bonding
  - Back-end-of-line compatible
  - Thermal budget <400°C



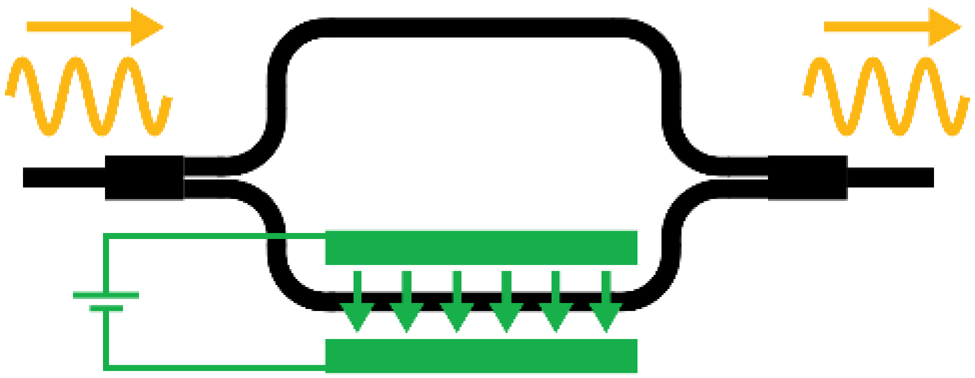
Eltes et al., IEDM, 2017

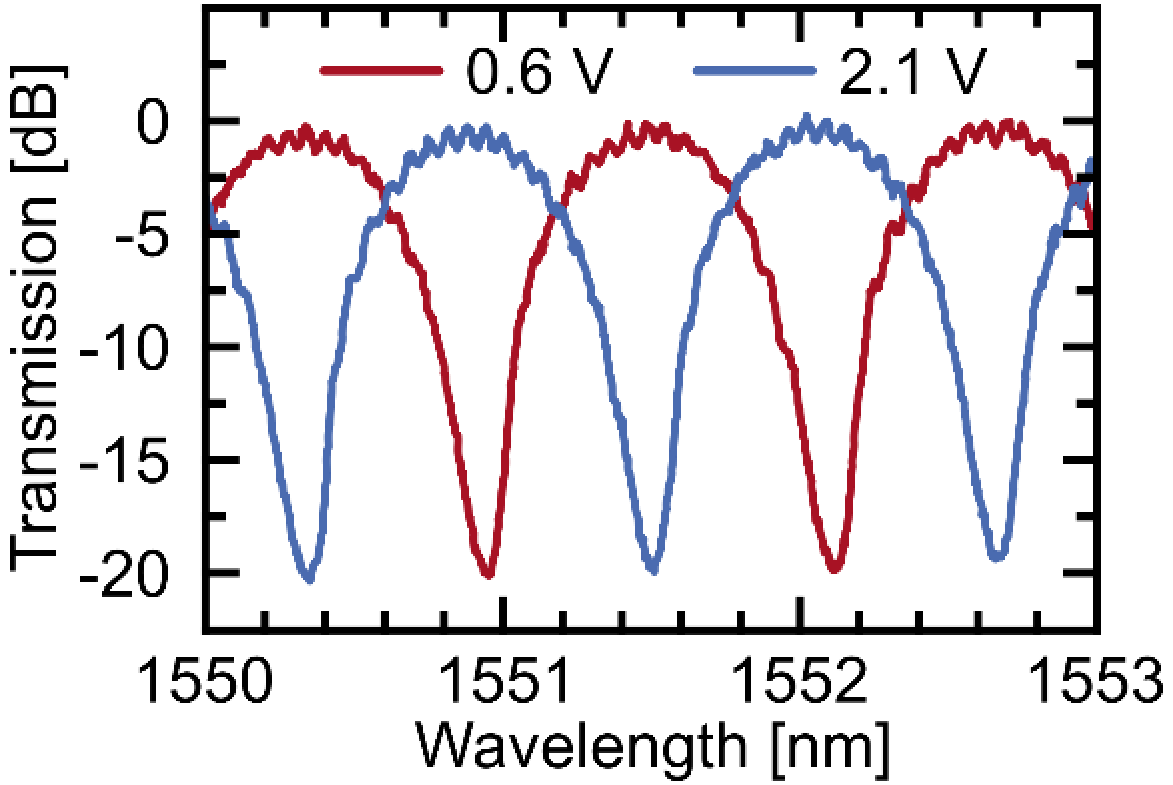


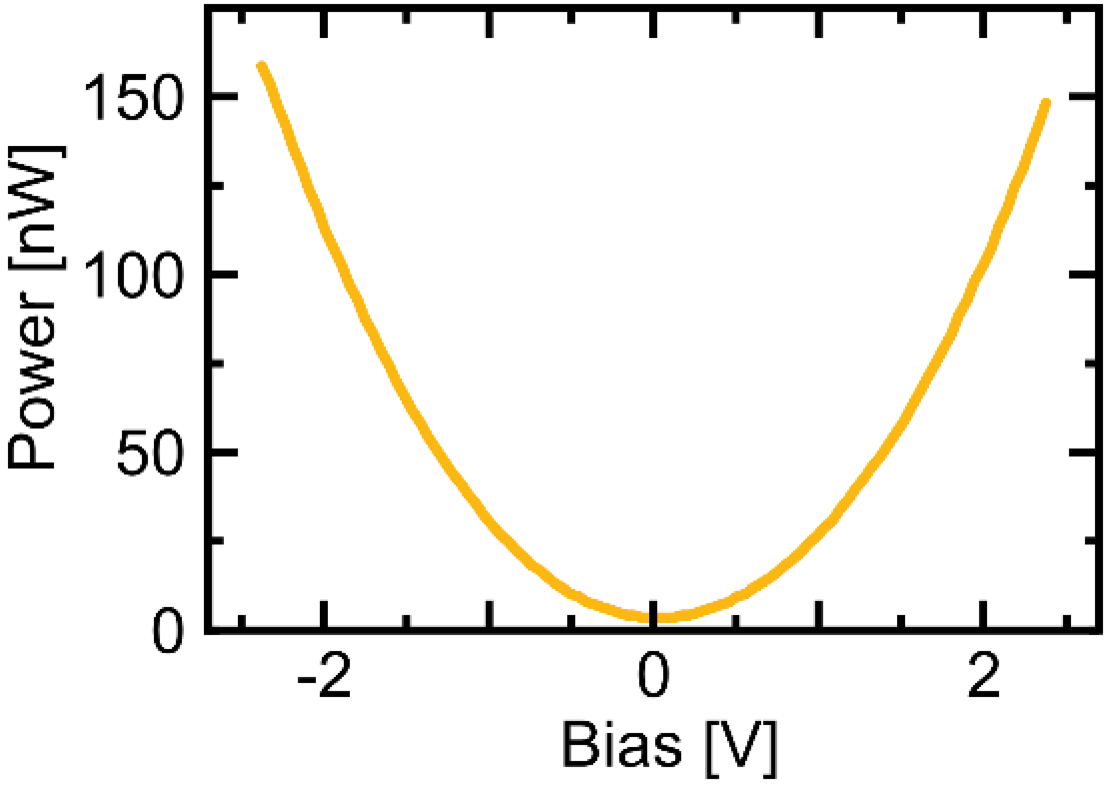
Bert Jan Offrein – ofb@zurich.ibm.com

Active devices : DC characterization

- Voltage induced phase shift
- Very low DC power





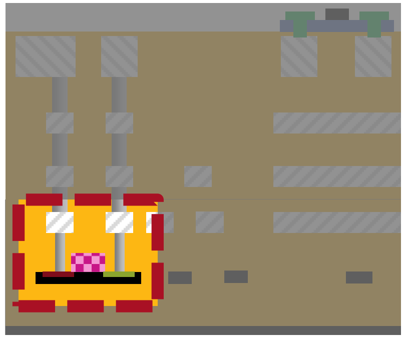


F. Eltes et al., IEDM Tech. Dogest 2017

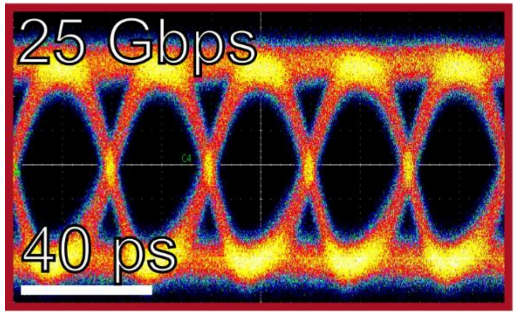
Bert Jan Offrein – ofb@zurich.ibm.com

Impact on FEOL components

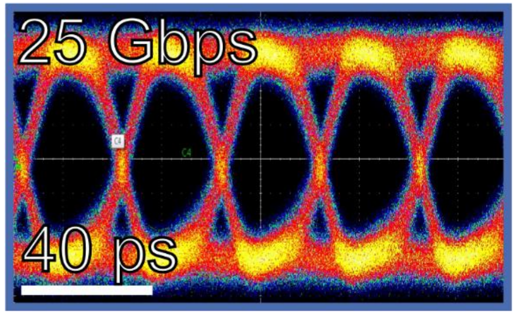
- No degradation of Ge photo diodes

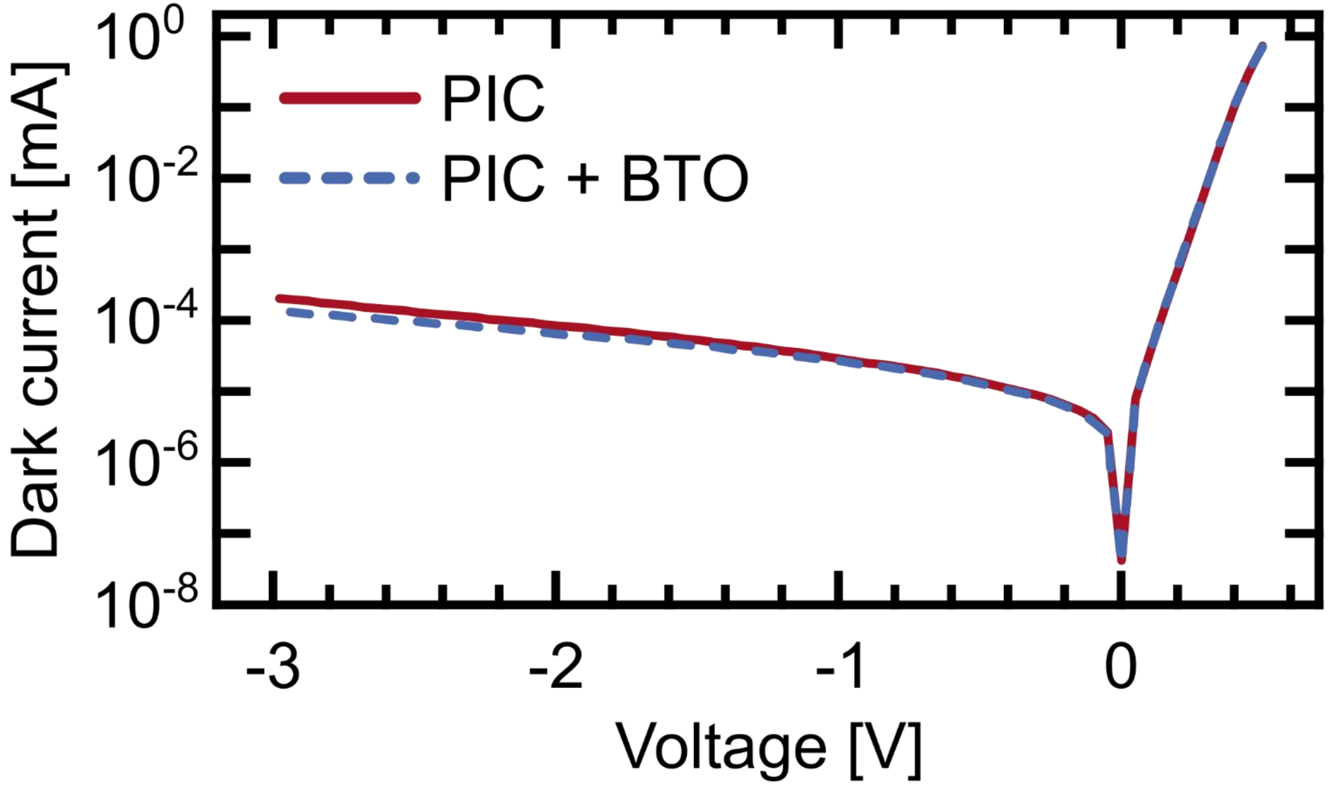


PIC



PIC + BTO





Eltes et al., IEDM, 2017

Bert Jan Offrein – ofb@zurich.ibm.com

Benchmark

Principle	Carrier modulation			Pockels
Structure	p-n	MOS	EAM	
Material	Si[1]	Si[2]	Ge[3]	BTO[6]
$V_{\pi}L$ [V·cm]	2.8	0.2	-	0.3
$V_{\pi}L\alpha$ [V·dB]	27	13	-	1.7
BW [GHz]	40	-	>50	2/20
Integration	+	+	+	+
Phase modulation only	-	-	-	+

Pursuing applications in

- High-speed modulators
- Ultra-low optical device tuning (pW)
- Cryogenic eo modulators for optical interfacing to quantum computers
- Non-volatile optical weights in photonic neuromorphic systems

[1]Patel, Opt. Exp. 2015 [2]Webster, OFC, 2015 [3]Srinivasan, JLT, 2016 [6]Eltes, IEDM, 2017

