Silicon and Biological Adaptive Neural Circuits

Gert Cauwenberghs Department of Bioengineering Institute for Neural Computation UC San Diego

Neuromorphic Engineering *"in silico" neural systems design*



Silicon and Biological Adaptive Neural Circuits

Today's Hottest Microchip

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:



- 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

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



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

gert@ucsd.edu

NeuroDyn: Biophysical Neurodynamics in Analog VLSI



Yu and Cauwenberghs 2009







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

Programmable Parameters: 384 total

Neurons V_i



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

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.

NeuroDyn Models and Architecture



The *NeuroDyn* chip emulates detailed neural and synaptic dynamics in silicon by implementing rate-based models of voltage-gated and ligand-gated channel kinetics.



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

NeuroDyn Synaptic Coupling



Mutual inhibitory synaptic coupling

Generalized Map-Based 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 Phasic Spiking



NeuroDyn analog emulation

NeuroDyn Tonic Bursting



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 _



Fast Rotation





Slow Rotation



Change Events Out

Video Out

gert@ucsd.edu

Change Detection APS: Compression and Reconstruction

Frame 0



Silicon and Biological Adaptive Neural Circuits

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

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



Silicon and Biological Adaptive Neural Circuits

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"

Address-Event Synaptic Connectivity

Goldberg, Cauwenberghs and Andreou, 2000



Silicon Membrane Array Transceiver

Vogelstein, Mallik and Cauwenberghs, 2004

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



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



Silicon Membrane Circuit



g_i(t) ion-specific membrane conductance

E_i ion-specific driving potential



Reconfigurable Silicon Large-Scale Neural Emulator

Vogelstein, Mallik and Cauwenberghs, 2007



- 9,600 neurons
 - 4 silicon membrane chips (IFAT)
- 4 million, 8-bit "virtual" synapses
 - 128MB (32bX4M) non-volatile **RAM**
- 1 million synaptic updates per second
 - 200MHz Spartan II Xilinx FPGA "MCU"
- Dynamically reconfigurable
 - Rewiring and synaptic plasticity (STDP etc.)
 - Driving potential (**DAC**) and conductance (**IFAT**)





Hierarchical Vision and Saliency-Based Acuity Modulation

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



Spike Timing-Dependent Plasticity



Bi and Poo, 1998

Spike Timing-Dependent Plasticity

in the Address Domain



Spike Timing-Dependent Plasticity on the IFAT

Vogelstein et al, NIPS*2002



Scaling of Task and Machine Complexity



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

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

Example: Board Games (Chess and Go)



- Complexity of typical strategic board games precludes exact solution through complete tree search for all but the simplest games (smallest boards).
 - Chess and Go are EXPTIME-complete: perfect strategy requires search time exponential in board size.
- Humans handle game complexity by pattern recognition and sequence recall, rather than tree search, acquired through extensive experience.
 - Novices routinely defeat computer Go, which fails to "see" the board like humans.
 - The need to "see" board patterns calls for adaptive neuromorphic approaches.

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

EE NanoE Phys

CS

CogSci

3-D Integrated Silicon Neuromorphic Processor

Park, Joshi, Yu, Maier, and Cauwenberghs, 2010





Top metal

TSV

Hierarchical address-event routing (HiAER)

Top metal

I/O pad

- 65,000, two-compartment neurons

 Conductance-based integrate and fire
 - array transceiver (**IFAT**)
 - 65 million, 32-bit "virtual" synapses
 - Conductance-based dynamical synapses
 - Dynamic table-look in embedded memory (2Gb **DRAM)**
 - Locally dense, globally sparse synaptic interconnectivity
 - Hierarchical address-event routing (HiAER)
 - Dynamically reconfigurable
 - Asynchronous spike event I/O interface



LIFAT (Analog CMOS)

HAER (Digital CMOS)

Phase Change Memory (PCM) Nanotechnology



Intel/STmicroelectronics (Numonyx) 256Mb multi-level phase-change memory (PCM) [Bedeschi et al, 2008]. Die size is 36mm2 in 90nm CMOS/Ge2Sb2Te5, and cell size is 0.097µm2. (a) Basic storage element schematic, (b) active region of cell showing crystalline and amorphous GST, (c) SEM photograph of array along the wordline direction after GST etch, (d) I-V characteristic of storage element, in set and reset states, (e) programming characteristic, (f) I-V characteristic of pnp bipolar selector.

