BENG 207 Special Topics in Bioengineering

Neuromorphic Integrated Bioelectronics

Week 1: Biophysical Foundations

Gert Cauwenberghs

Department of Bioengineering UC San Diego

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

Gert Cauwenberghs

BENG 207 Neuromorphic Integrated Bioelectronics

BENG 207 Neuromorphic Integrated Bioelectronics

Date	Торіс
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!

Lee Sedol vs. AlphaGo

Go World Champion vs. Google DeepMind ~ 100 W ~ 100 kW



Gert Cauwenberghs

Neuromorphic Engineering *"in silico" neural systems design*



Gert Cauwenberghs

The Computer and the Brain



John von Neumann



One of History's Hottest CPU Microchips Intel's Itanium 2



Source: IEEE ISSCC' 2002

The numbers ...

- 0.5 billion transistors in 120nm CMOS
- 1.6GHz clock, 64-bit instruction, 9MB L3 cache, 6.4GB/s I/O
- 2553 SPECfp_base2000
 (30% faster than 2.8GHz P4)
- 130 Watts
- ... and what they mean
 - Faster/cooler:
 - Scientific computing
 - Database search
 - Web surfing
 - Video games

What about intelligence?

Chips and Brains

- Itanium:
 - $3\,10^9$ floating op/s
 - 5 10⁸ transistors
 - 2 10⁹ Hz clock
 - 10^{10} Hz memory I/O
 - 128-b data bus @ 400MHz
 - 130 Watts

- Human brain:
 - 10^{15} synaptic op/s
 - 10¹⁵ synapses
- 1 Hz average firing rate
- 10¹⁰ Hz sensory/motor I/O
 - 10⁸ nerve fibers
- 25 Watts
- Silicon technology is approaching the *raw* computational power and bandwidth of the human brain.
- However, to emulate brain intelligence with chips requires a radical paradigm shift in computation:
 - Distributed representation in massively parallel architecture
 - Local adaptation and memory
 - Sensor and motor interfaces
 - Physical foundations of computing

Analysis by Synthesis



Richard Feynman



Carver Mead

Computational Systems Neuroscience

Brain 1 m

Systems

Maps





1 cm **Networks**







Neuromorphic Systems Engineering

1 Å



V_{Pre}

Φ

Ō

ð

0

V_{Post}

Drain

Synthesis

Multi-scale levels of investigation in analysis of the central nervous system (adapted from Churchland and Sejnowski 1992) and corresponding neuromorphic synthesis of highly efficient silicon cognitive microsystems. Boltzmann statistics of ionic and electronic channel transport provide isomorphic

G. Cauwenberghs, "Reverse Engineering the Cognitive Brain," PNAS, 2013

physical foundations.

Gert Cauwenberghs

Analysis

BENG 207 Neuromorphic Integrated Bioelectronics

gert@ucsd.edu

Physics of Computation

CMOS Silicon Technology





Voltage-dependent n-channel

- Electron transport between source and drain
- Gate controls energy barrier for electrons across the channel
- Boltzmann distribution of *electron energy* produces exponential *increase* in channel conductance with gate voltage





Cross-section of nMOS transistor in 0.18µm CMOS process (Intel, 2002)

Physics of Computation

CMOS Silicon Technology





Voltage-dependent p-channel

- Hole transport between source and drain
- Gate controls energy barrier for holes across the channel
- Boltzmann distribution of *hole energy* produces exponential *decrease* in channel conductance with gate voltage

Physics of Neural Computation

Silicon and Lipid Membranes Mead, 1989





Voltage-dependent *p*-channel

- Hole transport between source and drain
- Gate controls energy barrier for holes across the channel
- Boltzmann distribution of *hole energy* produces exponential *decrease* in channel conductance with gate voltage



Voltage-dependent conductance

- K^+/Na^+ transport across lipid bilayer
- Membrane voltage controls energy barrier for opening of ion-selective channels
- Boltzmann distribution of *channel energy* produces exponential increase in K^+/Na^+ conductance with membrane voltage

Physics of Neural Computation

Silicon and Biochemical Synapses Mead, 1989



Voltage-dependent *p*-channel

- Hole transport between source and drain
- Gate controls energy barrier for holes across the channel
- Boltzmann distribution of *hole energy* produces exponential *decrease* in channel conductance with gate voltage



(from Shepherd 1979)

Voltage-dependent quantal release

- K^+/Na^+ through postsynaptic membrane
- Presynaptic membrane voltage controls energy barrier for neurotransmitter release
- Boltzmann distribution in *quantal release* energy produces exponential dependence of postsynaptic K^+/Na^+ conductance

Why Develop "Neural" Silicon Chips?

