BEng 207 Special Topics in Bioengineering

Neuromorphic Integrated Bioelectronics

Week 7: Learning and Adaptation

Gert Cauwenberghs

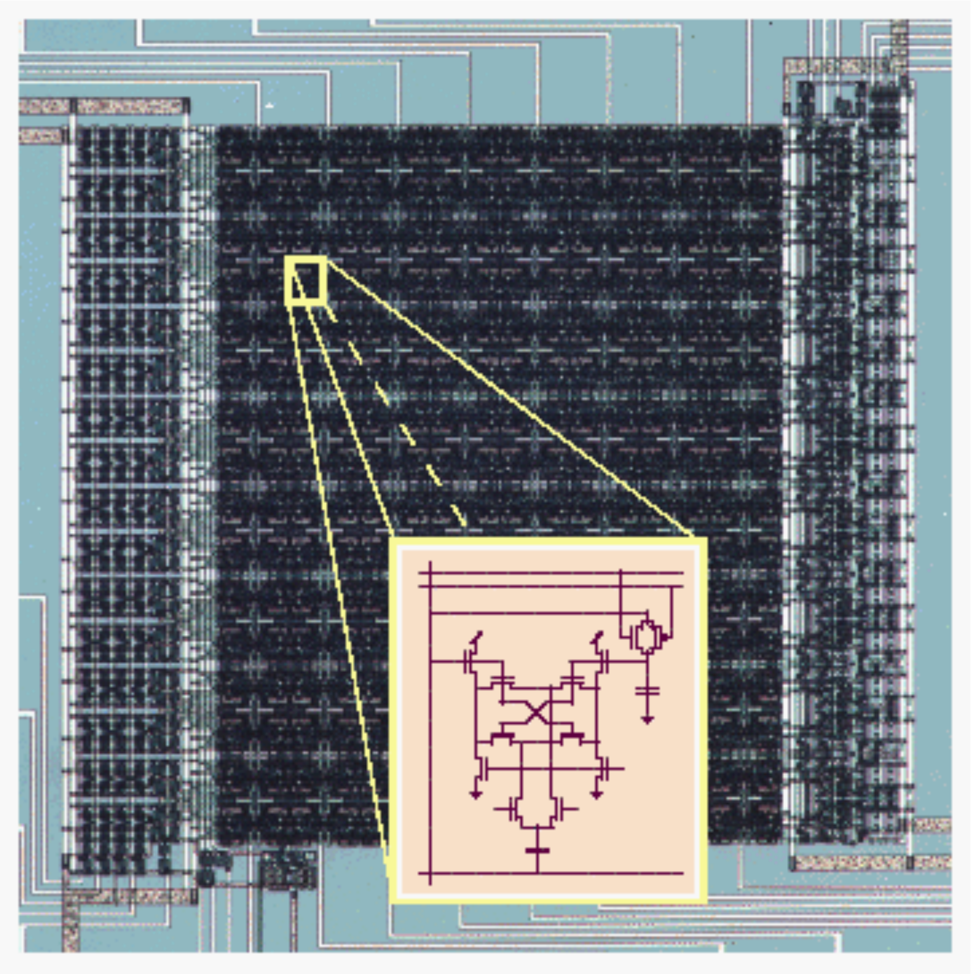
Department of Bioengineering
UC San Diego

http://isn.ucsd.edu/courses/beng207

BENG 207 Neuromorphic Integrated Bioelectronics

Date	Topic
9/27, 9/29	Biophysical foundations of natural intelligence in neural systems. Subthreshold MOS silicon models of membrane excitability. Silicon neurons. Hodgkin-Huxley and integrate-and-fire models of spiking neuronal dynamics. Action potentials as address events.
10/4, 10/6	Silicon retina. Low-noise, high-dynamic range photoreceptors. Focal-plane array signal processing. Spatial and temporal contrast sensitivity and adaptation. Dynamic vision sensors.
10/11, 10/13	Silicon cochlea. Low-noise acoustic sensing and automatic gain control. Continuous wavelet filter banks. Interaural time difference and level difference auditory localization. Blind source separation and independent component analysis.
10/18, 10/20	Silicon cortex. Neural and synaptic compute-in-memory arrays. Address-event decoders and arbiters, and integrate-and-fire array transceivers. Hierarchical address-event routing for locally dense, globally sparse long-range connectivity across vast spatial scales.
10/28, 11/1	Review. Modular and scalable design for neuromorphic and bioelectronic integrated circuits and systems. Design for full testability and controllability.
11/1, 11/3	Midterm due 11/2. Low-noise, low-power design. Fundamental limits of noise-energy efficiency, and metrics of performance. Biopotential and electrochemical recording and stimulation, lab-on-a-chip electrophysiology, and neural interface systems-on-chip.
11/8, 11/10	Learning and adaptation to compensate for external and internal variability over extended time scales. Background blind calibration of device mismatch. Correlated double sampling and chopping for offset drift and low-frequency noise cancellation.
11/15, 11/17	Energy conservation. Resonant inductive power delivery and data telemetry. Ultra-high efficiency neuromorphic computing. Resonant adiabatic energy-recovery charge-conserving synapse arrays.
11/22, 11/24	Guest lectures
11/29, 12/1	Project final presentations. All are welcome!

Large-Scale Mixed-Signal Sensory Computation



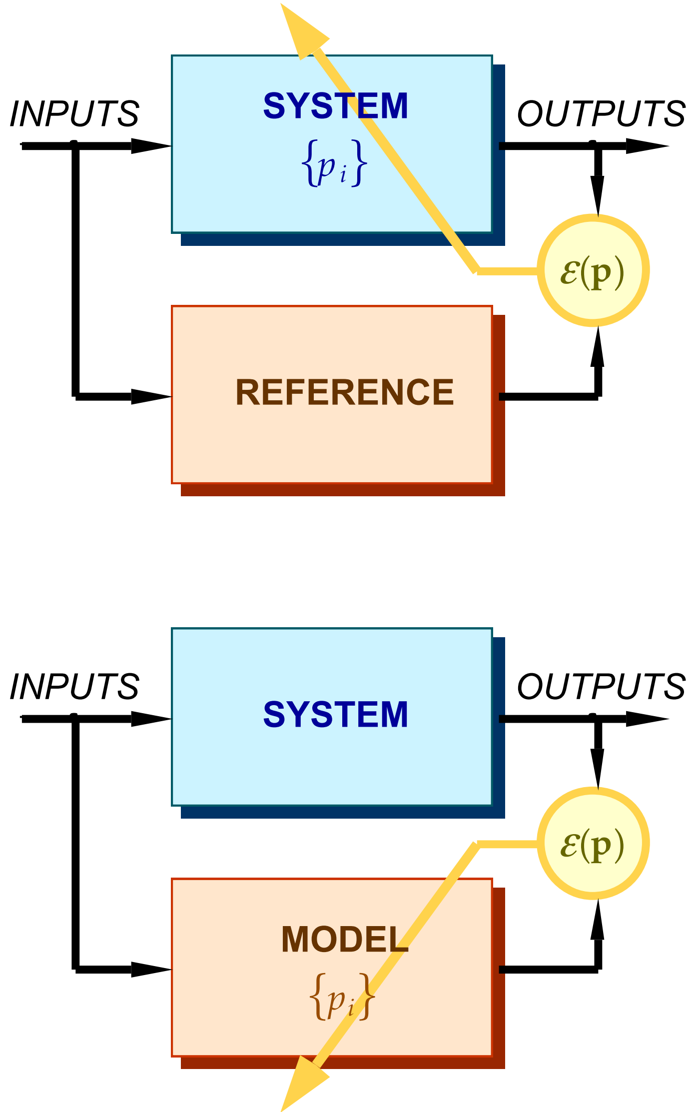
Example: VLSI Analog-to-digital vector quantizer (Cauwenberghs and Pedroni, 1997)

- **Massive Parallelism**
  - distributed representation
  - local memory and adaptation
  - analog sensory interface
  - physical computation
  - analog accumulation on single wire
- **Scalable**

  silicon area and power scale linearly with throughput
- **Highly Efficient**

  factor 100 to 10,000 less energy/operation than DSP
- **Limited Precision**
  - analog mismatch and nonlineary (WYDINWYG)
  - fix: adaptation in redundancy

Learning on Silicon



Adaptation:

- necessary for robust performance under variable conditions and in unpredictable environments
- also compensates for imprecision in analog computation
- avoids ad-hoc programming, tuning, and manual parameter adjustment

Learning:

- generalization of output to previously unknown, although similar, stimuli
- system identification to extract relevant environmental parameters

Cauwenberghs & Bayoumi, Eds., *Learning on Silicon*, Kluwer 1999.

Adaptive Elements

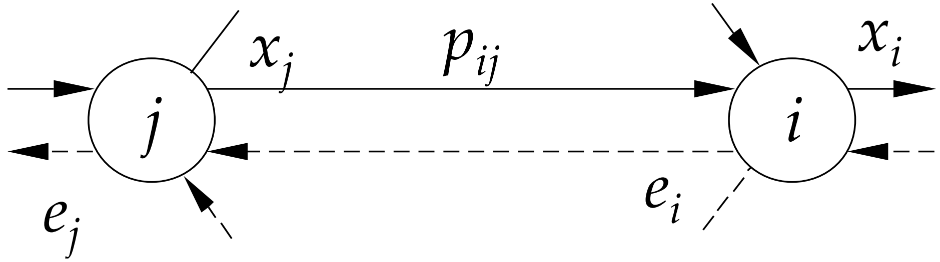
Adaptation:

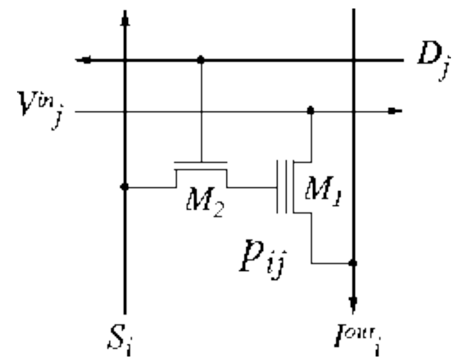
Autozeroing (high-pass filtering)	*outputs*
Offset Correction	*outputs*
e.g. Image Non-Uniformity Correction	
Equalization /Deconvolution	*inputs, outputs*
e.g. Source Separation; Adaptive Beamforming	

Learning:

Unsupervised Learning	*inputs, outputs*
e.g. Adaptive Resonance; LVQ; Kohonen	
Supervised Learning	*inputs, outputs, targets*
e.g. Least Mean Squares; Backprop	
Reinforcement Learning	*reward/punishment*

Incremental Outer-Product Learning in Neural Nets





Multi-Layer Perceptron: $x_i = f(\sum_j p_{ij} x_j)$

Outer-Product Learning Update: $\Delta p_{ij} = \eta \; x_j \cdot e_i$

- Hebbian *(Hebb, 1949)*: $e_i = x_i$
- LMS Rule *(Widrow-Hoff, 1960)*: $e_i = f'_i \cdot \left(x_i^{\text{target}} - x_i\right)$
- Backpropagation *(Werbos, Rumelhart, LeCun)*: $e_j = f'_j \cdot \sum_i p_{ij} e_i$

Technology

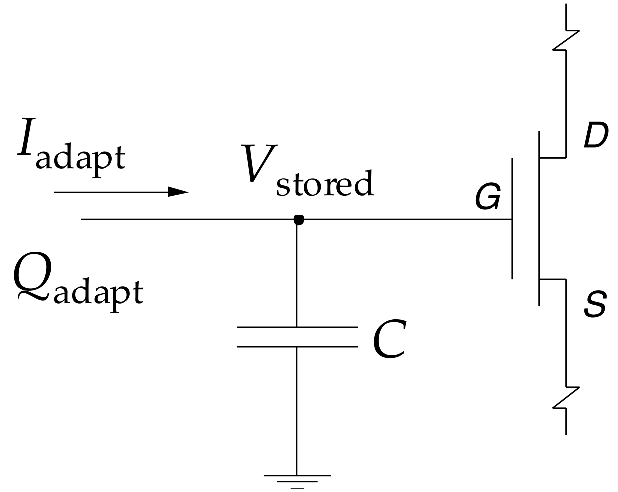
Incremental Adaptation:

- Continuous-Time:

$$C \frac{\mathrm{d}}{\mathrm{d}t} V_{\mathrm{stored}} = I_{\mathrm{adapt}}$$

- Discrete-Time:

$$C \, \Delta V_{\mathrm{stored}} = Q_{\mathrm{adapt}}$$



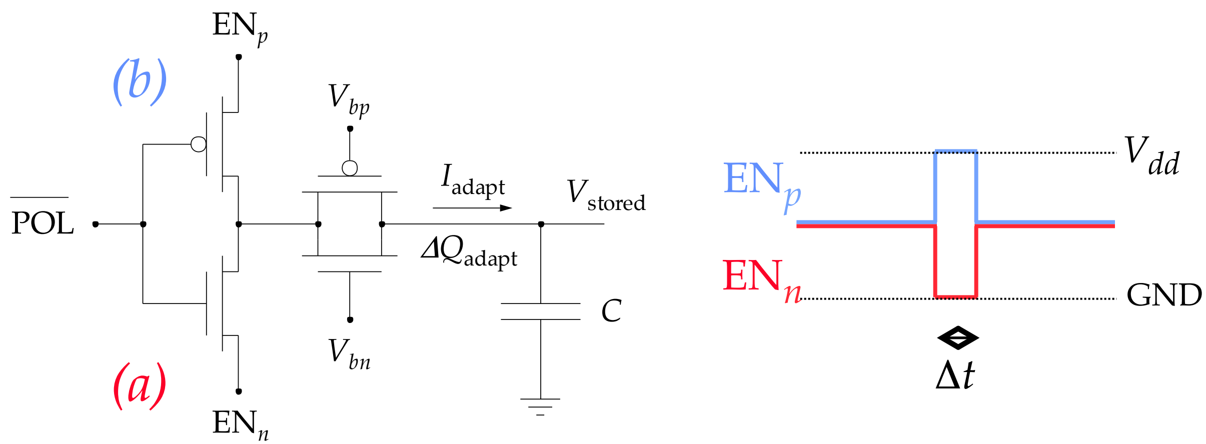
Storage:

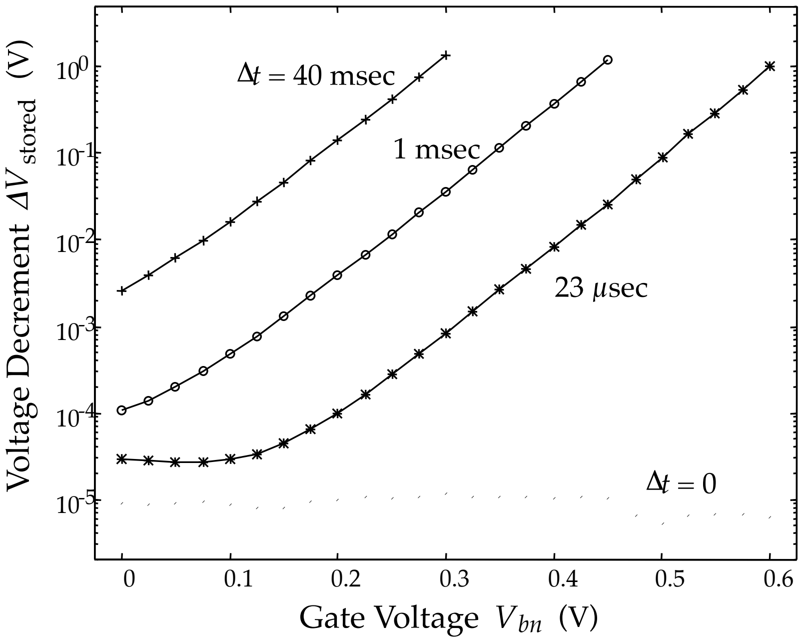
- Volatile capacitive storage (incremental refresh)
- Non-volatile storage (floating gate)

Precision:

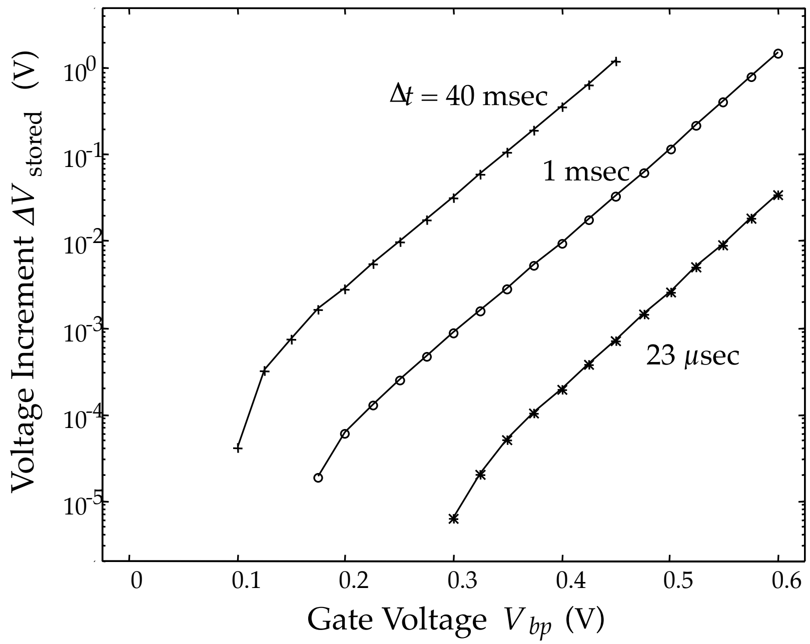
- Only polarity of the increments is critical (not amplitude).
- Adaptation compensates for inaccuracies in the analog implementation of the system.

Dynamic Memory and Incremental Adaptation





(a)

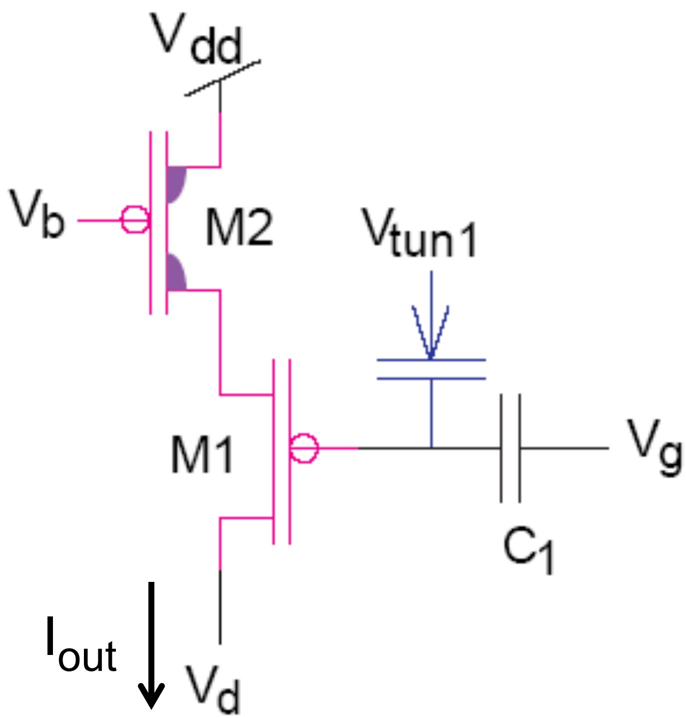


(b)

Cauwenberghs, ALOG 1998

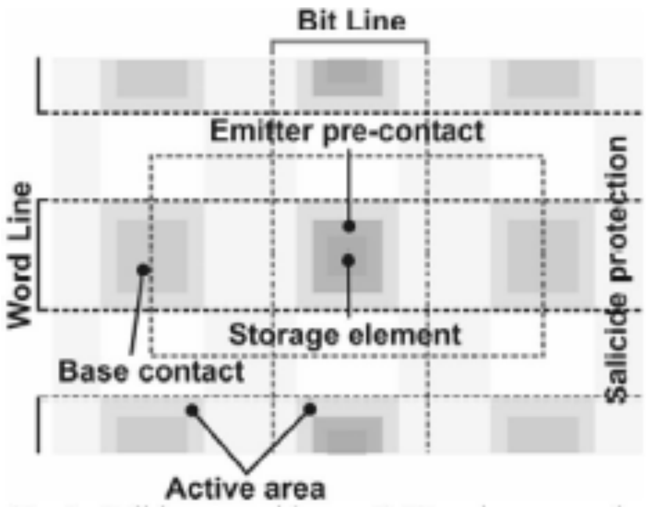
Floating-Gate Non-Volatile Memory and Adaptation

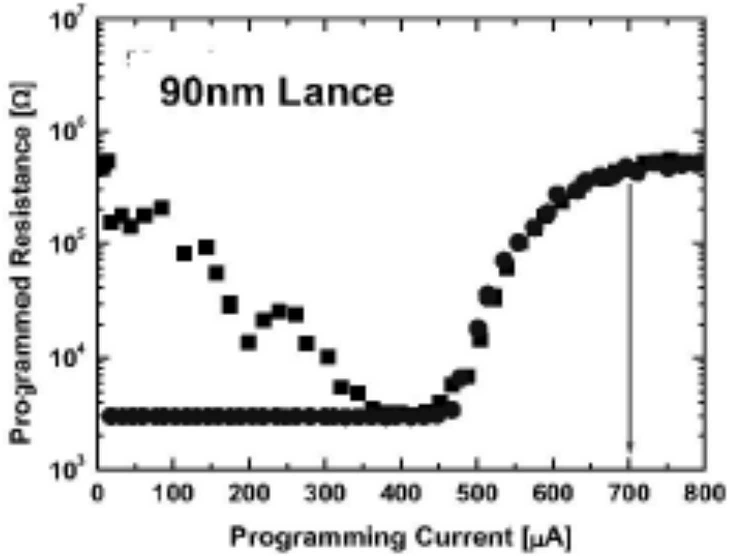
Paul Hasler, Chris Diorio, Brad Minch, Carver Mead, …

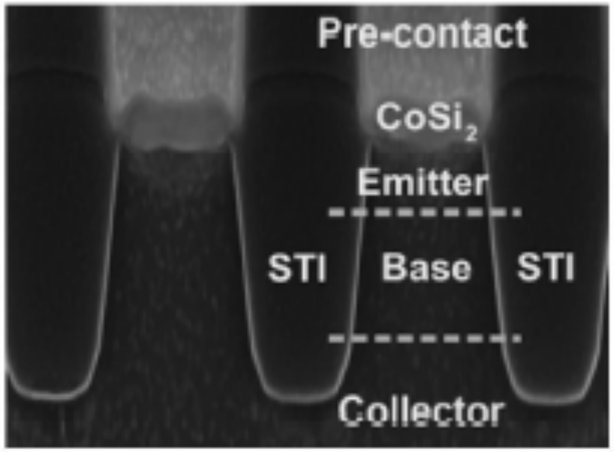


- **Hot electron injection**
  - 'Hot' electrons injected from drain onto floating gate of M1.
  - Injection current is proportional to drain current and exponential in floating-gate to drain voltage (~5V).
- **Tunneling**
  - Electrons tunnel through thin gate oxide from floating gate onto high-voltage (~30V) n-well.
  - Tunneling voltage decreases with decreasing gate oxide thickness.
- **Source degeneration**
  - Short-channel M2 improves stability of closed-loop adaptation (Vd open-circuit).
  - M2 is not required if adaptation is regulated (Vd driven).
- **Current scaling**
  - In subthreshold, Iout is exponential both in the floating gate charge, and in control voltage Vg.

Phase Change Memory Technology





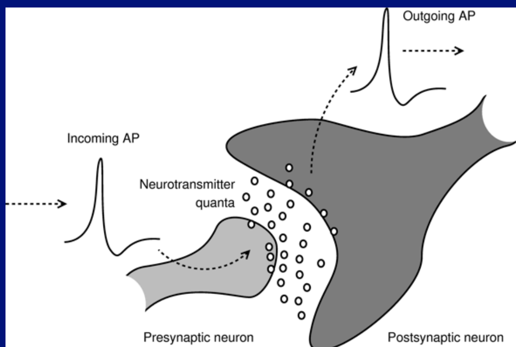


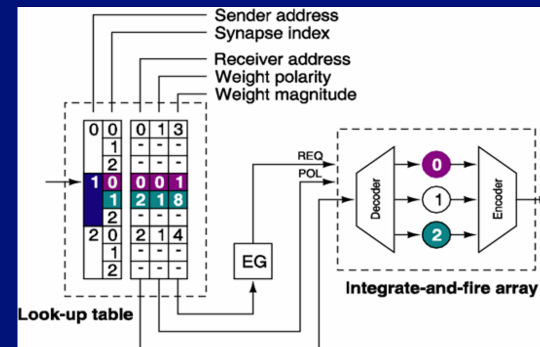
- Analog switch
  - *100Ω – 1MΩ resistance range*
- Fast write and read times (~nsec)
- Radiation hard

G. Atwood, R. Bez, "90nm Phase Change Technology with μTrench and Lance Cell Elements," VLSI Symp, 2007.

Reconfigurable Synaptic Connectivity and Plasticity

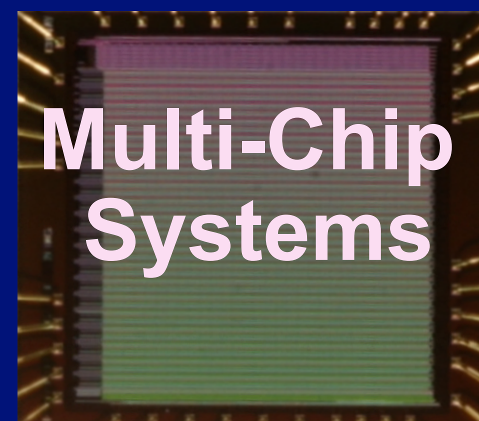
From Microchips to Large-Scale Neural Systems





