How to outperform a supercomputer with neuromorphic chips

Kwabena Boahen
Stanford Bioengineering
boahen@stanford.edu

July 2007 Telluride

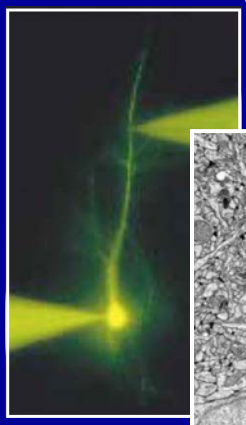
Acknowledgements

Collaborators
Bert Shi
Rajit Manohar

Support
Packard Foundation
NSF CAREER & BITS
ONR YIP & MURI
CRCNS - NIMH R01
Pioneer Award - NIH

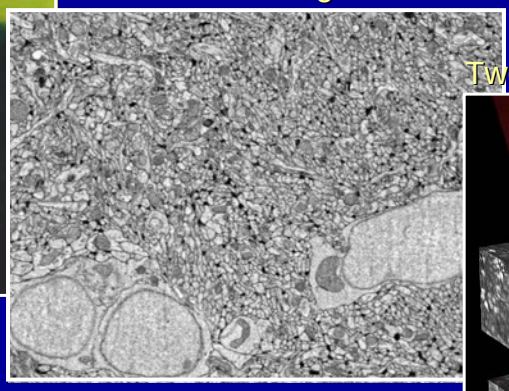
July 2007 Telluride Workshop

Dendritic recording