Biology Motives:

- *In silico* emulation of neural and sensory-motor systems
 - Real-time computational power
 - Accounts for noise and imprecision in neural elements
- Analysis by synthesis
 - Emulating form and structure of neural systems provides better understanding, accounting for physical and architectural constraints
- Interfacing silicon with neurons and synapses in vivo
 - Allows to observe and control neural and synaptic activity

Engineering Motives:

- Efficiency of implementation
 - Lower power, smaller size
- Real-world interface
 - Integrated sensors and actuators
 - Analog, continuous-time dynamics
 - Intelligent brain-machine interfaces!

Neuromorphic Systems Design Flow



Silicon Model of Visual Cortical Processing



Neural model of boundary contour representation in V1, one orientation shown (Grossberg, Mingolla, and Williamson, 1997) Single-chip focal-plane implementation (Cauwenberghs and Waskiewicz, 1999)

Gert Cauwenberghs

Event-Coding Silicon Retina

Zaghloul and Boahen, 2006



- Models coding and communication of visual events in the mammalian retina and optic nerve
 - Integrated photosensors (rods)
 - On and off transient and sustained ganglia cell outputs
 - Spatiotemporal compressed coding and communication in optic nerve
 - Address-event coding of spikes

Change Threshold Detection APS CMOS Imager

Chi, Mallik, Clapp, Choi, Cauwenberghs and Etienne-Cummings (2007)



- Event-driven video compression
 - Change detection and threshold encoding on the focal plane
- 6T pixel combines APS and change event coding
- 4.3mW power at 3V and 30fps





Video Out



Change Events Out





BENG 207 Neuromorphic Integrated Bioelectronics

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



Gert Cauwenberghs

Scaling of Task and Machine Complexity



Achieving (or surpassing) human-level machine intelligence requires a convergence between:

- Advances in computing resources approaching connectivity and energy efficiency levels of computing and communication in the brain;
- Advances in deep learning methods, and supporting data, to adaptively reduce algorithmic complexity.

Scaling and Complexity Challenges

- Scaling the event-based neural systems to performance and efficiency approaching that of the human brain will require:
 - Scalable advances in silicon integration and architecture
 - Scalable, locally dense and globally sparse interconnectivity
 - *Hierarchical address-event routing*
 - High density (10¹² neurons, 10¹⁵ synapses within 5L volume)
 - Silicon nanotechnology and 3-D integration
 - High energy efficiency (10¹⁵ synOPS/s at 15W power)
 - Adiabatic switching in event routing and synaptic drivers
 - Scalable models of neural computation and synaptic plasticity
- Neuro

CS

EE

NanoE

Phys

- Convergence between cognitive and neuroscience modeling
 - Modular, neuromorphic design methodology
- **CogSci** Data-rich, environment driven evolution of machine complexity

Large-Scale Reconfigurable Neuromorphic Computing Technology and Performance Metrics

	Stromatias 2013 SpiNNaker Manchester	Davies 2018 Loihi Intel	Merolla 2014 TrueNorth IBM	Schemmel 2010 FACETS/BrainScaleS Heidelberg	Benjamin 2014 NeuroGrid Stanford	Park 2014 IFAT UCSD
Technology (nm)	130	14	28	180	180	90
Die Size (mm ²)	102	60	430	50	168	16
Neuron Type	Digital Arbitrary	Digital Conductance Integrate & Fire	Digital Accumulate & Fire	Analog Conductance Integrate & Fire	Analog Shared-Dendrite Conductance I&F	Analog 2-Compartment Conductance I&F
# Neurons	5216 ¹	128k ²	1M ²	512	65k	65k
Neuron Area (μm²)	N/A ¹	240 (240k) ²	14 (3325) ²	1500	1800	140
Peak Throughput (Events/s)	5M	3.4G	1G	65M	91M	73M
Energy Efficiency (J/SynEvent)	8n	24p	26p	N/A	31p	22p