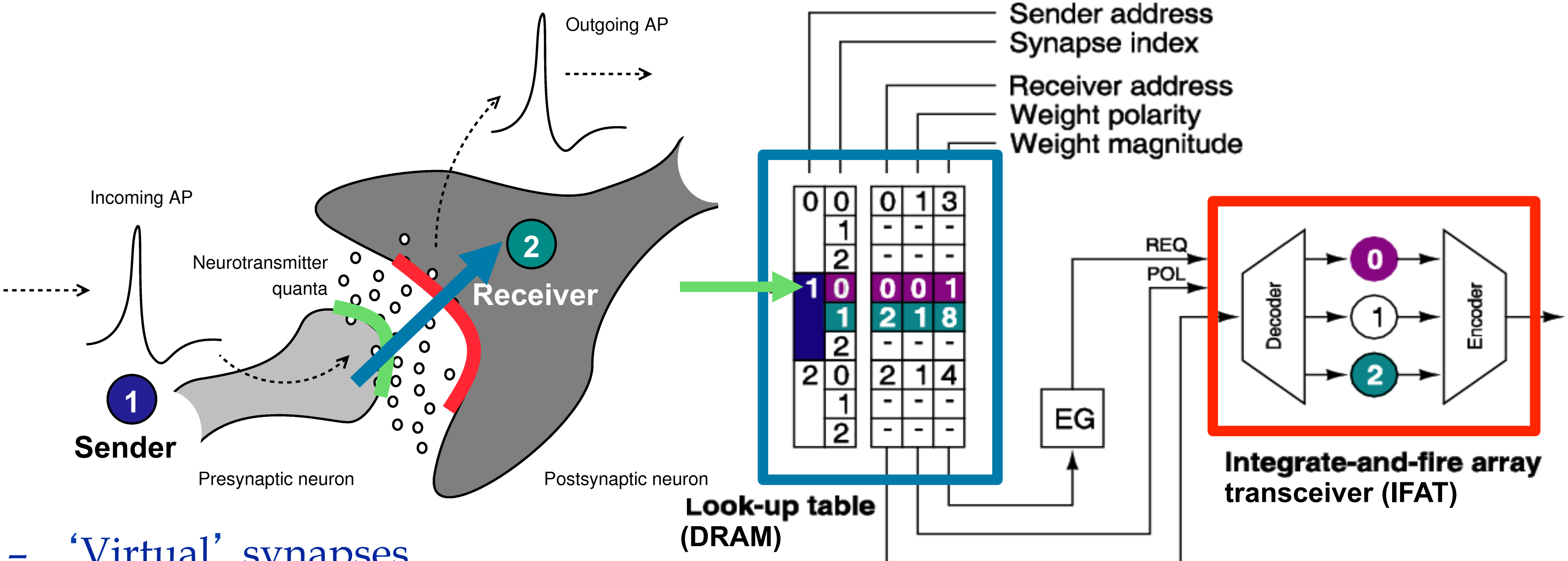
Address-Event Representation

Synaptic Plasticity & Wiring



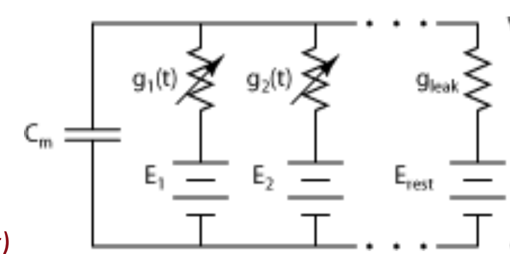
Address-Event Synaptic Connectivity and Plasticity

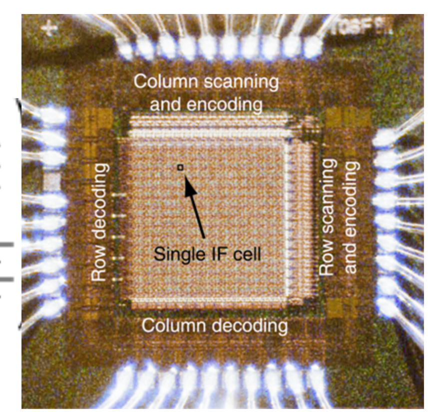
Goldberg, Cauwenberghs and Andreou, 2000



0	0	0	1	3
	1	-	-	-
	2	-	-	-
1	0	0	0	1
	1	2	1	8
	2	-	-	-
2	0	2	1	4
	1	-	-	-
	2	-	-	-

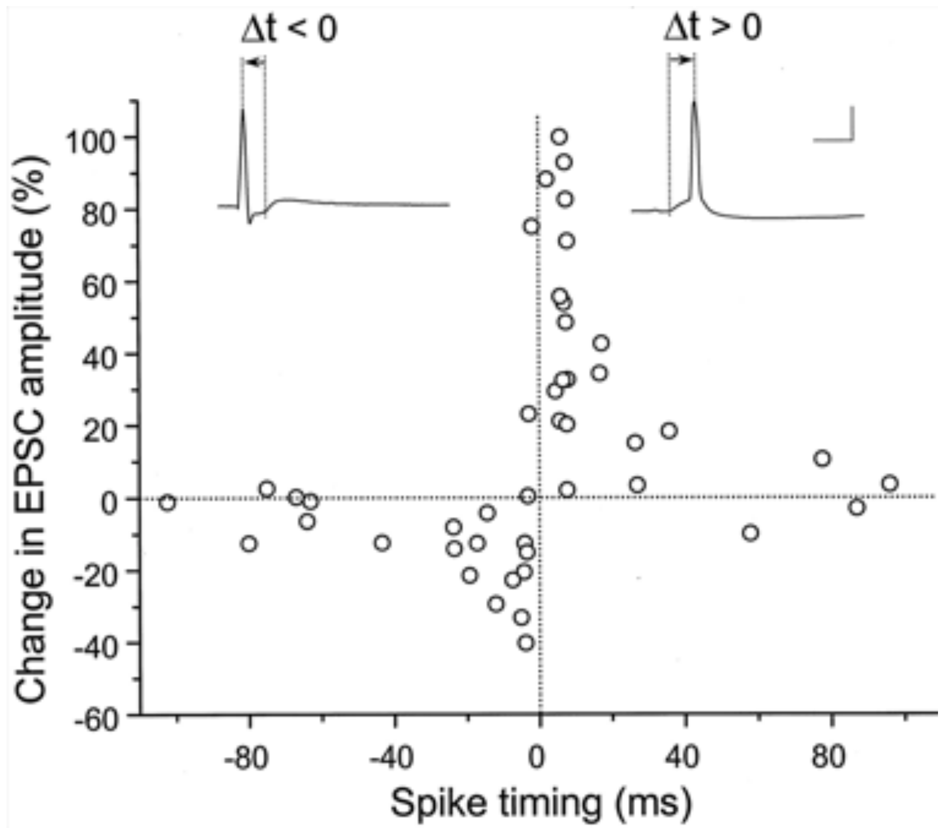
- ‘Virtual’ synapses
  - *Dynamically reconfigurable*
  - *Wide-ranging connectivity*
  - *Rewiring and synaptic plasticity*
- Quantal release: $R = n\,p\,q$
  - *n: multiplicity* *(repeat event)*
  - *p: probability of release* *(toss a coin)*
  - *q: quantity released* *(set amplitude)*





IFAT2 (2000)

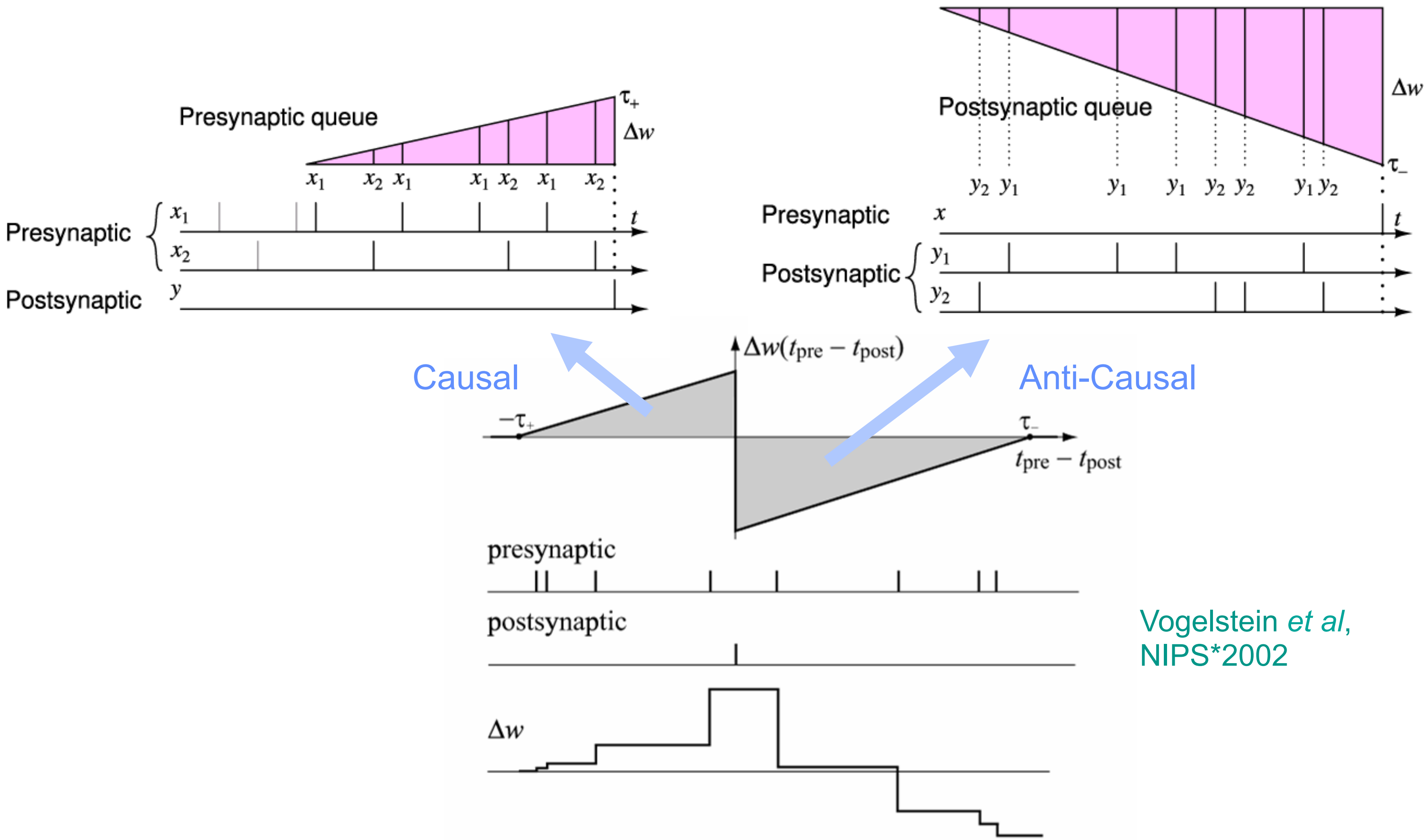
Spike Timing-Dependent Plasticity



Bi and Poo, 1998

Spike Timing-Dependent Plasticity

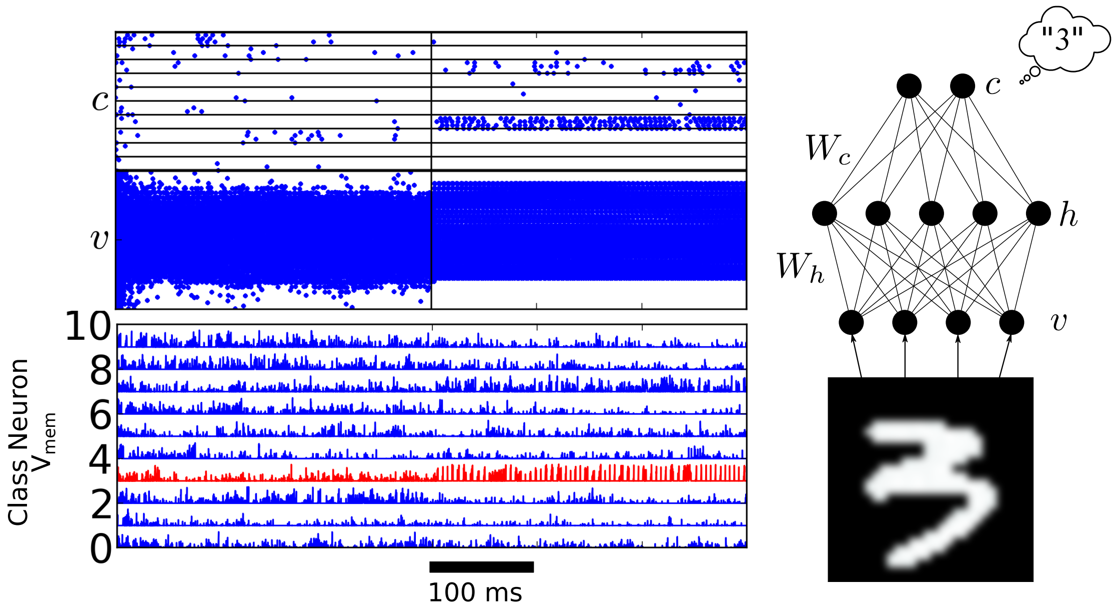
in the Address Domain



Vogelstein *et al*, NIPS*2002

Deep Learning in Spike-Based Neuromorphic Systems

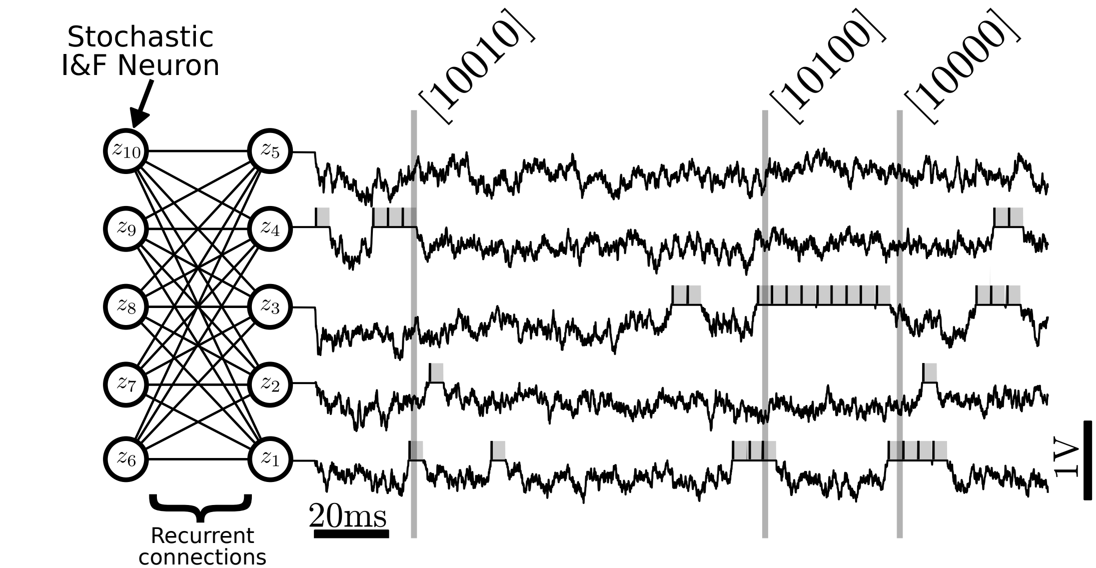
- *Neural Sampling:* Integrate & Fire (I&F) neurons can perform MCMC sampling of a Boltzmann distribution
- Restricted Boltzmann Machines can be trained using STDP