Bert Jan Offrein – ofb@zurich.ibm.com Felix Eltes – fee@zurich.ibm.com

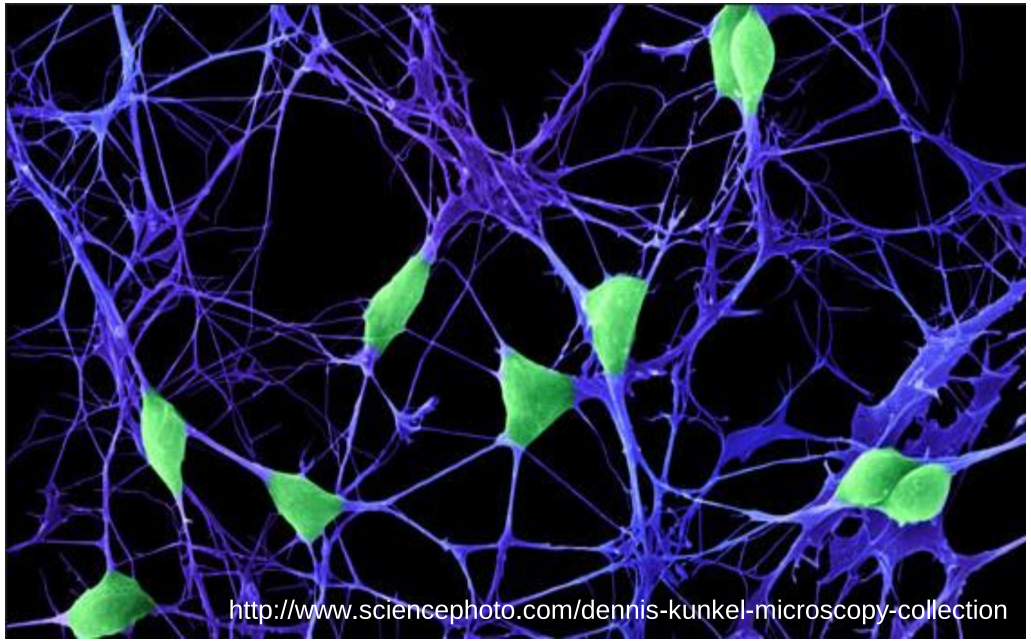
Outline

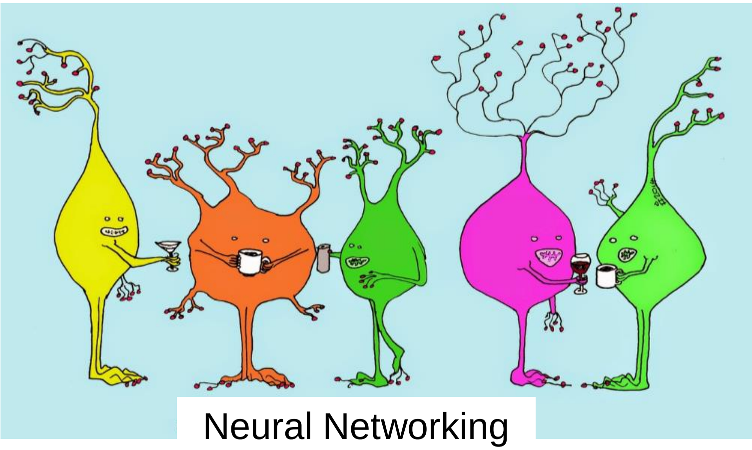
- Facilitate traditional scaling- Optical interconnects
  - CMOS Silicon Photonics
    - Pockels modulators, integrating $BaTiO_3$
    - Embedded lasers and detectors, integrating III-V stacks
- New computing paradigms – Neuromorphic computing
  - Dedicated hardware for neural networks
    - Resistive Synaptic Processor
    - Photonic Synaptic Processor
- Conclusions

Bert Jan Offrein – ofb@zurich.ibm.com

Neuromorphic Computing

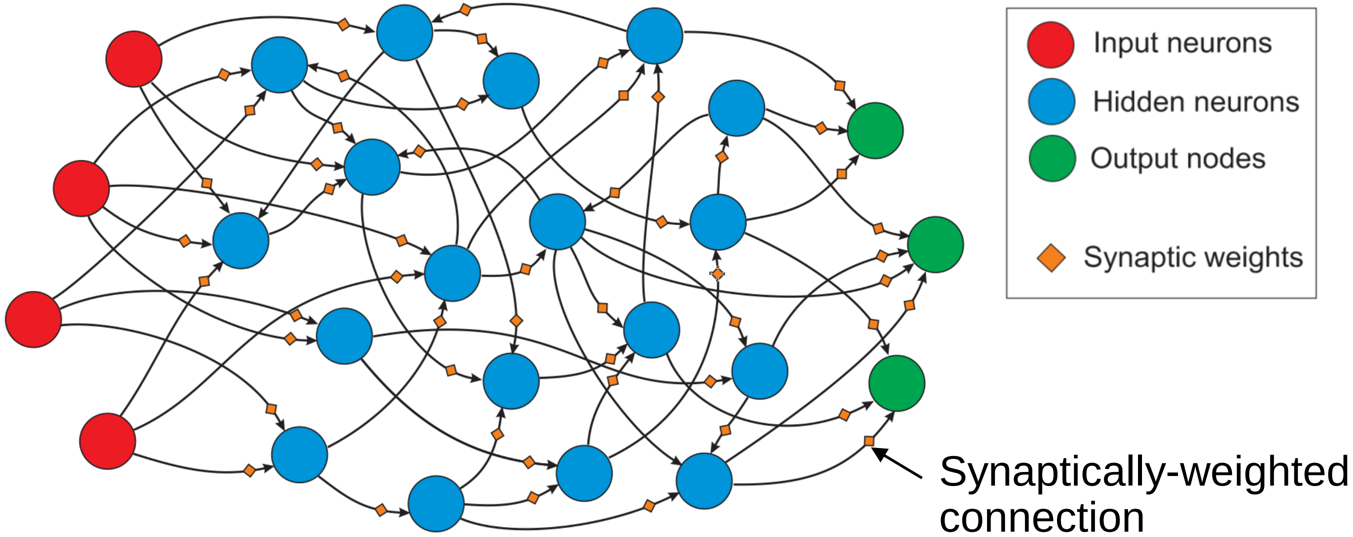
Brain at neural network level





Neural Networking

Complex model of brain-like neural network



This brain model is too complex to be fully and precisely modelled

⇩

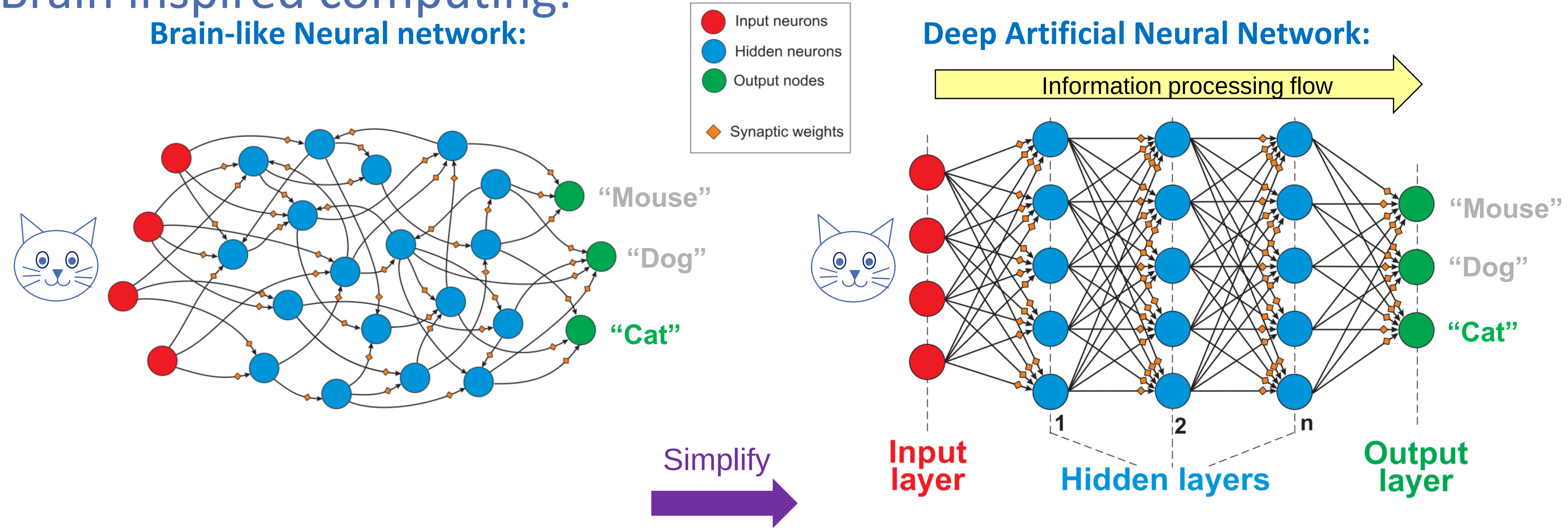
Breakthrough in implementation and applicability of neuromorphic computing came with the concept of using simplified Artificial Neural Networks (ANNs)