¹ Software-instantiated neuron model

² Time-multiplexed neuron processor

- Benjamin, B., P. Gao, E. McQuinn, S. Choudhary, A. Chandrasekaran, J. Bussat, R. Alvarez-Icaza, J. Arthur, P. Merolla, and K. Boahen, "Neurogrid: A mixed analog-digital multichip system for large-scale neural simulations," *Proc. IEEE*, 102(5):699–716, 2014.
- Davies, M. et al., "Loihi: A Neuromorphic Manycore Processor with On-Chip Learning," IEEE Micro, vol. 38 (1), pp. 82-99, 2018.
- Merolla, P.A., J.V. Arthur, R. Alvarez-Icaza, A S. Cassidy, J. Sawada, F. Akopyan, B.L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S.K. Esser, R. Appuswamy, B. Taba, A. Amir, M.D. Flickner, W.P. Risk, R. Manohar, and D. S. Modha, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, 345(6197):668–673, 2014.
- Park, J., S. Ha, T. Yu, E. Neftci, and G. Cauwenberghs, "65k-neuron 73-Mevents/s 22-pJ/event asynchronous micro-pipelined integrate-and-fire array transceiver," Proc. 2014 IEEE Biomedical Circuits and Systems Conf. (BioCAS), 2014.
- Schemmel, J., D. Bruderle, A. Grubl, M. Hock, K. Meier, and S. Millner, "A waferscale neuromorphic hardware system for large-scale neural modeling," *Proc. 2010 IEEE Int. Symp. Circuits and Systems (ISCAS)*, 1947–1950, 2010.
- Stromatias, E., F. Galluppi, C. Patterson, and S. Furber, "Power analysis of largescale, real-time neural networks on SpiNNaker," *Proc. 2013 Int. Joint Conf. Neural Networks (IJCNN)*, 2013.

Large-Scale Reconfigurable Neuromorphic Computing



"CRI: CI-NEW: Trainable Reconfigurable Development Platform for Large-Scale Neuromorphic Cognitive Computing," National Science Foundation CNS-1823366, G. Cauwenberghs (PI), E. Neftci, and A. Majumdar, 8/2018-7/2021.

Provides open access to large-scale reconfigurable neuromorphic computing hardware and software as an experimental testbed and development platform with up to 128M neurons and 32B synapses for the research community at large.



Neural-Synaptic Array Transceiver (Detorakis et al, Frontiers in Neuroscience, 2018)

NeuroDyn (2021 Telluride Neuromorphic Workshop)

Closing the Loop: Interactive Neural/Artificial Intelligence



Gert Cauwenberghs

gert@ucsd.edu



Neuromorphic Systems Engineering

F. Broccard, S. Joshi, J. Wang and G Cauwenberghs, "Neuromorphic neural interfaces: from neurophysiological inspiration to biohybrid coupling with nervous systems," *JNE*, 2017

Gert Cauwenberghs

Integrated Systems Neuroengineering



BENG 207 Neuromorphic Integrated Bioelectronics

Silicon Neurons



Hodgkin-Huxley neuronal dynamics Integrate-and-fire models Action potentials as address events

