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

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/11, 10/13</td>
<td>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.</td>
</tr>
<tr>
<td>10/28, 11/1</td>
<td>Review. Modular and scalable design for neuromorphic and bioelectronic integrated circuits and systems. Design for full testability and controllability.</td>
</tr>
<tr>
<td>11/8, 11/10</td>
<td>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.</td>
</tr>
<tr>
<td>11/22, 11/24</td>
<td>Guest lectures</td>
</tr>
<tr>
<td>11/29, 12/1</td>
<td>Project final presentations. All are welcome!</td>
</tr>
</tbody>
</table>
Large-Scale Mixed-Signal Sensory Computation

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

- **Scalable**
  silicone area and power scale linearly with throughput

- **Highly Efficient**
  factor 100 to 10,000 less energy/operation than DSP

- **Limited Precision**
  - analog mismatch and nonlinearity (WYDINWYG)
  - fix: adaptation in redundancy

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

Adaptive Elements

Adaptation:
- Autozeroing (high-pass filtering) \( \text{outputs} \)
- Offset Correction \( \text{outputs} \)
  - \textit{e.g.} Image Non-Uniformity Correction
- Equalization / Deconvolution \( \text{inputs, outputs} \)
  - \textit{e.g.} Source Separation; Adaptive Beamforming

Learning:
- Unsupervised Learning \( \text{inputs, outputs} \)
  - \textit{e.g.} Adaptive Resonance; LVQ; Kohonen
- Supervised Learning \( \text{inputs, outputs, targets} \)
  - \textit{e.g.} Least Mean Squares; Backprop
- Reinforcement Learning \( \text{reward/punishment} \)
Incremental Outer-Product Learning in Neural Nets

Multi-Layer Perceptron:

Outer-Product Learning Update:

- Hebbian (*Hebb, 1949*):

- LMS Rule (*Widrow-Hoff, 1960*):

- Backpropagation (*Werbos, Rumelhart, LeCun*):

\[
\Delta p_{ij} = \eta \ x_j \cdot e_i
\]

\[
e_i = x_i - f'(x_i)(x_i^{\text{target}} - x_i)
\]

\[
e_j = f'(x_j) \sum_i p_{ij} e_i
\]
Technology

Incremental Adaptation:
- Continuous-Time:
  \[ C \frac{d}{dt} V_{\text{stored}} = I_{\text{adapt}} \]
- Discrete-Time:
  \[ C \Delta V_{\text{stored}} = Q_{\text{adapt}} \]

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

\[ \Delta Q_{\text{adapt}} \]

\[ V_{\text{stored}} \]

\[ \Delta t = 40 \text{ msec} \]
\[ 1 \text{ msec} \]
\[ 23 \mu\text{sec} \]

\[ \Delta t = 0 \]

\[ \Delta t \]

\[ \Delta V_{\text{stored}} \]

\[ V_{\text{bn}} \]

\[ V_{\text{bp}} \]

\[ V_{\text{dd}} \]

\[ \text{GND} \]

\[ \text{Cauwenberghs, ALOG 1998} \]
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, I_{out} is exponential both in the floating gate charge, and in control voltage V_g.
Phase Change Memory Technology

- Analog switch
  - $100\Omega$ – $1M\Omega$ resistance range
- Fast write and read times (~nsec)
- Radiation hard

Reconfigurable Synaptic Connectivity and Plasticity
From Microchips to Large-Scale Neural Systems

Address-Event Representation

Neural Systems

Synaptic Plasticity & Wiring

Multi-Chip Systems
- ‘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)
Spike Timing-Dependent Plasticity

Bi and Poo, 1998
Spike Timing-Dependent Plasticity

*in the Address Domain*

Causal

Anti-Causal

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

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.