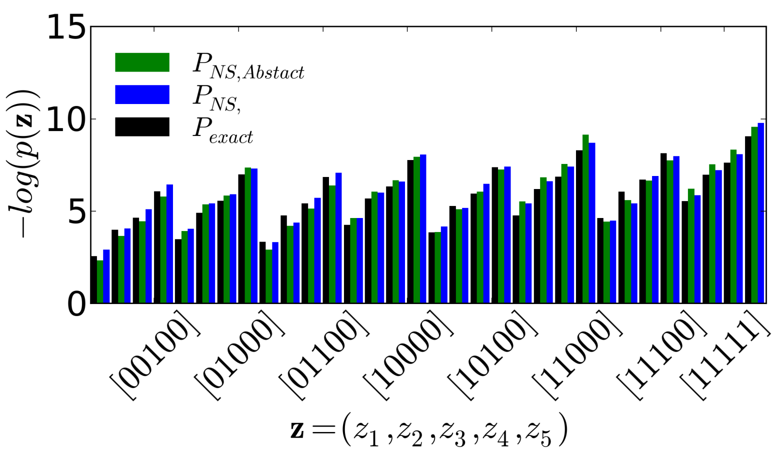


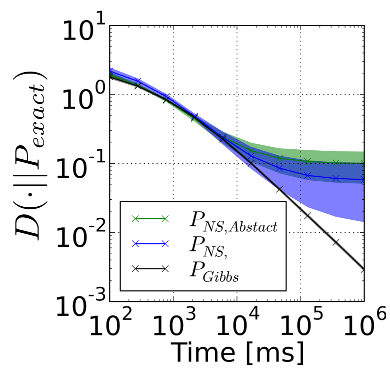
- 92% accuracy on MNIST hand-written digit recognition task

E. Neftci *et al*, "Event-driven contrastive divergence for spiking neuromorphic systems", *Frontiers in Neuroscience*, doi: 10.3389/fnins.2013.00272, 2014

Neural Sampling with Noisy Integrate-and-Fire Neurons





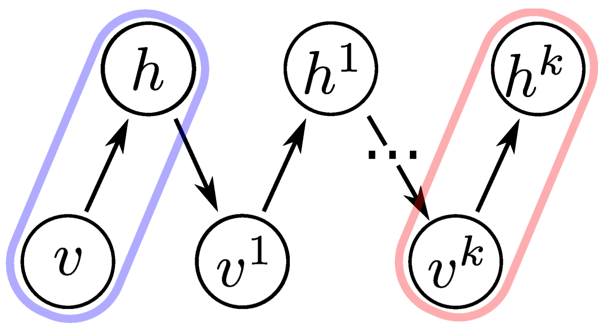


- *We identified conditions under which spike trains from general integrate-and-fire neurons in the presence of noise generate Monte-Carlo Markov Chain (MCMC) samples from a Boltzmann distribution*
- *This framework provides the foundation for event-driven on-line stochastic learning using contrastive divergence in Boltzmann machines*

E. Neftci *et al*, "Event-driven contrastive divergence for spiking neuromorphic systems", *Frontiers in Neuroscience*, doi: 10.3389/fnins.2013.00272, 2014

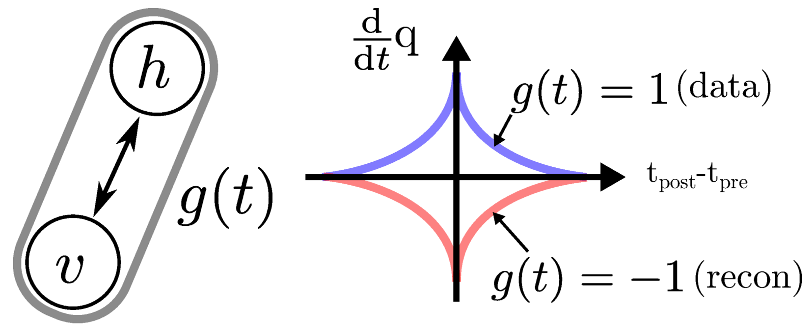
Event-Driven Contrastive Divergence

On-line Training of Boltzmann Machines Using STDP



$$\Delta w \propto \langle vh \rangle_{\mathrm{data}} - \langle v^k h^k \rangle_{\mathrm{recon}}$$

CD training with standard RBM



$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{q} = g(t)\mathrm{STDP}(v(t), h(t))$$

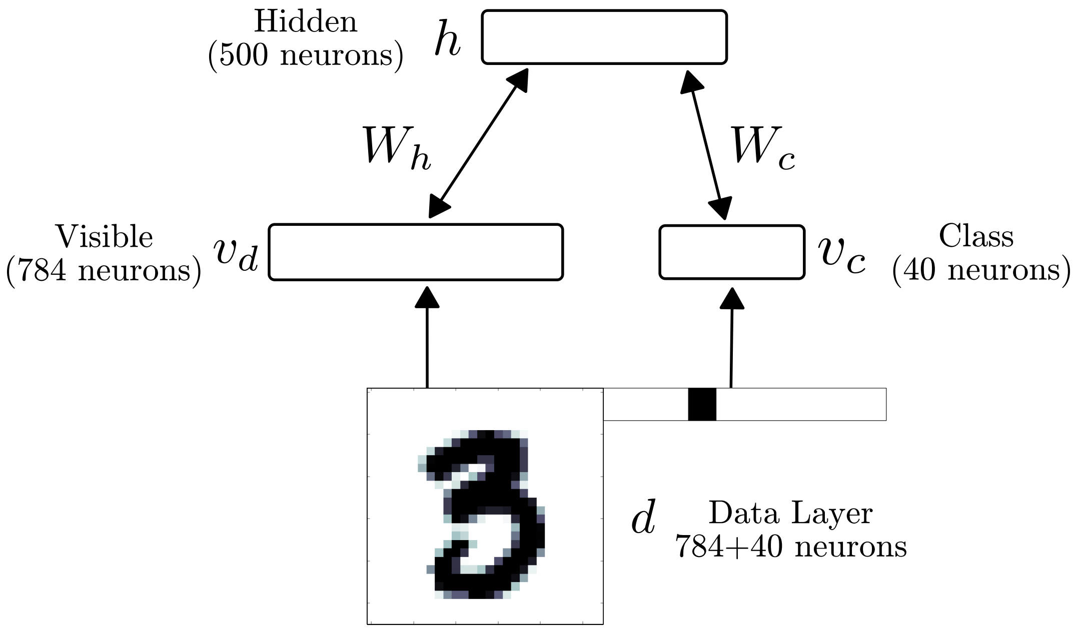
eCD on-line training with I&F RBM

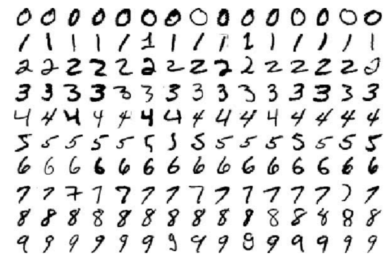
- Emulates contrastive divergence (CD) for training standard Restricted Boltzmann Machines (RBMs) using neural sampling with integrate-and-fire neurons.
- On-line spike event-driven training using spike-timing dependent plasticity (STDP)
  - *Temporally symmetric form produces the correlations* $\langle vh \rangle$ *in on-line form*
  - *Modulation* $g(t)$ *controls wake-sleep phases (data vs. reconstruction)*

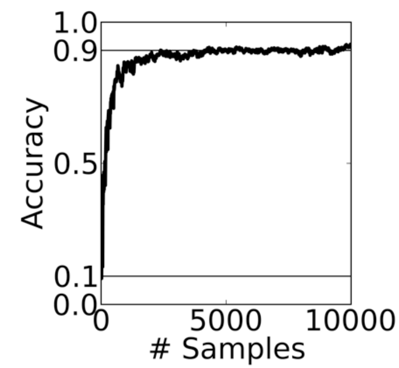
E. Neftci *et al*, "Event-driven contrastive divergence for spiking neuromorphic systems", *Frontiers in Neuroscience*, doi: 10.3389/fnins.2013.00272, 2014

Event-Driven Contrastive Divergence

Learning a Model of MNIST Hand-Written Digits





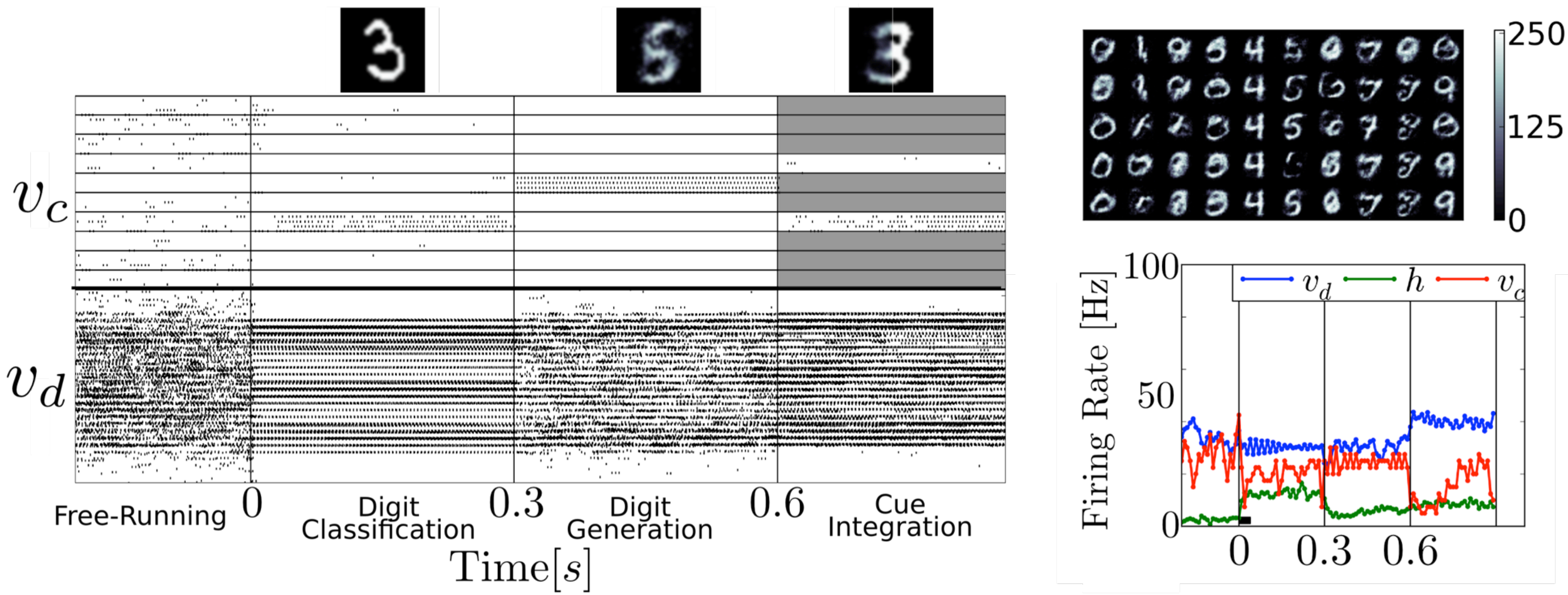


- MNIST hand-written digit recognition accuracy:
  - *CD with standard RBM: 93.6%*
  - *eCD with neural sampling: 91.9%*
- Extends to deep learning across multiple RBM layers for greater accuracy

E. Neftci *et al*, "Event-driven contrastive divergence for spiking neuromorphic systems", *Frontiers in Neuroscience*, doi: 10.3389/fnins.2013.00272, 2014

Event-Driven Contrastive Divergence

Inference, Generation, and Cue Integration



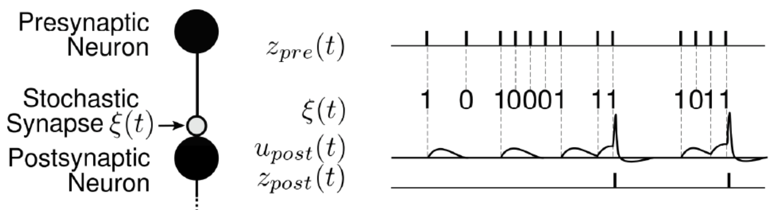
- Generative power of the Boltzmann machine model:
  - *Bottom-up: Classification of incoming data*
  - *Top-down: Generation of prototypical data for a class label*
  - *Hybrid: Cue integration with missing data based on class label priors*

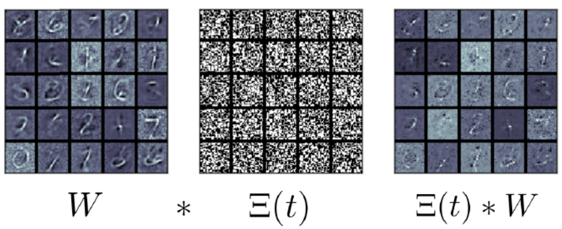
E. Neftci *et al*, "Event-driven contrastive divergence for spiking neuromorphic systems", *Frontiers in Neuroscience*, doi: 10.3389/fnins.2013.00272, 2014