Gert Cauwenberghs

BENG 207 Neuromorphic Integrated Bioelectronics

gert@ucsd.edu

Nernst Potentials



Gert Cauwenberghs

lon

 K^+

CI-

Reverse Engineering the Cognitive Brain in Silicon

Nernst-Planck Equilibrium



Membrane Capacitance



BENG 260 Neurodynamics — <u>http://isn.ucsd.edu/courses/beng260</u>

Membrane Conductance



Voltage Gated Ion Channels

n: slow K+ activation; **m**: fast Na+ activation; **h**: slow Na+ inactivation



Figure 5.8 Gating of membrane channels. In both figures, the interior of the neuron is to the right of the membrane, and the extracellular medium is to the left. **(A)** A cartoon of gating of a persistent conductance. A gate is opened and closed by a sensor that responds to the membrane potential. The channel also has a region that selectively allows ions of a particular type to pass through the channel, for example, K+ ions for a potassium channel. **(B)** A cartoon of the gating of a transient conductance. The activation gat is coupled to a voltage sensor (denoted by a circled +) and acts like the gate in A. A second gate, denoted by the ball, can block that channel once it is open. The top figure shows the channel in a deactivated (and deinactivated) state. The middle panel shows an activated channel, and the bottom panel shows an inactivated channel. Only the middle panel corresponds to an open, ion-conducting state. (A from Hille, 1992; B from Kandel et al., 1991.)

Dayan & Abbott 2001, p. 169

BENG 260 Neurodynamics — <u>http://isn.ucsd.edu/courses/beng260</u>

Gert Cauwenberghs

Hodgkin-Huxley Model

Squid axon:

$$C_{m} \frac{dV_{m}}{dt} = I_{ext} - \underbrace{\bar{g}_{K} n^{4} (V_{m} - E_{K})}_{I_{K}} - \underbrace{\bar{g}_{Na} m^{3} h (V_{m} - E_{Na})}_{I_{Na}} - \underbrace{g_{L} (V_{m} - E_{L})}_{I_{L}}$$

$$C_{m} = 1\mu F/cm^{2} \qquad E_{K} = -12 \, mV \qquad E_{Na} = 120 \, mV \qquad E_{L} = 10.6 \, mV$$

$$\bar{g}_{K} = 36 \, mS/cm^{2} \qquad \bar{g}_{Na} = 120 \, mS/cm^{2} \qquad g_{L} = 0.3 \, mS/cm^{2}$$

$$(mS = \mu A/mV)$$

$$\begin{array}{rcl} \frac{dn}{dt} = & \alpha_n \left(1 - n\right) - \beta_n n = & \frac{n_{\infty} - n}{\tau_n}; & \alpha_n \left(V_m\right) = & \frac{10 - V_m}{100 \left(e^{1 - V_m / 10} - 1\right)}; & \beta_n \left(V_m\right) = & \frac{1}{8} e^{-V_m / 80} \\ \frac{dm}{dt} = & \alpha_m \left(1 - m\right) - \beta_m m = & \frac{m_{\infty} - m}{\tau_m}; & \alpha_m \left(V_m\right) = & \frac{25 - V_m}{100 \left(e^{2.5 - V_m / 10} - 1\right)}; & \beta_m \left(V_m\right) = & 4e^{-V_m / 18} \\ \frac{dh}{dt} = & \alpha_h \left(1 - h\right) - \beta_h h = & \frac{h_{\infty} - h}{\tau_h}; & \alpha_h \left(V_m\right) = & \frac{7}{100} e^{-V_m / 20}; & \beta_h \left(V_m\right) = & \frac{1}{1 + e^{3 - V_m / 10}} \end{array}$$

 α_x and β_x in units 1/ms; V_m in units mV* V_m slightly shifted by 65mV so that $E_{rest} \equiv 0mV$

Hodgkin & Huxley, 1952

BENG 260 Neurodynamics — http://isn.ucsd.edu/courses/beng260

Gert Cauwenberghs

Hodgkin-Huxley Model



BENG 260 Neurodynamics — http://isn.ucsd.edu/courses/beng260

Postsynaptic Conductance

Postsynaptic membrane current and voltage:

...

$$C_m \frac{dV_m}{dt} = I_{ext} - I_K - I_{Na} - I_L - \sum_i I_{syn_i}$$
$$I_{syn_i} = g_{syn_i}(t) \cdot (V_m - E_{syn_i})$$



BENG 260 Neurodynamics — http://isn.ucsd.edu/courses/beng260

NeuroDyn: Biophysical Neurodynamics in Analog VLSI Yu and Cauwenberghs 2009

Generalized Hodgkin-Huxley neural and conductance-based synapse membrane dynamics in silico:



Each parameter is individually addressable and programmable through 10-bit DACs.