Bert Jan Offrein – ofb@zurich.ibm.com

Brain inspired computing:

Brain-like Neural network:

Deep Artificial Neural Network:



- Omni-directional signal flow
- A-synchronous pulse signals
- Information encoded in signal timing

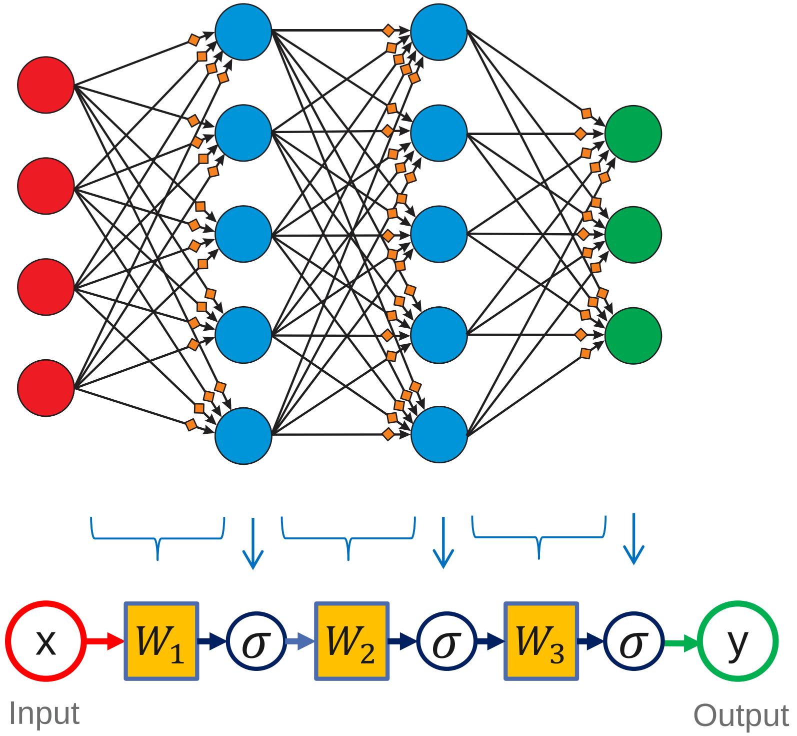
➔ Difficult to implement efficiently on standard computer hardware

- Feed-forward sequential processing
- Information encoded in signal amplitude
- Neuron activation: Accumulate + Threshold
- **Training: Backpropagation Algorithm**

Bert Jan Offrein – ofb@zurich.ibm.com

Training: Backpropagation algorithm

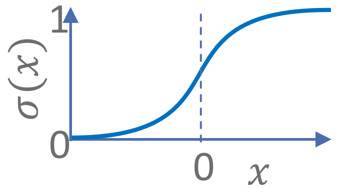
Neural net as chain of vector operations:



→ : Signal vector

W_n : Synaptic weight matrix

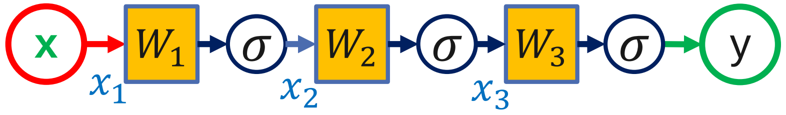
σ : Per-element neural activation function (sigmoid)



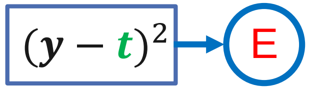
Backpropagation algorithm:

For many training cases **x** with target response **t**:

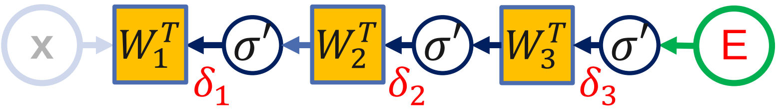
1. Forward Propagate:



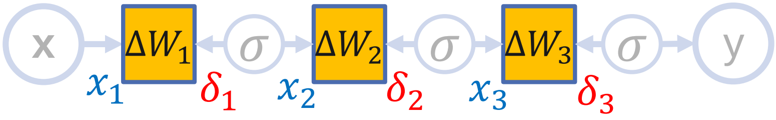
2. Determine output error:



3. Backward Propagate: Determine neuron input influence δ on error **E**:



4. Adjust the active weights, proportional to their influence on the error: $\Delta W = -\eta\, \boldsymbol{x} \otimes \boldsymbol{\delta}$



Bert Jan Offrein – ofb@zurich.ibm.com

Efficient training of Deep Artificial Neural Networks:

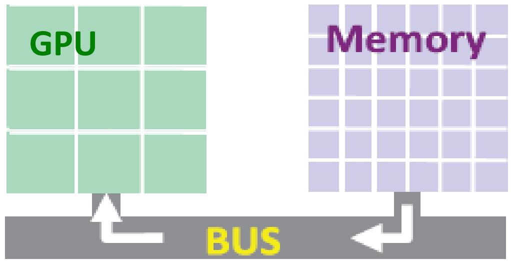
Training by Backpropagation Method:

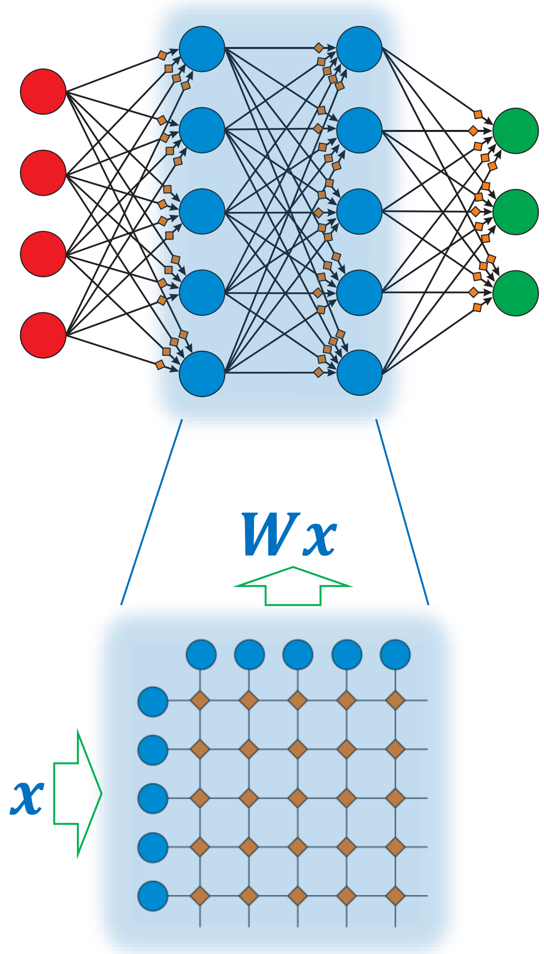
- **Processing dominated by many large matrix operations**
  - Forward Propagation: $W_{1,2..}$
  - Backward Propagation: $W^T_{1,2..}$
  - Weight Update: $\Delta W_{1,2..}$

Scale $\propto N^2$

Neurons/layer

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





For fast and efficient neural network data processing:

- Fully parallel processing
- Tight integration of processing and memory
- Analog signal processing