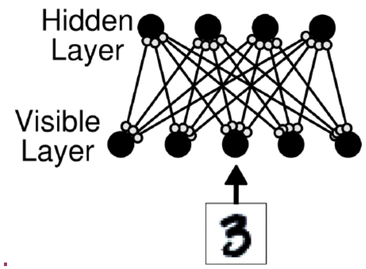
Spiking Synaptic Sampling Machine (S^3M)

Biophysical Synaptic Stochasticity in Inference and Learning

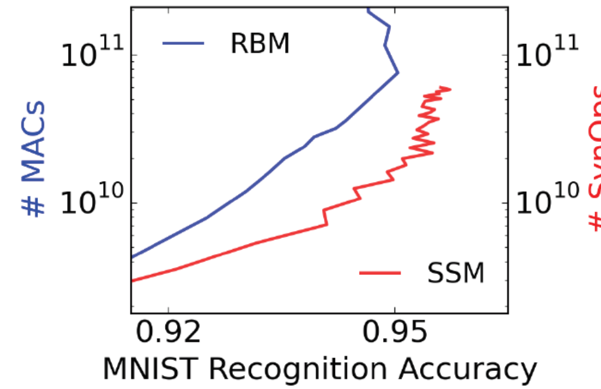
Synaptic stochasticity as biophysical model of continuous *DropConnect*







Time-varying Bernoulli random masking of weights



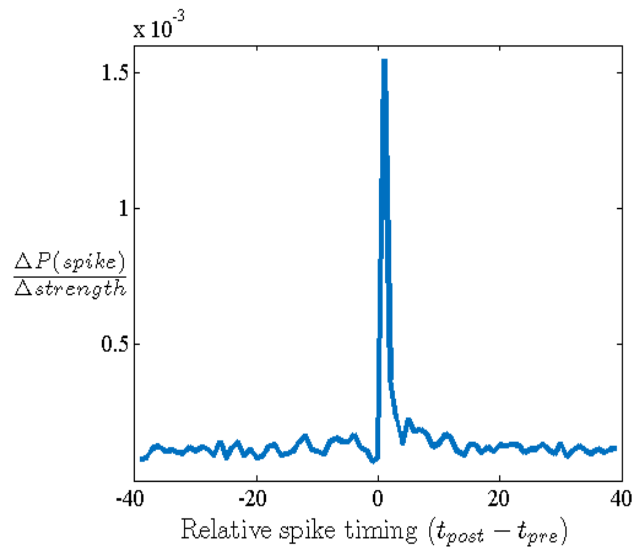
The S^3M requires fewer synaptic operations (SynOps) than the equivalent Restricted Boltzmann Machine (RBM) requires multiply-accumulate (MAC) operations at the same accuracy.

- Stochastic synapses for spike-based Monte Carlo sampling
  - *Models biophysical origins of noise in neural systems*
  - *Activity dependent noise: multiplicative synaptic sampling rather than additive neural sampling*
  - *Sparsity in neural activity and in synaptic connectivity*
- Online unsupervised learning with STDP
  - *Biophysical model of spike-based learning*
  - *Event-driven contrastive divergence*

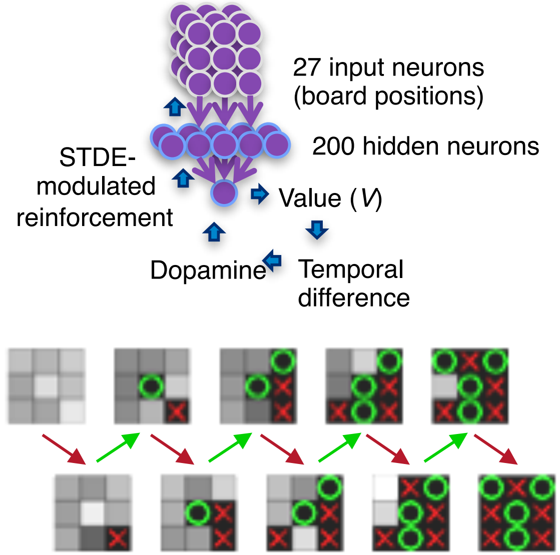
Emre O. Neftci, Bruno U. Pedroni, Siddharth Joshi, Maruan Al-Shedivat, Gert Cauwenberghs, "Stochastic Synapses Enable Efficient Brain-Inspired Learning Machines," *Frontiers in Neuroscience*, vol. 10, pp. 3389:1-16 (DOI: 10.3389/fnins.2016.00241), 2016.

Spike-Timing Dependent Eligibility

Reinforcement Learning by Reward Modulation of STDP



Spike timing-dependent eligibility (STDE)

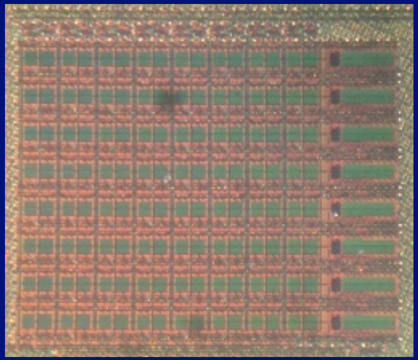


STDE-based temporal-difference reinforcement learning of the game of Tic-Tac-Toe

- Spike timing-dependent eligibility (STDE):
  - *Variant on biologically inspired spike timing-dependent plasticity (STDP)*
  - *Quantifies the sensitivity of post-synaptic spiking probability, conditioned on timed pre-synaptic spike input, to synaptic strength*
  - *Direct replacement for input activity term in Hebb-type incremental outerproduct update rules for gradient-based learning in rate-based ANNs*
- Temporal-difference reinforcement learning
  - *STDE-based Dopamine modulation of reward*

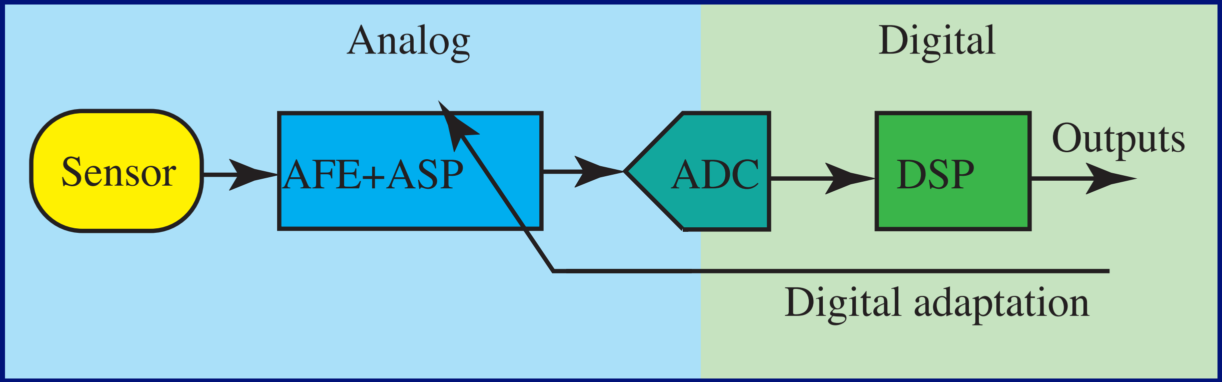
P. Frady et al, 2009

Adaptive Low-Power Sensory Systems



2pJ/MAC 14b 8×8 Linear Transform Mixed-Signal Spatial Filter in 65nm CMOS with 84dB Interference Suppression

S. Joshi et al, ISSCC 2017

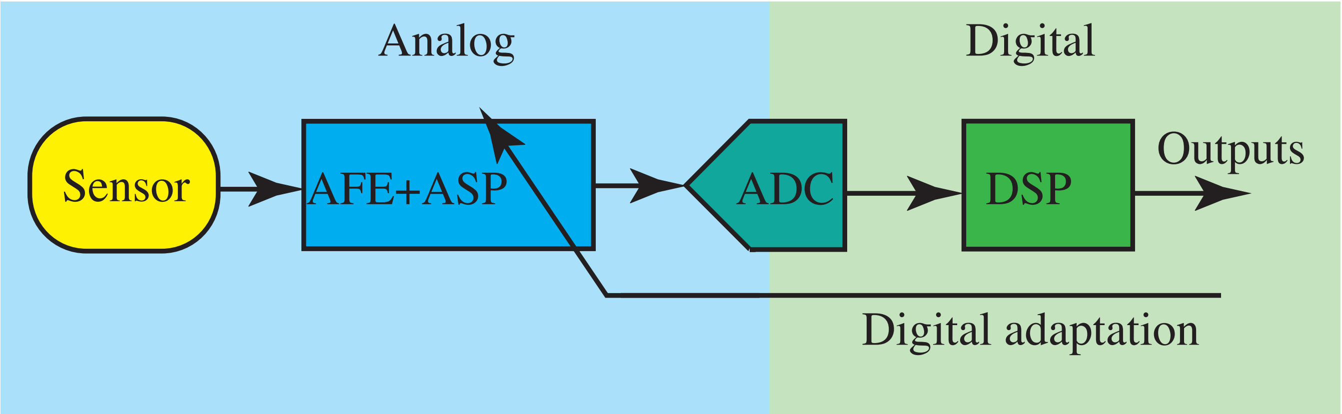


Charge-domain Analog Signal Processing

Low-dimensional, Low-resolution Digital Coding

Digital Adaptation

Linear Transform Analog and Mixed-Signal Sensory Processing

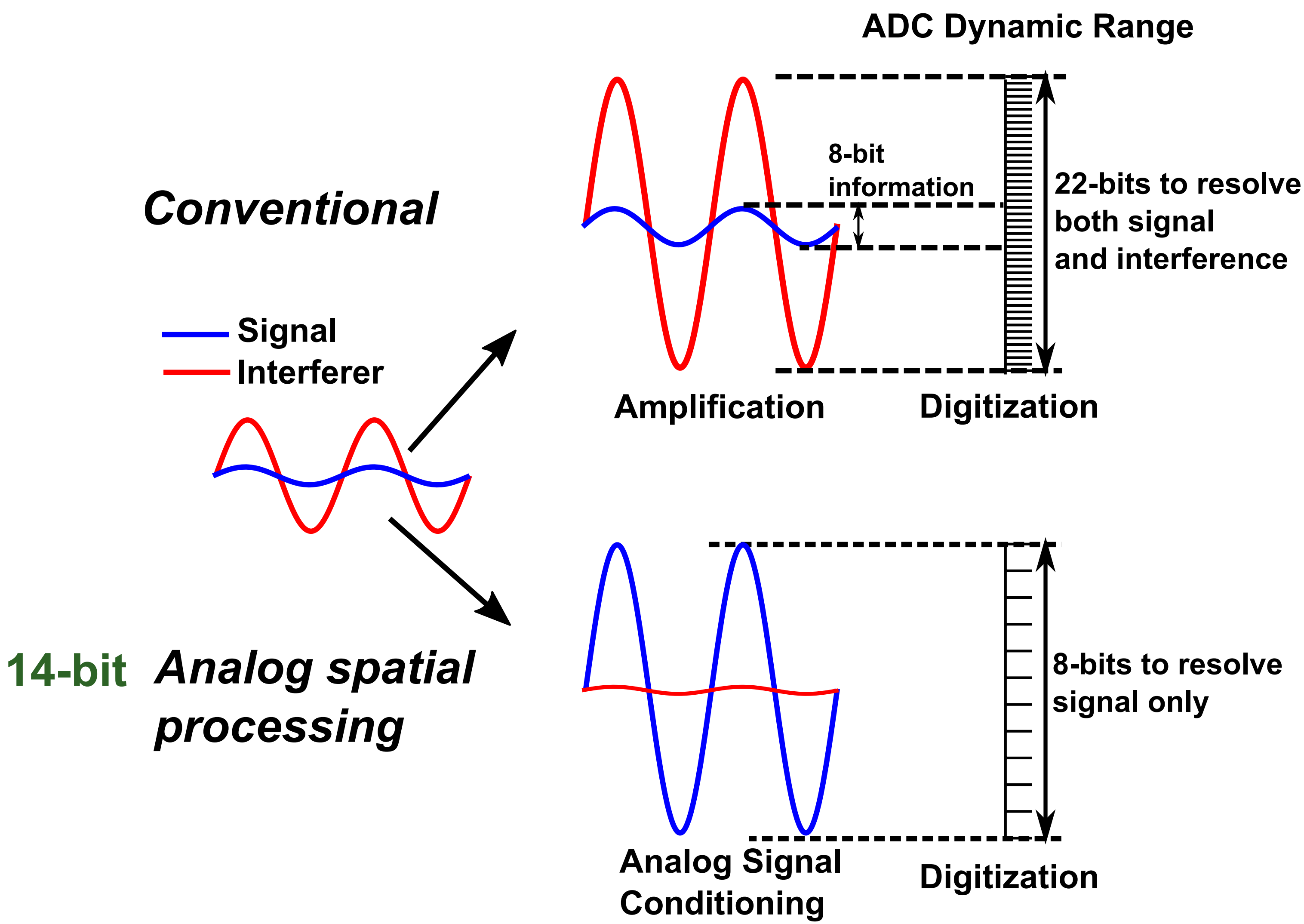


- Application Enabler
- Lower Power
- Analog ***processing gain*** lowers A/D requirements

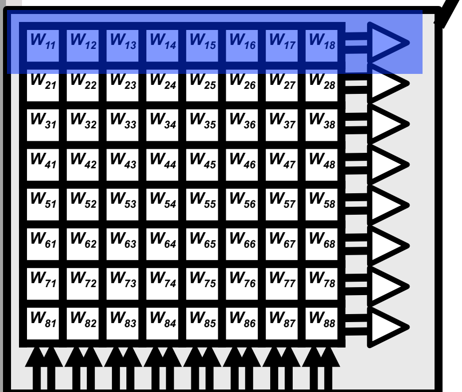
Processing gain: Improvement in SNR/DR due to ASP