NeuroDyn: Biophysical Neurodynamics in Analog VLSI

Yu and Cauwenberghs 2009



Programmable Parameters: 384 total

Neurons V_i $\alpha_{n_i}(V)$ m_i h_i $4x3x7^*$

Synapses s_{ii}

 $\alpha_{r_{ij}}(V_{pre})^{2}$ 12x7*



*All rates α , β are 7-point sigmoidal spline regression functions $\alpha_{.}(V_{k}), \beta_{.}(V_{k}), \quad k = 1,...7$

E_{Na} Ki Li

The NeuroDyn Board consists of 4 neurons fully connected through 12 synapses. All parameters are individually programmable and have biophysically-based parameters governing the conductances, reversal potentials, and voltage-dependance of the channel kinetics.





Recorded dynamics of action potential and channel kinetics for one HH neuron.

NeuroDyn Synaptic Coupling



Uncoupled

Mutual inhibitory synaptic coupling

Generalized Map-Based I&F Neural Dynamics

Izhikevich 2003; Rulkov, Timofeev & Bazhenov 2004; Mihalas & Niebur 2009



Electronic version of the figure and reproduction permissions are freely available at www.izhikevich.com

Generalized HH/ML Neural Dynamics



NeuroDyn Tonic Spiking



NeuroDyn analog emulation

NeuroDyn Phasic Spiking



NeuroDyn analog emulation

NeuroDyn Tonic Bursting



NeuroDyn analog emulation

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

Address-Event Representation (AER)

Lazzaro et al., 1993; Mahowald, 1994; Deiss 1994; Boahen 2000

- AER emulates extensive connectivity between neurons by communicating spiking events time-multiplexed on a shared data bus.
- Spikes are represented by two values:
 - Cell location (address)
 - Event time (implicit)
- All events within Δt are "simultaneous"

Silicon Membrane Array Transceiver

Vogelstein, Mallik and Cauwenberghs, 2004

- Voltage-controlled membrane conductance
 - Event-driven activation
 - Dynamically reconfigurable:
 - *conductance g*
 - driving potential E

 Address-event encoding of pre-and post-synaptic action potentials

Silicon Membrane Circuit

Goldberg, Cauwenberghs and Andreou, 2000 Vogelstein, Mallik and Cauwenberghs, 2004

g_i(t) ion-specific membrane conductance

 E_i ion-specific reversal potential

Hierarchical Vision and Saliency-Based Acuity Modulation

Vogelstein, Mallik, Culurciello, Cauwenberghs, and Etienne-Cummings, NECO 2007

OR image

Simple cell response

Saliency map

Large-Scale Reconfigurable Neuromorphic Computing

Hierarchical Address-Event Routing (HiAER) Integrate-and-Fire Array Transceiver (IFAT) for scalable and reconfigurable neuromorphic neocortical processing. (a) Biophysical model of neural and synaptic dynamics. (b) Dynamically reconfigurable synaptic connectivity is implemented across IFAT arrays of addressable neurons by routing neural spike events locally through DRAM synaptic routing tables. (c) Each neural cell models conductance based membrane dynamics in proximal and distal compartments for synaptic input with programmable axonal delay, conductance, and reversal potential. (d) Multiscale global connectivity through a hierarchical network of HiAER routing nodes. (e) HiAER-IFAT board with 4 IFAT custom silicon microchips, serving 256k neurons and 256M synapses, and spanning 3 HiAER levels (L0-L2) in connectivity hierarchy. (f) The IFAT neural array multiplexes and integrates (top traces) incoming spike synaptic events to produce outgoing spike neural events (bottom traces). The latest IFAT microchip measured energy consumption is 22 pJ per spike event, several orders of magnitude more efficient than emulation on CPU/GPU platforms.

Yu et al, BioCAS 2012; Park et al, BioCAS 2014; Park et al, TNNLS 2017; Broccard et al, JNE 2017

BENG 207 Neuromorphic Integrated Bioelectronics

Date	Торіс
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!