Crossbar arrays

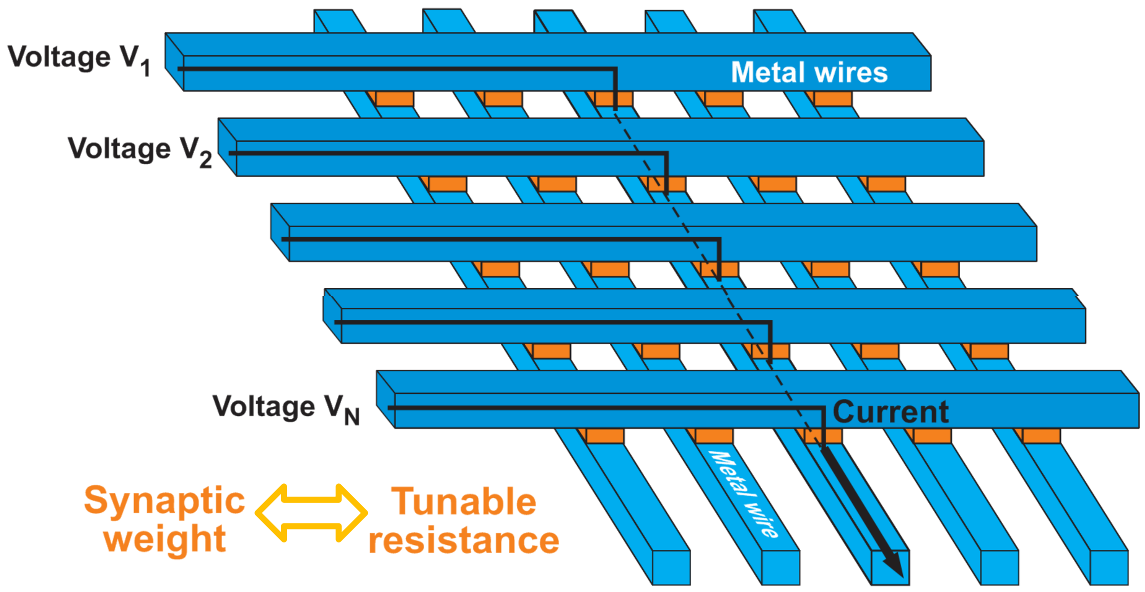
- Electrical
- Optical

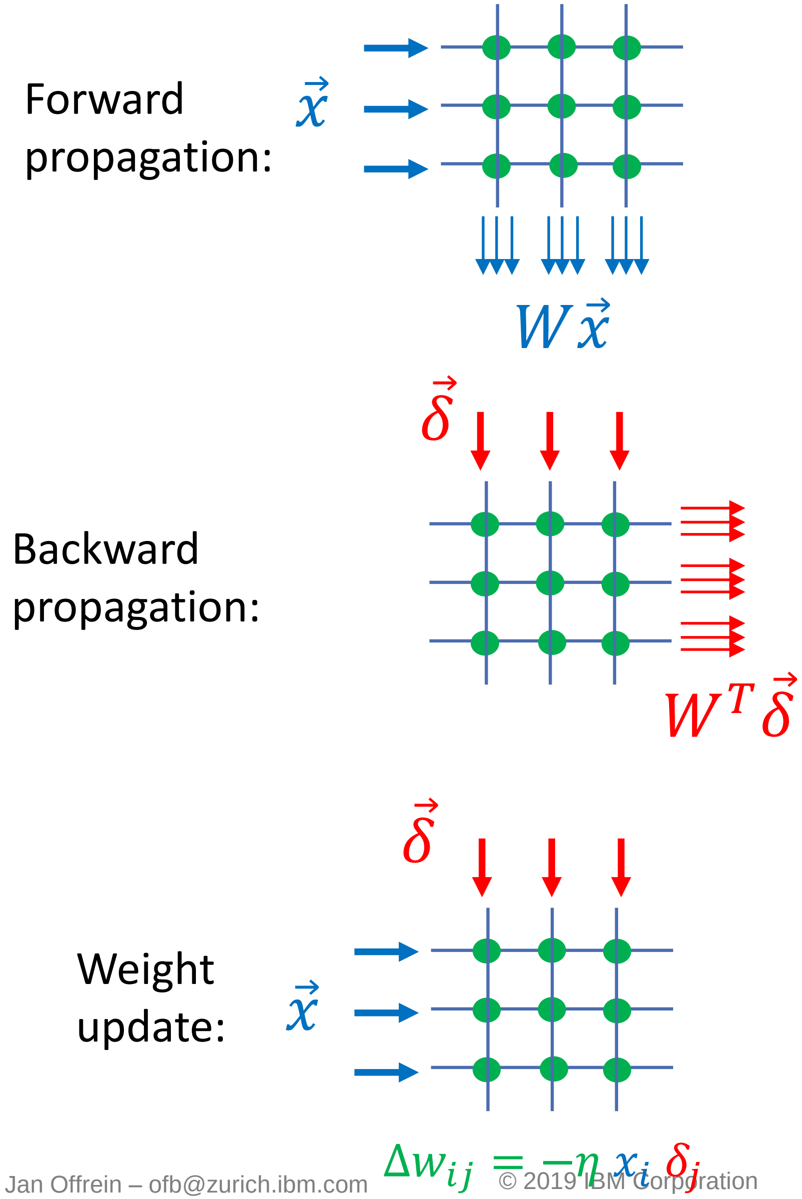
📖 G. W. Burr et al., "Tech. Dig. - Int. Electron Devices Meet. IEDM, vol. 2016–Febru, no. 408, p. 4.4.1-4.4.4, 2016.
📖 T. Gokmen and Y. Vlasov, Front. Neurosci., vol. 10, no. JUL, pp. 1–13, 2016.

Bert Jan Offrein – ofb@zurich.ibm.com

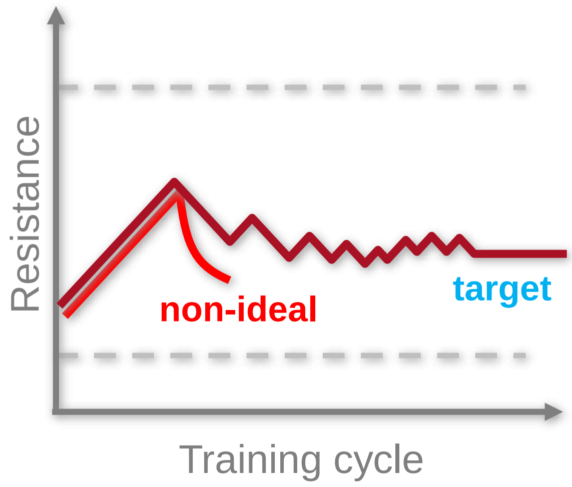
Analog crossbar arrays:

Electrical crossbar array:





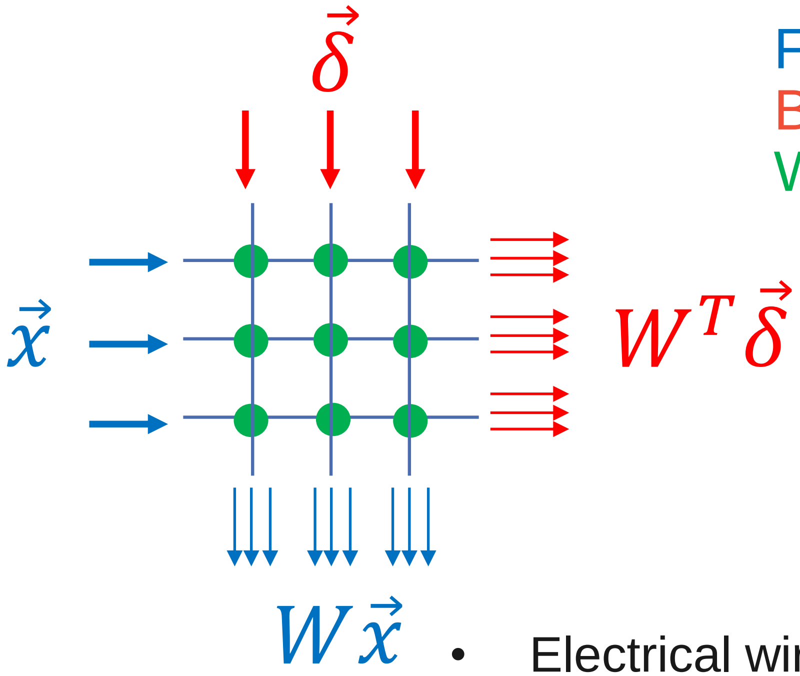
- **Weight update:** proportional to signals on row and column
  - Symmetric increase and decrease of weight
  - ~1000 analog levels required
- **Physical challenge**: Identify material systems that meet these requirements