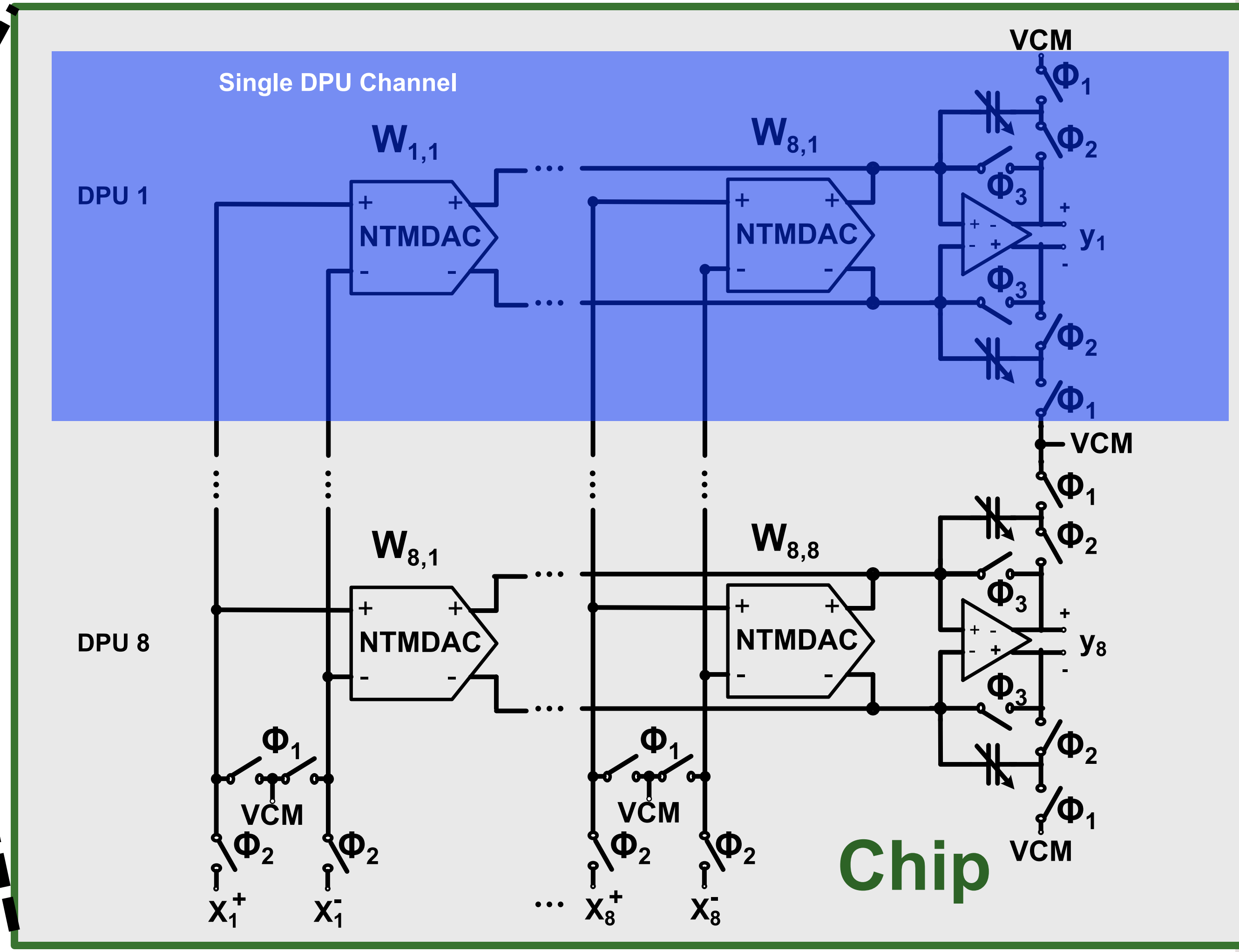
S. Joshi et al, “2pJ/MAC 14b 8×8 Linear Transform Mixed-Signal Spatial Filter in 65nm CMOS with 84dB Interference Suppression,” ISSCC 2017

Spatial Processing Gain



Dot Product Unit

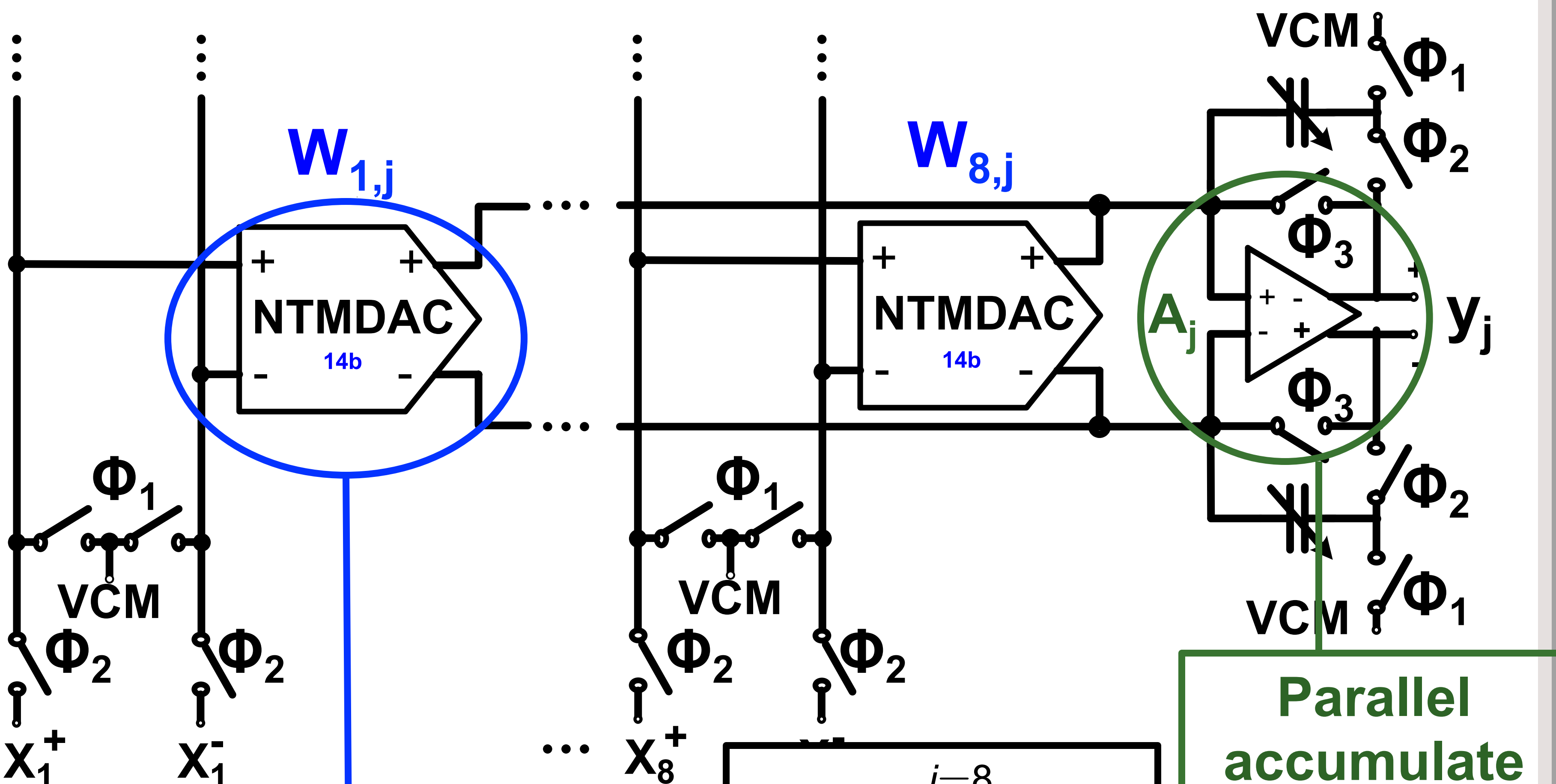




The spatial filter is composed of 8 arrayed Dot Product Units (DPUs)

S. Joshi et al, “2pJ/MAC 14b 8×8 Linear Transform Mixed-Signal Spatial Filter in 65nm CMOS with 84dB Interference Suppression,” ISSCC 2017