- Scalable to high density and energy efficiency
 - < 100nm cell size in 32nm CMOS
 - < pJ energy per synapse operation

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

Inherently 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



Trainable Modular Vision Systems: The SVM Approach Papageorgiou, Oren, Osuna and Poggio, 1998



Support vector machine (SVM) classification for pedestrian and face object detection

- Support vector machine (SVM) with mathematical foundations in *Statistical Learning Theory* (Vapnik, 1995)
- The training process selects a small fraction of prototype *support vectors* from the data set, located at the *margin* on both sides of the classification boundary (e.g., barely faces vs. barely non-faces)



Gert Cauwenberghs

Silicon and Biological Adaptive Neural Circuits

gert@ucsd.edu

Trainable Modular Vision Systems: The SVM Approach Papageorgiou, Oren, Osuna and Poggio, 1998

Testing

Training







Overcomplete Representation Overcomplete Representation SVM Classifier person non-person ROC curve for various image representations and dimensions The number of support vectors, in relation to the number of training samples and the vector dimension, determine the generalization performance

 Both training and runtime performance are severely limited by the computational complexity of evaluating kernel functions

Gert Cauwenberghs

Silicon and Biological Adaptive Neural Circuits

Kerneltron: Adiabatic Support Vector "Machine"

Karakiewicz, Genov and Cauwenberghs, 2007





Karakiewicz, Genov, and Cauwenberghs, VLSI' 2006; CICC' 2007

- 1.2 TMACS / mW
 - adiabatic resonant clocking _ conserves charge energy
 - energy efficiency on par with _ human brain (10¹⁵ SynOP/S at 15W)



 $Vout_{m}^{(i)}$

Vdd/2 Vdd

Vdd/2

Vdd n

Vdd/2

Compute Vdd

0

M3

CID

M2

Resonant Charge Energy Recovery



Sub-Micropower Analog VLSI Adaptive Sequence Decoding

Chakrabartty and Cauwenberghs, 2004



Forward decoding MAP sequence estimation



Biometric verification





Adaptive Machine Intelligence

Training Machines towards Human Performance through Games



Silicon and Biological Adaptive Neural Circuits

Competitive Games: Humans and Machines



- Learning through experience in two-player zero-sum games:
 - Humans to humans: Novices learn from experts to become experts.
 - Humans to machines: Towards human-level machine performance.
 - Machines to machines: Beyond human-level machine performance.
- Heterogeneous competitive ranking:
 - ELO score ranks humans and machines alike.
 - *Turing test.*

Web-Based Competitive Games Human Players



- Existing, extensively developed game infrastructure
- Readily available, large pool of human subjects

Web-Based Competitive Games Humans and Machines



- Event codec adapter and machine interface
- Central logging, ranking, and matchmaking at external game server

Web-Based Competitive Games Humans Tutoring Machines



- Machine learns by observing actions *and* internal representation (EEG brain activity) of human expert.
- Neuromorphic: trained machine approaches human brain function *and* form.

Extensions of Interface and Benchmark Infrastructure General Game Environments



Game boxes

- Specialized computers with advanced graphics for games
 - Virtual game environments
 - Multi-player capable through internet
- Examples:
 - Sony Playstation II
 - Microsoft Xbox 360
 - Nintendo Wii

Robots

- Physical interface to sensory input and motor output
 - Real-world game environments
- Examples:
 - NSI Darwin
 - K-Team Khepera III
 - WowWee Robosapien



http://www.gearsofwar.com

http://www.wowwee.com

gert@ucsd.edu

Closing the Loop: Interactive Neural/Artificial Intelligence



Gert Cauwenberghs

Silicon and Biological Adaptive Neural Circuits

gert@ucsd.edu



NSF EFRI 2012 – Mind, Machines and Motor Control (M3C)

Distributed Brain Dynamics of Human Motor Control



G. Cauwenberghs, K. Kreutz-Delgado, T.P. Jung, S. Makeig, H. Poizner, T. Sejnowski, F. Broccard, D. Peterson, M. Arnold, A. Akinin, C. Stevenson, J. Menon



Brain Computer Interfaces and Motor Control



The brain's motor commands ...