Bert Jan Offrein – ofb@zurich.ibm.com

Photonic crossbar unit - operating principle

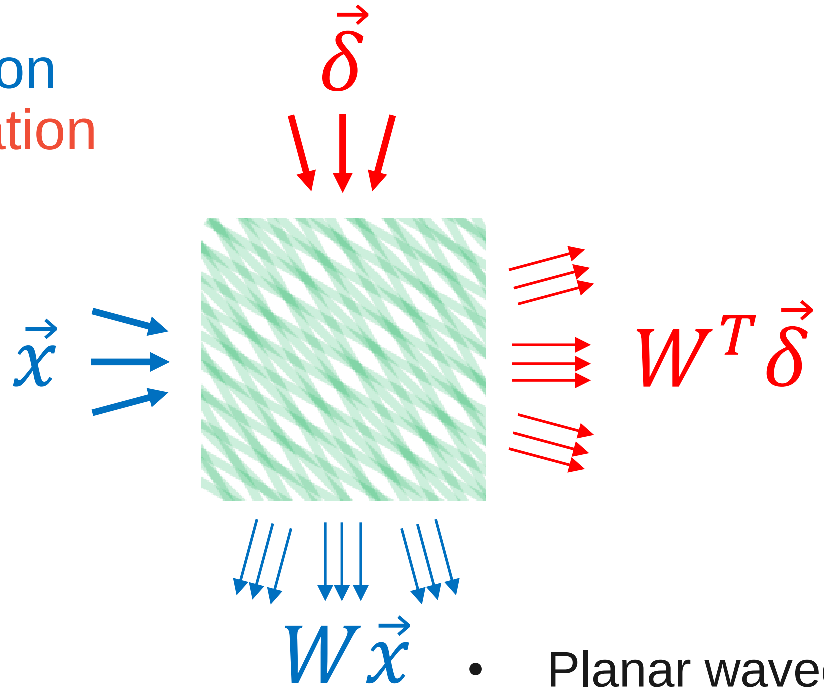
Electrical crossbar



Forward propagation
Backward propagation
Weights

- Electrical wires
- Local weights
- Resistance tuning

Photonic crossbar



- Planar waveguiding
- Distributed weights
- Refractive index tuning

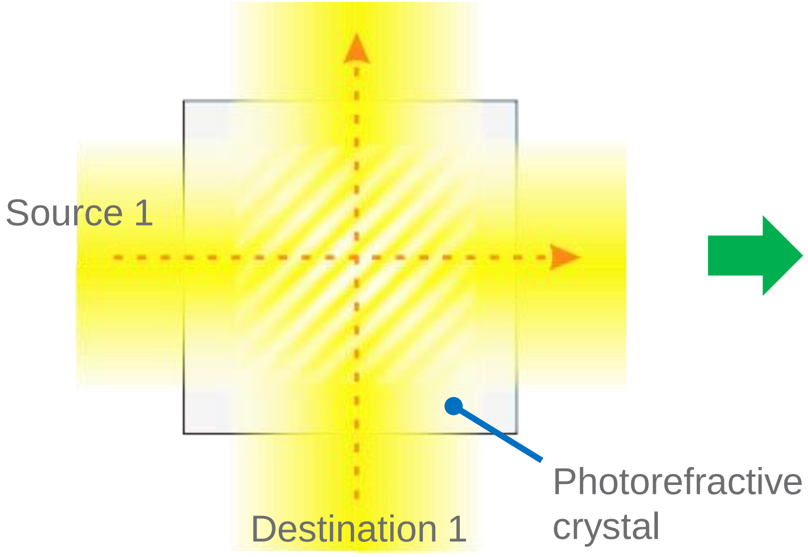
Writable photorefractive gratings provide the same functionality as the tunable resistive elements in a crossbar unit

Bert Jan Offrein – ofb@zurich.ibm.com

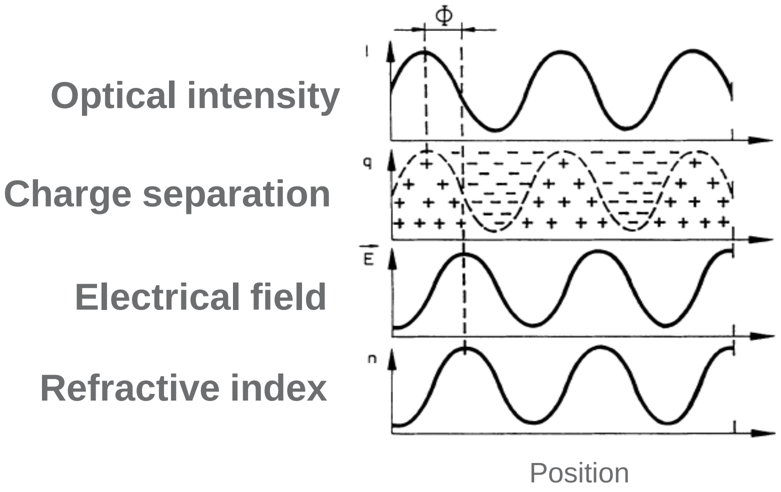
Optical crossbar arrays: Holographic storage and signal processing

Weight Storage:

Interference pattern:

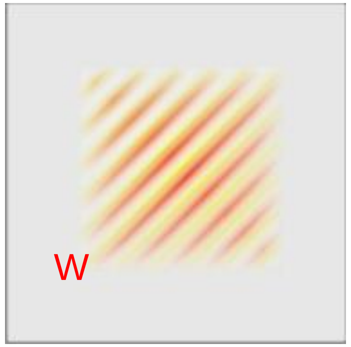


Photorefractive effect:



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

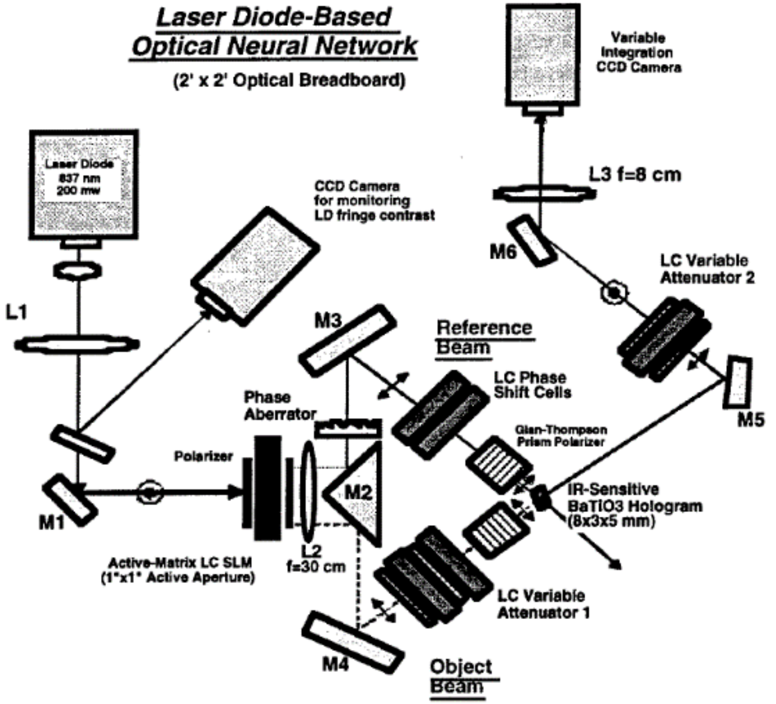
Bert Jan Offrein – ofb@zurich.ibm.com

IBM

Optical crossbar arrays: Integrated Solution

Concept demonstrated in bulk optics

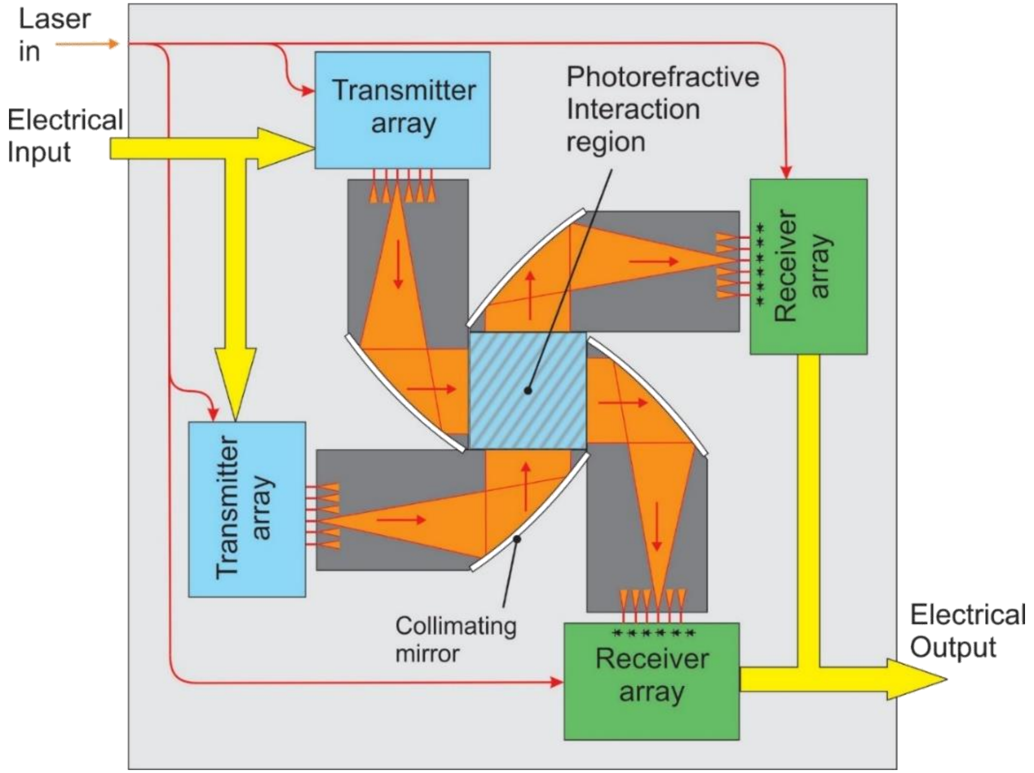
- Backpropagation training of neural networks with hidden layers
- Large setup, slow electro-optics, stability issues



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

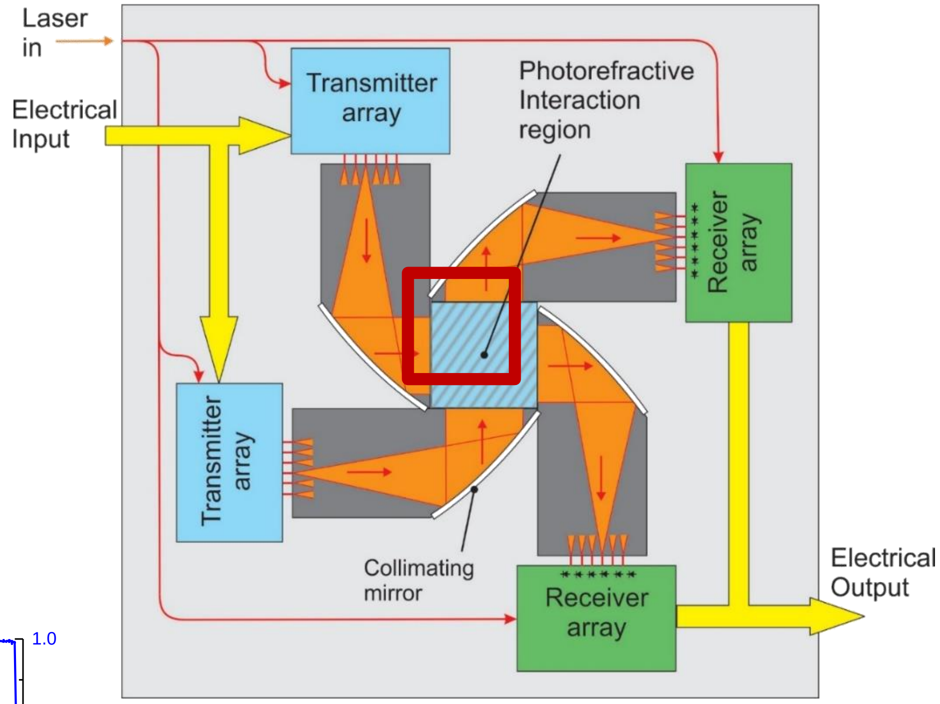
Our approach: Miniaturize using Integrated Optics

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



Bert Jan Offrein – ofb@zurich.ibm.com

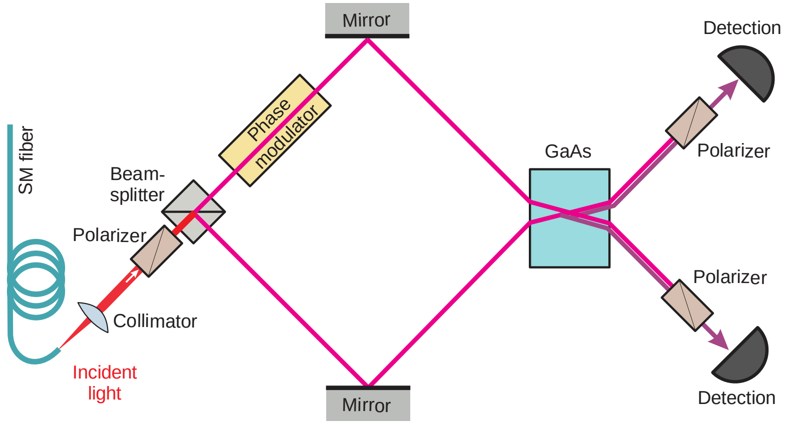
Photonic weight processing unit: Building blocks

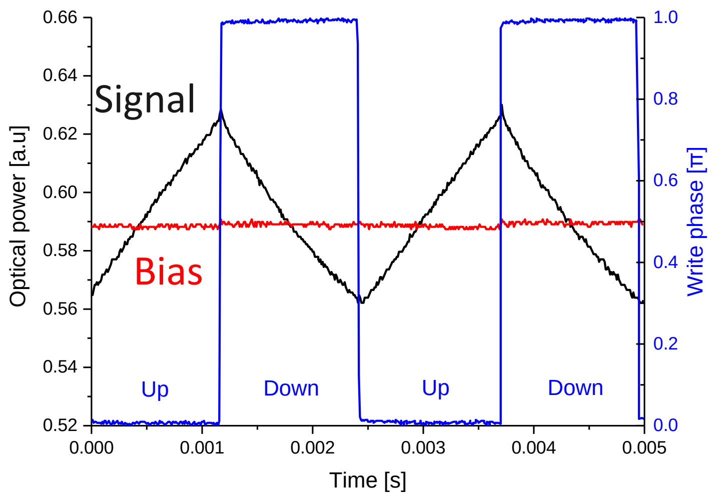


Photorefractive interaction region:

- Stores synaptic weights as refractive index gratings
- Photorefractive material: Semi-Insulating GaAs
  - Matches Si-Photonics wavelength range
  - Compatible with III-V on Silicon processes

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





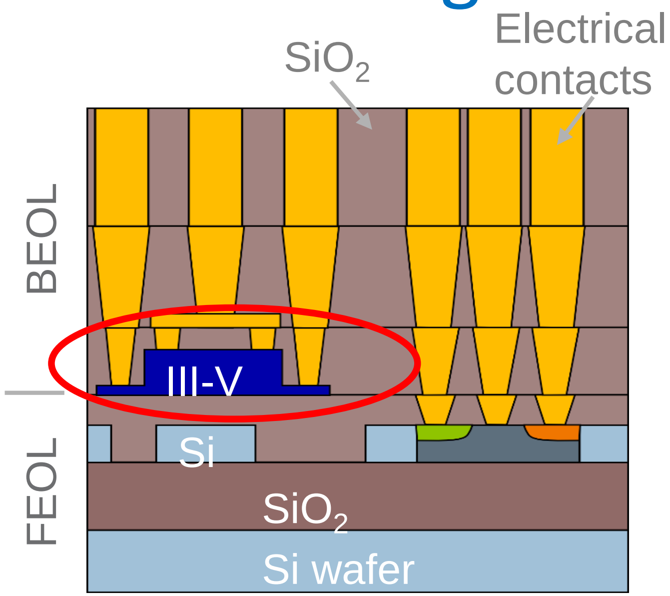
- Low asymmetry
- High dynamic range
- Bipolar weight storage
- To be confirmed in thin film

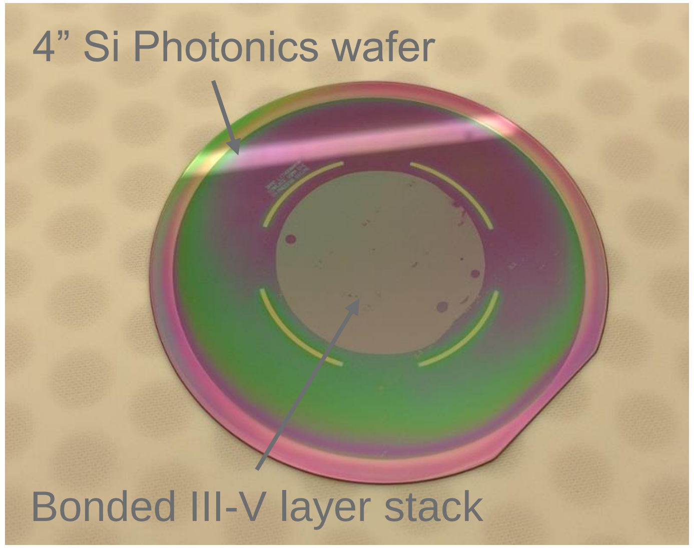
Bert Jan Offrein – ofb@zurich.ibm.com

Photonic weight processing unit: Building blocks

Integration of the photorefractive layer:

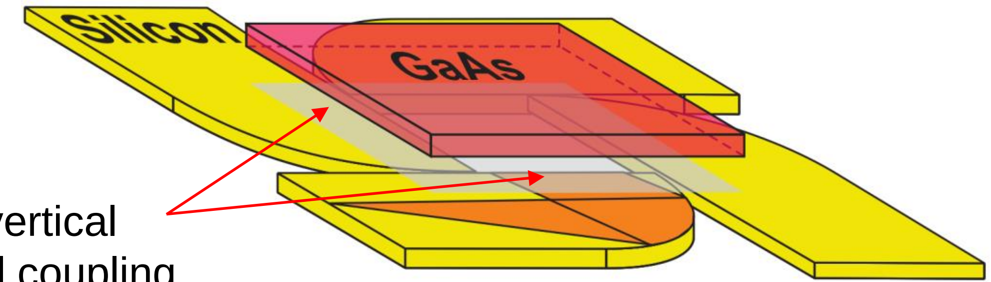
- Bonding technology as demonstrated for other III-V on Si projects:
  - Gain layers for integrated light sources
- Oxide bonding to Silicon-photonics stack



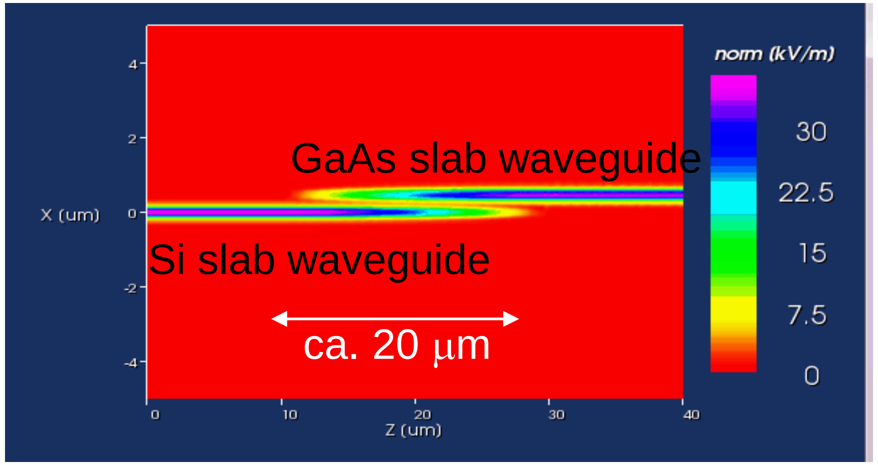


M. Seifried *et al.*, "Monolithically Integrated CMOS-Compatible III–V on Silicon Lasers“ doi: 10.1109/JSTQE.2018.2832654.

- Vertical directional coupling for efficient coupling of light between Si-photonic and GaAs layers



Vertical directional coupling



Bert Jan Offrein – ofb@zurich.ibm.com

Summary

- Silicon technology remains the basis for computing devices
  - Leverage existing processes, infrastructure and know-how
  - Continuous extending of materials and function

- Silicon Photonics is currently becoming a major driver for employing optical interconnects
  - Intra- and inter-datacenter communication
  - Integration is functions is a must for cost and assembly reasons

- New computing paradigms – Neuromorphic computing - provides a path to handle unstructured data
  - Analog signal processing in crossbar arrays
  - Parallel processing of key algorithms in neural networks
  - Electrical and optical implementations

Bert Jan Offrein – ofb@zurich.ibm.com

Acknowledgments

IBM Research – Zurich, Switzerland
R. Dangel, F. Horst, J. Fompeyrine, S. Abel, F.Eltes, N. Meier, D. Caimi, K. Moselund, L. Czornomaz, M. Sousa, H. Siegwart, C. Caer, Y. Baumgartner, D. Jubin, A. La Porta, J. Weiss, U. Drechsler

Co-funded by the European Union Horizon 2020 Programme and the Swiss National Secretariat for Education, Research and Innovation (SERI)

NEURAM, ULPEC, PLACMOS
NEUROTECH, ChipAI

Thank you
for your attention

Bert Jan Offrein – ofb@zurich.ibm.com