Dot Product Unit

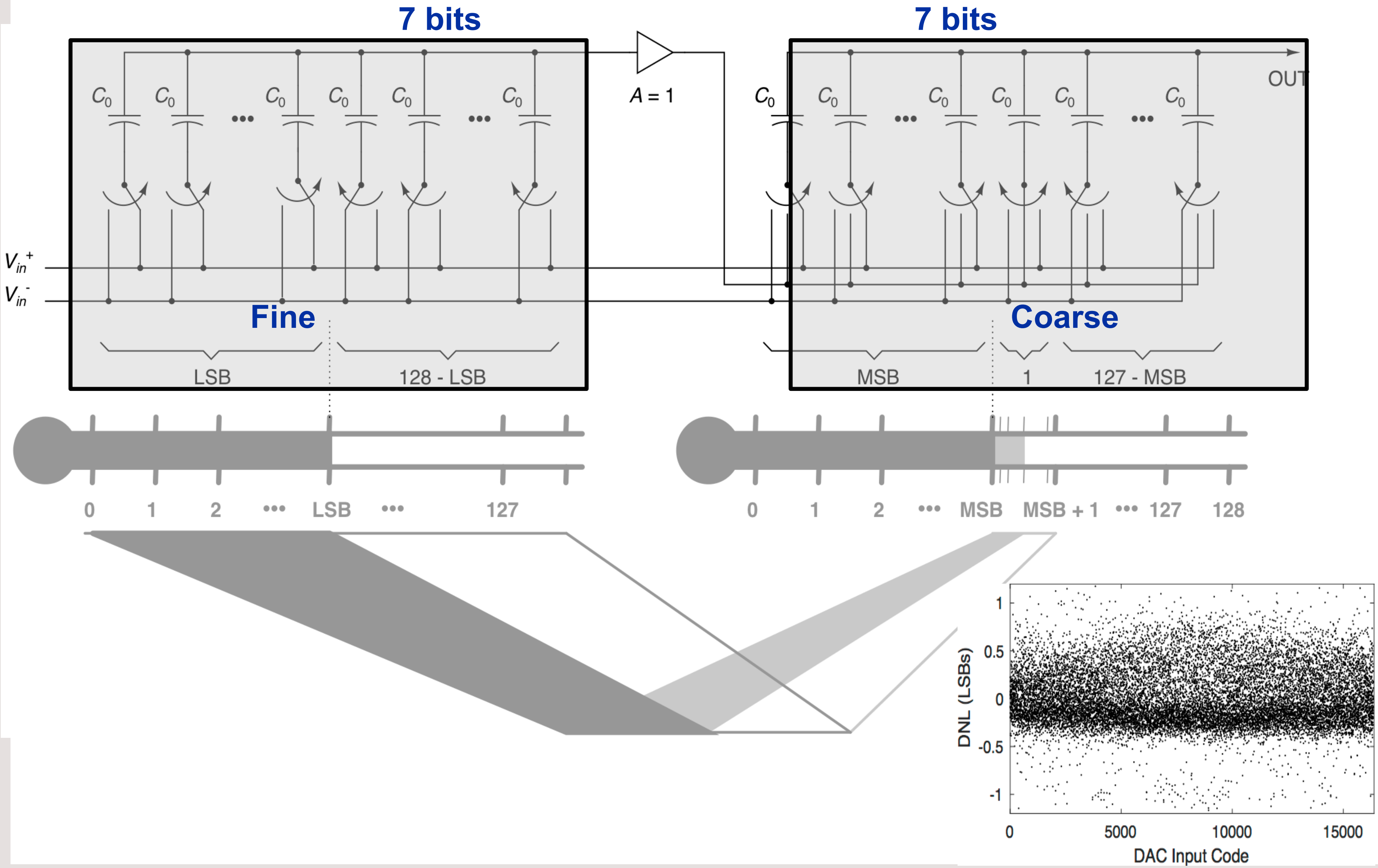


Nested Thermometer Multiplying DAC

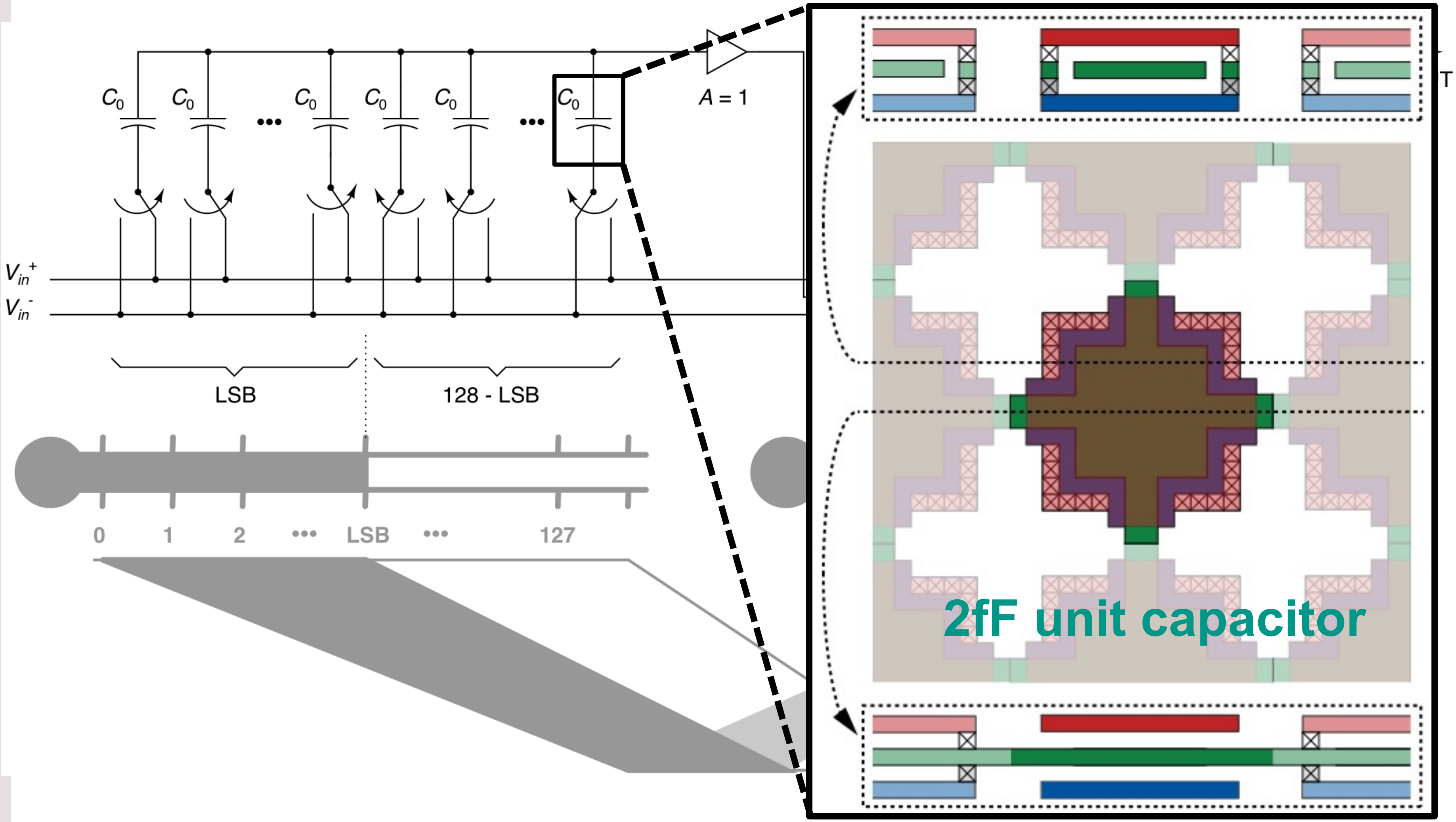
$$y_j = A_j \sum_{i=1}^{i=8} x_i W_{i,j}$$

Parallel accumulate at VGA input.

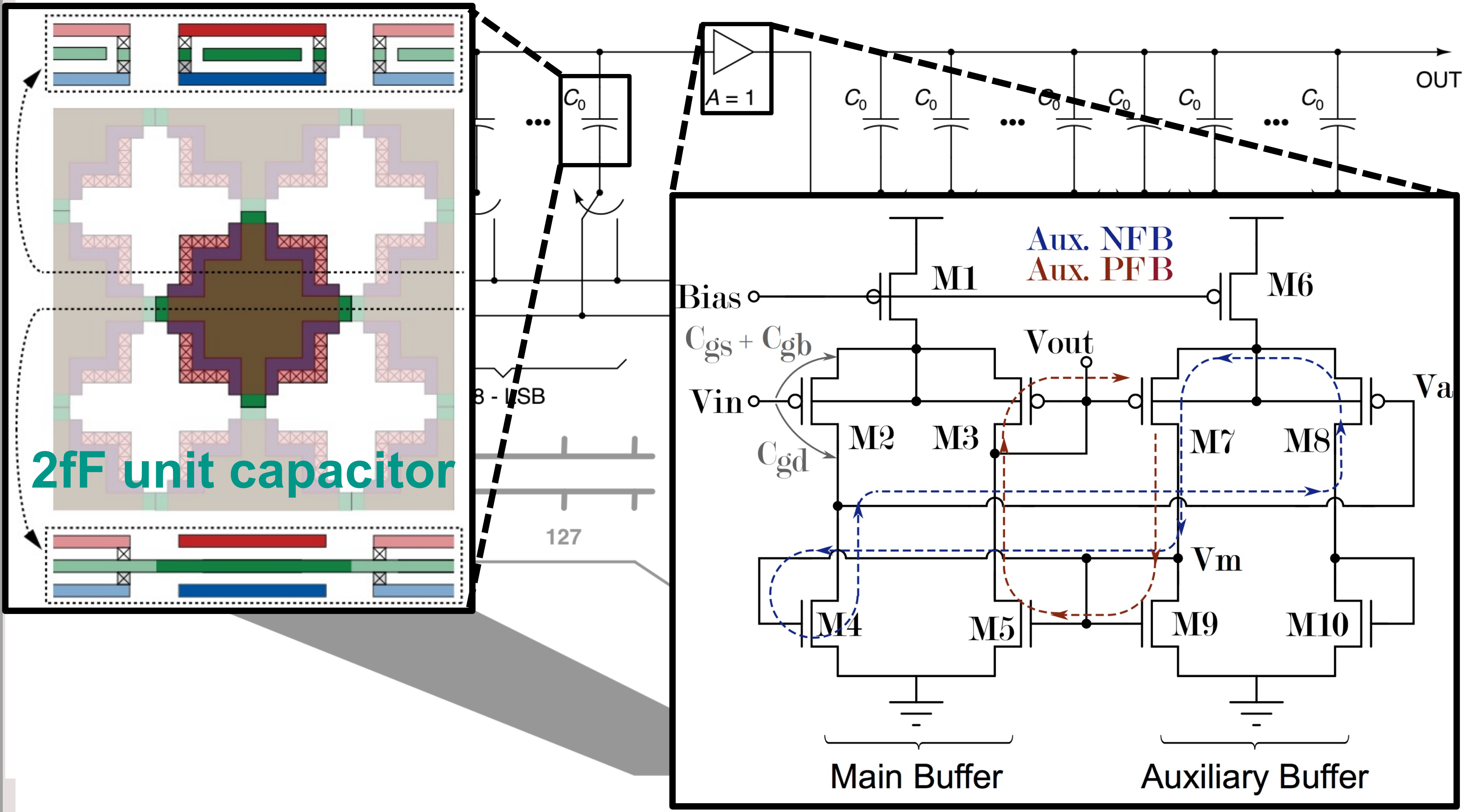
Nested Thermometer Multiplying DAC



Nested Thermometer Multiplying DAC

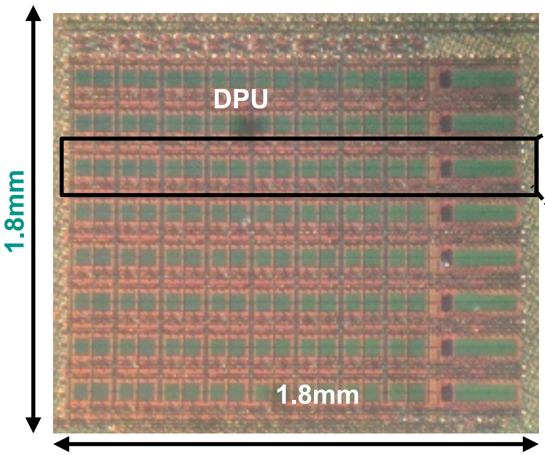


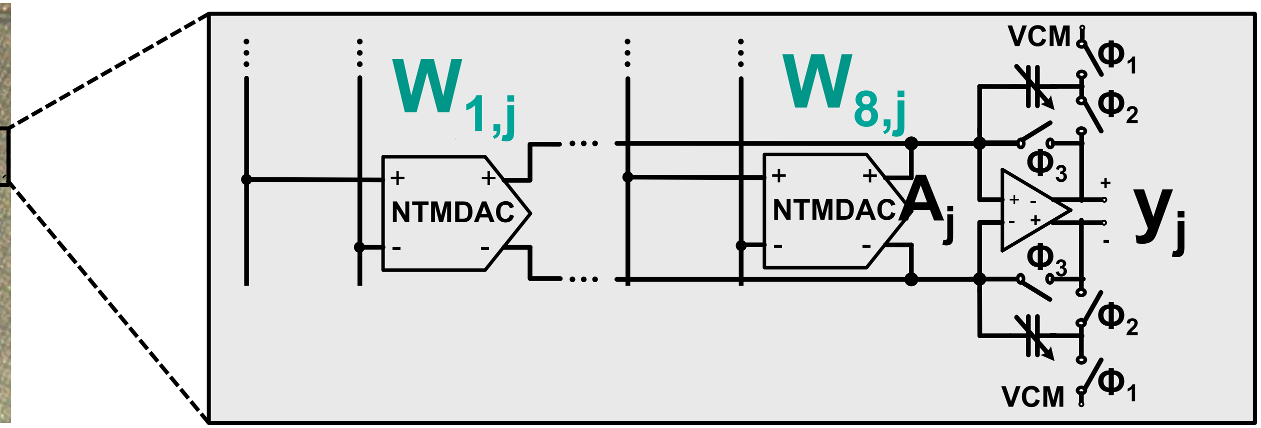
Nested Thermometer Multiplying DAC

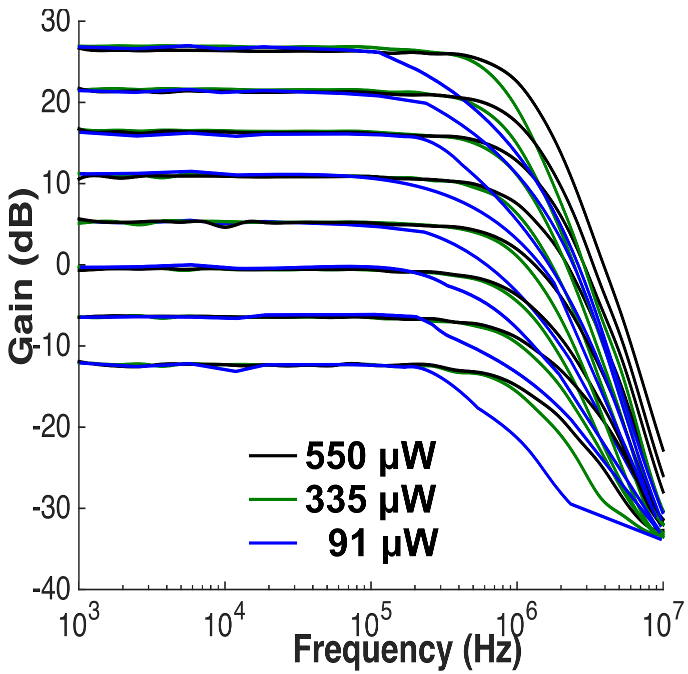


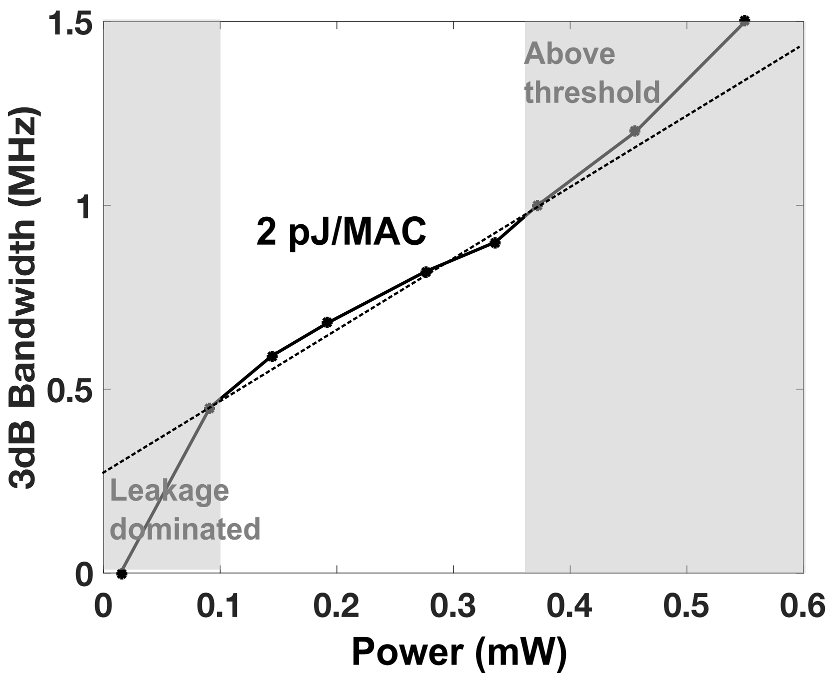
[Joshi TCAS-II 2016]