- Parietal/frontal cortex
 - Implanted electrodes
 - Electroencephalogram (EEG)
 - Cortical signals, noninvasive
 - Low bandwidth (seconds)
- Nerve signals
 - Spinal cord electrodes
 - Electromyogram (EMG)
 - Muscle signals, noninvasive
 - Higher bandwidth (milliseconds)

. translated into motor actions

- Machine learning/signal processing
- Neuromorphic approaches
 - Central pattern generators (CPGs)

Nicolelis, Nature Rev. Neuroscience 4, 417, 2003

Wireless Non-Invasive, Orthotic Brain Machine Interfaces



Calit2 StarCAVE immersive 3-D virtual reality environment



Yu Mike Chi, 2010 TATRC Grand Challenge

- Mind-machine interfaces for augmented human-computer interaction
- Body sensor networks for mobile health monitoring and augmented situation awareness

Wireless EEG/ICA Neurotechnology

with Tom Sullivan, Steve Deiss, Tzyy-Ping Jung and Scott Makeig



Integrated EEG/ICA wireless EEG recording system

- Scalable towards 1000+ channels
- Dry contact electrodes
- Wireless, lightweight
- Integrated, distributed independent component analysis (ICA)

Wireless Non-Contact Biopotential Sensors

Mike Yu Chi and Gert Cauwenberghs, 2010





- Easy access to hair-covered areas of the head without gels or slap-contact
- EEG data available only from the posterior
 - P300 (Brain-computer control, memory recognition)
 - SSVP (Brain-computer control)





EEG/ECoG/EMG Amplification, Filtering and Quantization

Mollazadeh, Murari, Cauwenberghs and Thakor (2009)



0.1

Time (s)

0.15

- Low noise
 - 21nV/√Hz input-referred noise
 - 2.0µVrms over 0.2Hz-8.2kHz
- Low power
 - 100µW per channel at 3.3V
 - Reconfigurable
 - 0.2-94Hz highpass, analog adjustable
 - 140Hz-8.2kHz lowpass, analog adjustable
 - 34dB-94dB gain, digitally selectable
- High density
 - 16 channels
 - 3.3mm X 3.3mm in 0.5μm 2P3M CMOS
 - 0.33 sq. mm per channel



Gert Cauwenberghs

0.05

-300^L

0.2

Distributed Sensing of Dopamine Activity

Murari, Stanacevic, Cauwenberghs, and Thakor (2005)

Electrochemical detection

Carbon-probe redox current



"In vitro" Dopamine monitoring by the chip using microfabricated electrode array as working electrode.

VLSI potentiostat array for distributed electrochemical sensing (*Murari, Stanacevic, Cauwenberghs, and Thakor, 2004*)





Carbon electrodes for Dopamine sensing (Murari, Rege, Paul, and Thakor, 2002)



Integrated Microfluidics Electrochemical Sensing

Naware, Rege, Genov, Stanacevic, Cauwenberghs and Thakor (ISCAS' 2004)

"In vitro" nitric oxide (NO) sensing

- emulation of the shear stress regulated NO release pathway observed in endothelial cells
- current observed by multi-channel VLSI potentiostat



Sensor Interface Conditioning Telemetry Chip Sauer, Stanacevic, Cauwenberghs, and Thakor (2005) Inductor Coil Implantable probe with electrochemical sensors, VLSI potentiostat and power **Telemetry** harvesting telemetry chip. **Potentiostat** SoS released probe body **Electrodes** Data Clock CLK Receiver Extraction VDD Power Rectification Regulation Transmitter GND Data **♦ |**♦ _{Data} Modulation Encoding Telemetry chip (1.5mm X 1.5mm) Power delivery and data transmission over the same inductive link

Cortical Surface Microvascular and Functional Imaging

with K. Murari, N. Thakor, J. Driscoll, D. Kleinfeld and T. Sejnowski



Laser speckle functional imaging of microvascular neural activity on cortical surface, through thinned skull.





Gert Cauwenberghs

Silicon and Biological Adaptive Neural Circuits

Sample

Speckle backscatter

CMOS Imaging in Awake Behaving Rats

Murari, Etienne-Cummings, Cauwenberghs, and Thakor (2010)







— 180 μm



- Minute 0 Minute 12 Minute 30 Minute 60
- First simultaneous behavioral and cortical imaging from untethered, freely-moving rats.

Integrated Systems Neuroengineering



Gert Cauwenberghs

Silicon and Biological Adaptive Neural Circuits

gert@ucsd.edu