Event-Driven Contrastive Divergence

On-line Training of Boltzmann Machines Using STDP

- 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 $<vh>$ in on-line form
  - Modulation $g(t)$ controls wake-sleep phases (data vs. reconstruction)

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

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

Spiking Synaptic Sampling Machine (S$^3$M)

Biophysical Synaptic Stochasticity in Inference and Learning

- 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

Spike-Timing Dependent Eligibility

Reinforcement Learning by Reward Modulation of STDP

- 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

Charge-domain Analog Signal Processing

Low-dimensional, Low-resolution Digital Coding

Digital Adaptation

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

S. Joshi et al, ISSCC 2017
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**

Spatial Processing Gain

Conventional

14-bit Analog spatial processing

ADC Dynamic Range

22-bits to resolve both signal and interference

8-bit information

Amplification

Digitization

8-bits to resolve signal only

Analog Signal Conditioning

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

Dot Product Unit

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

Parallel accumulate at VGA input.

Nested Thermometer Multiplying DAC
Nested Thermometer Multiplying DAC

A folded cascode OTA, biased in subthreshold.

7 bits

Coarse

Fine

7 bits

MSB

127 - MSB

LSB

128 - LSB

\[ V_{\text{in}}^{+} \]

\[ V_{\text{in}}^{-} \]

\[ A = 1 \]

DAC Input Code

DNL (LSBs)
Nested Thermometer Multiplying DAC

A folded cascode OTA, biased in subthreshold.
nested thermometer multiplying DAC

2fF unit capacitor

[Image: Nested Thermometer Multiplying DAC diagram]

[Joshi TCAS-II 2016]
System Measurements

Power (mW)

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>10^3</th>
<th>10^4</th>
<th>10^5</th>
<th>10^6</th>
<th>10^7</th>
</tr>
</thead>
<tbody>
<tr>
<td>550 μW</td>
<td>20</td>
<td>15</td>
<td>10</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>335 μW</td>
<td>-10</td>
<td>-15</td>
<td>-20</td>
<td>-25</td>
<td>-30</td>
</tr>
<tr>
<td>91 μW</td>
<td>-30</td>
<td>-35</td>
<td>-40</td>
<td>-45</td>
<td>-50</td>
</tr>
</tbody>
</table>

Gain (dB)

- 550 μW
- 335 μW
- 91 μW

Frequency (Hz) vs. Gain (dB)

Above threshold
2 pJ/MAC
Leakage dominated

3dB Bandwidth (MHz)

Power (mW) vs. 3dB Bandwidth (MHz)
Measurements: Angular Resolution

Finite gain of OTA affects performance below 10°

Expected suppression @ 14b
90% confidence bound @ 14b
Measured Suppression

Signal Source
Signal To Interferer Ratio = -18dB

Interferer Source
Experimental setup.
Measurements: SIR

- Performance maintained at 0dBm interferer power.
- Input switch nonlinearity limits performance.
- $P_{\text{int}} = +6\text{dBm}$
Application: MIMO Communication

Spatial filtering to separate signal mixture

Constellation

\[ \frac{\theta^c}{2} \]

\[ \frac{\theta^o}{2} \]

\[ \lambda \]

\[ W \]

\[ \begin{align*} Q & \quad 0.1 \\ \bar{Q} & \quad 0 \\ -0.1 & \quad I (V) \end{align*} \]

64-QAM resolved
RMS EVM 2.9%

16-QAM resolved
RMS EVM 3.1%
## Application: MIMO Communication

### Beamforming Performance (baseband only)

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

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

---

# BENG 207 Neuromorphic Integrated Bioelectronics

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/11, 10/13</td>
<td>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.</td>
</tr>
<tr>
<td>10/28, 11/1</td>
<td>Review. Modular and scalable design for neuromorphic and bioelectronic integrated circuits and systems. Design for full testability and controllability.</td>
</tr>
<tr>
<td>11/8, 11/10</td>
<td>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.</td>
</tr>
<tr>
<td>11/22, 11/24</td>
<td>Guest lectures</td>
</tr>
<tr>
<td>11/29, 12/1</td>
<td>Project final presentations. All are welcome!</td>
</tr>
</tbody>
</table>