System Measurements

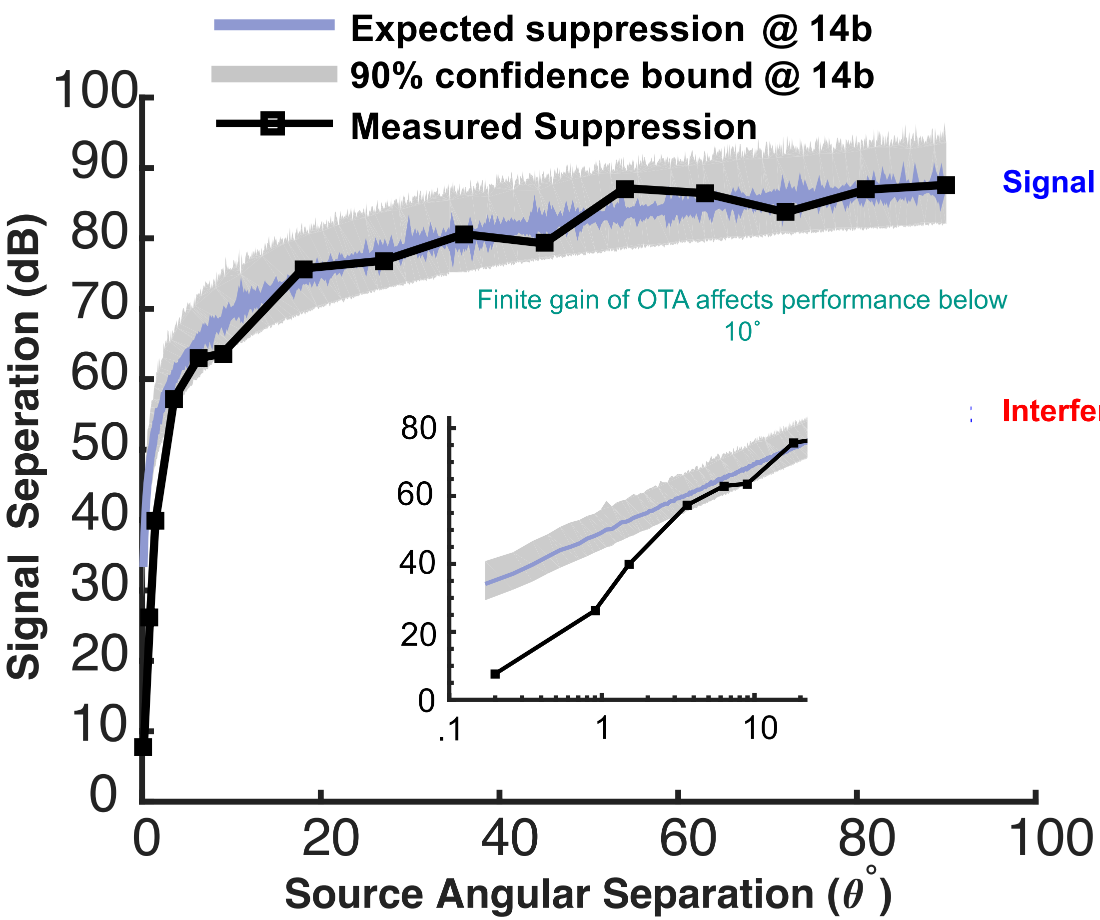


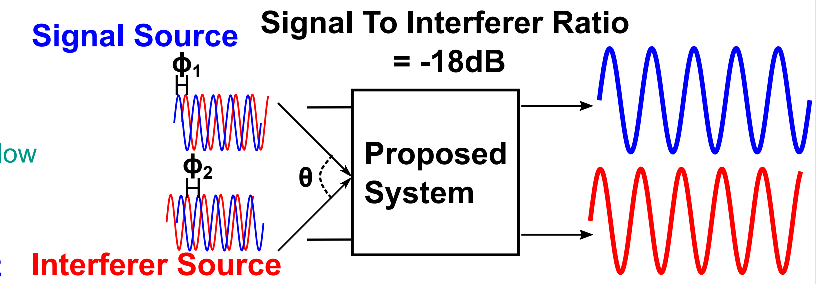






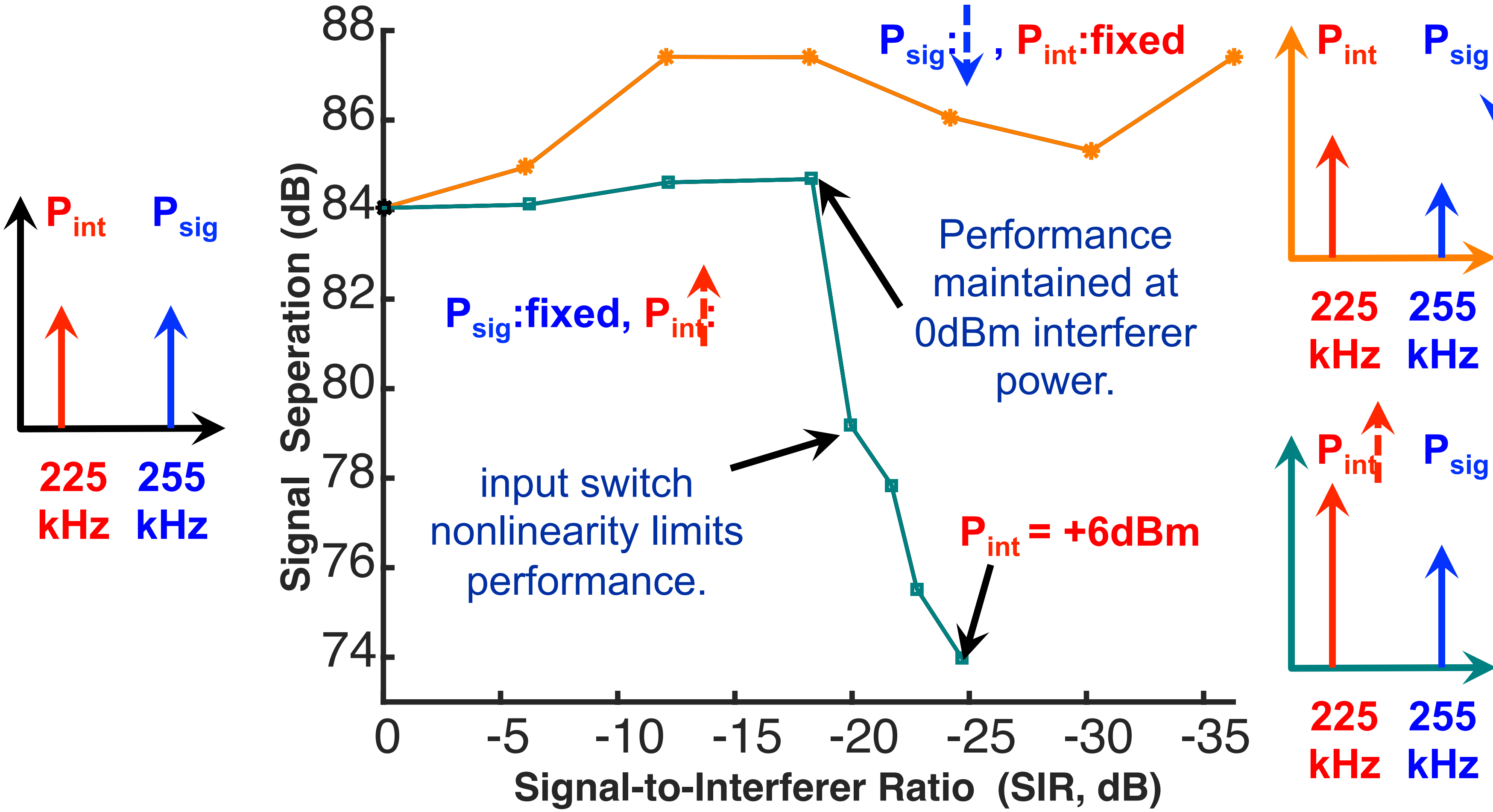
Measurements: Angular Resolution





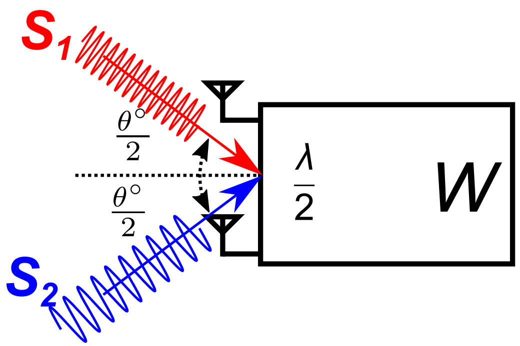
Experimental setup.

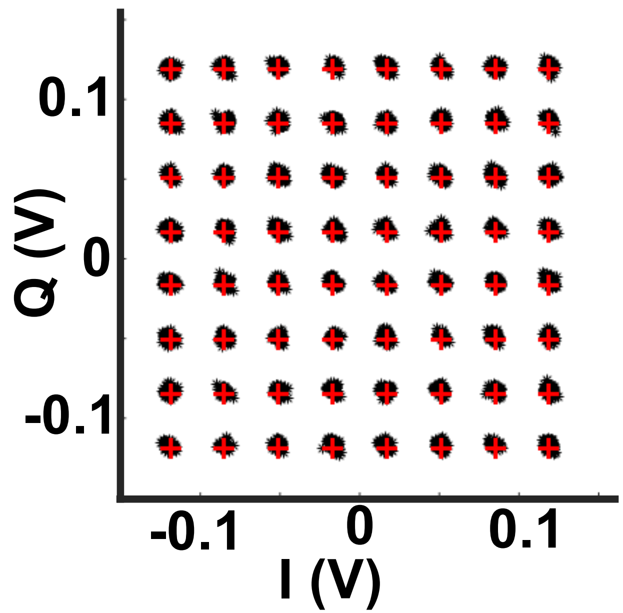
Measurements: SIR

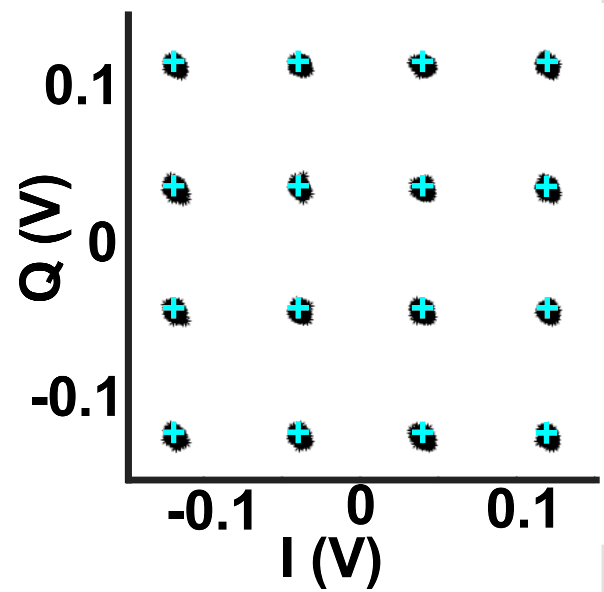


Application: MIMO Communication

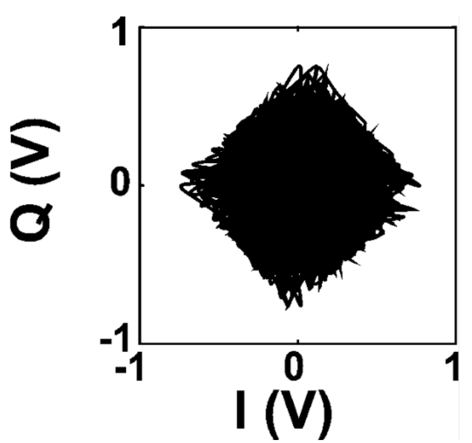
Spatial filtering to separate signal mixture







Constellation



64-QAM resolved
RMS EVM 2.9%

16-QAM resolved
RMS EVM 3.1%

Application: MIMO Communication

Beamforming Performance (baseband only)

	Tseng et. al. JSSC 2010	Ghaffari et. al. JSSC 2014	Kim et. al. JSSC 2015	This work
Received EVM (dB)	-25	-	-28.8	-30.8
Effective number of bits	5	5	8	14
Angular Resolution (°)	22.5	22.5	<5[a]	<1[a]
Interferer Cancellation (dB)	30[b]	15[b,c]	48[b]	>80[b]
CMOS Technology (nm)	90	65	65	65
Power at Baseband (mW)	10[d]	68-195[e]	1.3	0.396
Bandwidth at Baseband (MHz)	20	5	3	2.4

[a]Greater than 15 dB cancellation, [b]Cancellation at 45° angular separation, [c]Out of beam, [d]LO power only, [e]Total power reported baseband power not reported

S. Joshi et al, “2pJ/MAC 14b 8×8 Linear Transform Mixed-Signal Spatial Filter in 65nm CMOS with 84dB Interference Suppression,” ISSCC 2017

BENG 207 Neuromorphic Integrated Bioelectronics

Date	Topic
9/27, 9/29	Biophysical foundations of natural intelligence in neural systems. Subthreshold MOS silicon models of membrane excitability. Silicon neurons. Hodgkin-Huxley and integrate-and-fire models of spiking neuronal dynamics. Action potentials as address events.
10/4, 10/6	Silicon retina. Low-noise, high-dynamic range photoreceptors. Focal-plane array signal processing. Spatial and temporal contrast sensitivity and adaptation. Dynamic vision sensors.
10/11, 10/13	Silicon cochlea. Low-noise acoustic sensing and automatic gain control. Continuous wavelet filter banks. Interaural time difference and level difference auditory localization. Blind source separation and independent component analysis.
10/18, 10/20	Silicon cortex. Neural and synaptic compute-in-memory arrays. Address-event decoders and arbiters, and integrate-and-fire array transceivers. Hierarchical address-event routing for locally dense, globally sparse long-range connectivity across vast spatial scales.
10/28, 11/1	Review. Modular and scalable design for neuromorphic and bioelectronic integrated circuits and systems. Design for full testability and controllability.
11/1, 11/3	Midterm due 11/2. Low-noise, low-power design. Fundamental limits of noise-energy efficiency, and metrics of performance. Biopotential and electrochemical recording and stimulation, lab-on-a-chip electrophysiology, and neural interface systems-on-chip.
11/8, 11/10	Learning and adaptation to compensate for external and internal variability over extended time scales. Background blind calibration of device mismatch. Correlated double sampling and chopping for offset drift and low-frequency noise cancellation.
11/15, 11/17	Energy conservation. Resonant inductive power delivery and data telemetry. Ultra-high efficiency neuromorphic computing. Resonant adiabatic energy-recovery charge-conserving synapse arrays.
11/22, 11/24	Guest lectures
11/29, 12/1	Project final presentations. All are welcome!