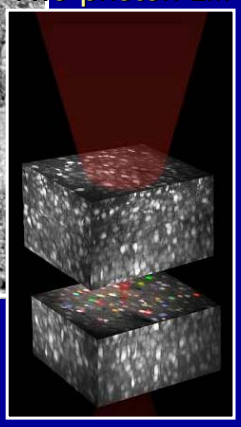
Hausser et al 1997

Serial Scanning EM



Denk et al 2005

Two-photon LM



Reid et al 2005

July 2007

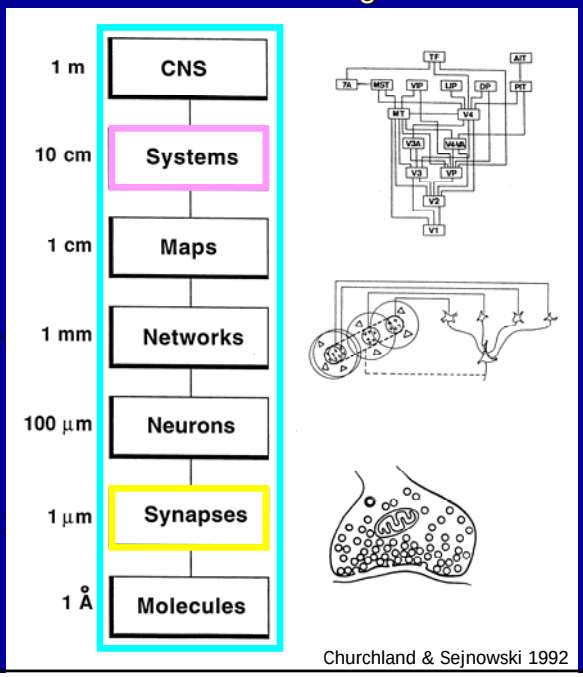
Telluride Workshop

Levels of Investigation

Analyze

Simulate

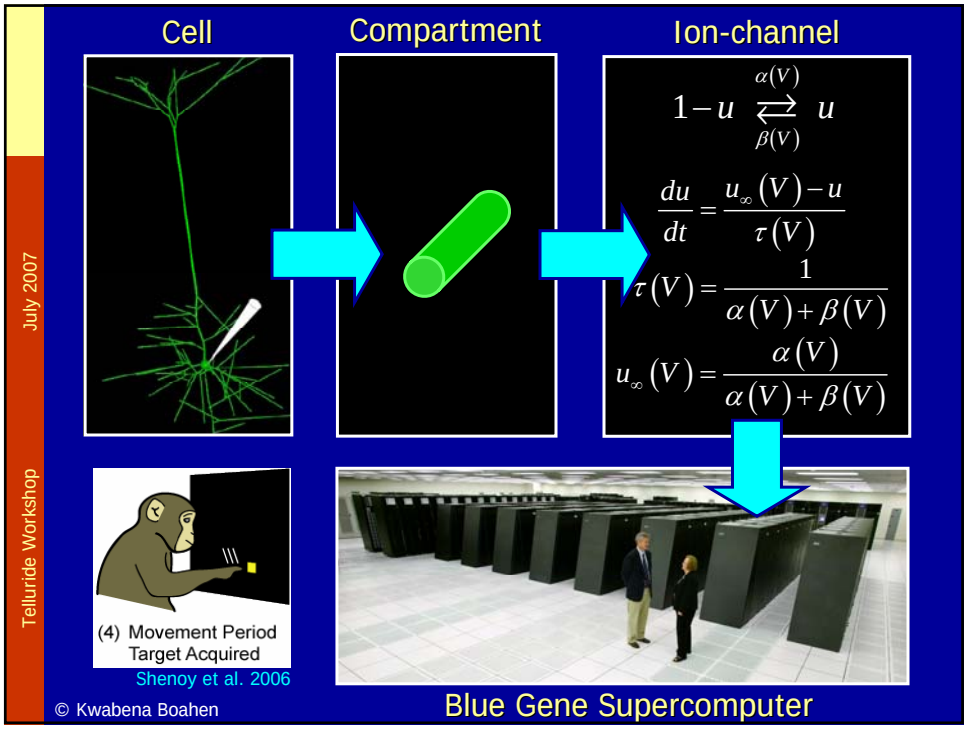
Experiment

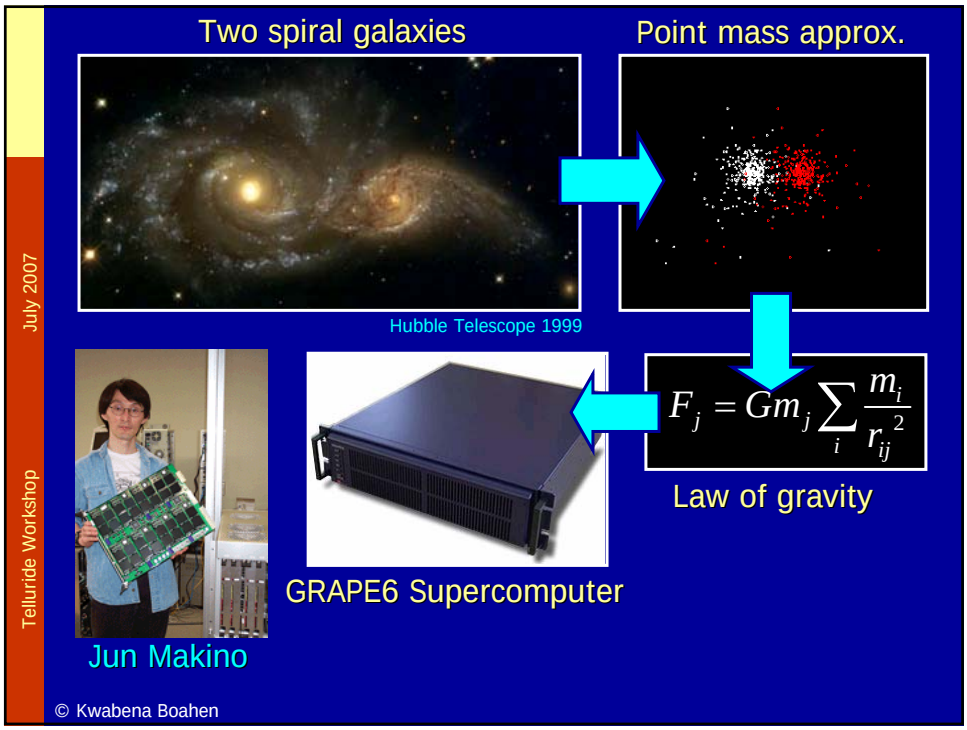


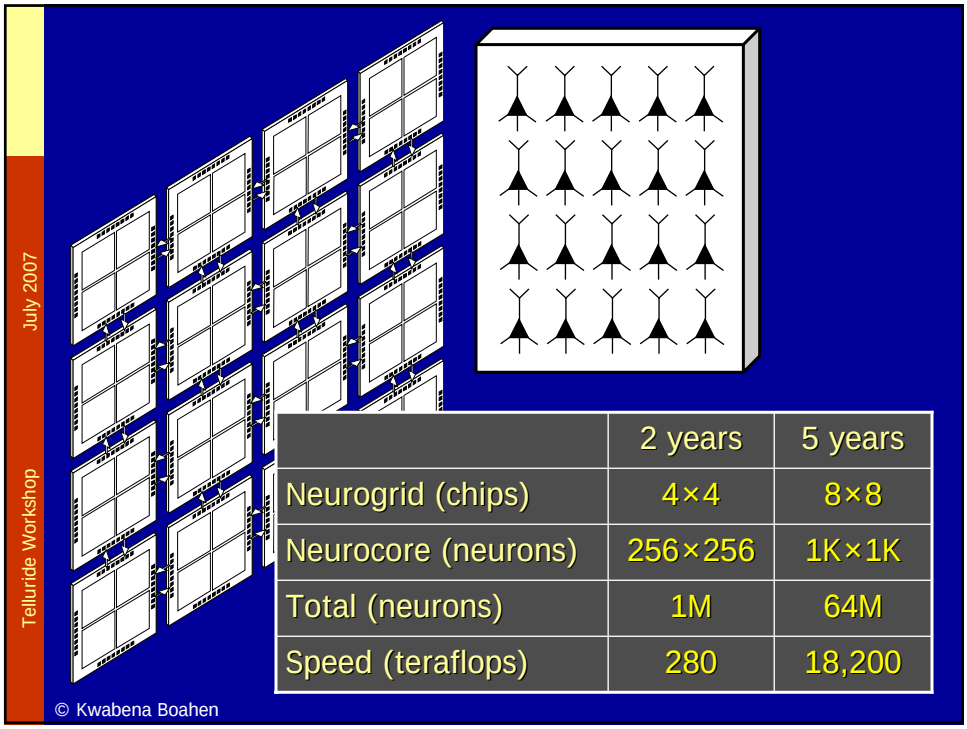
Churchland & Sejnowski 1992

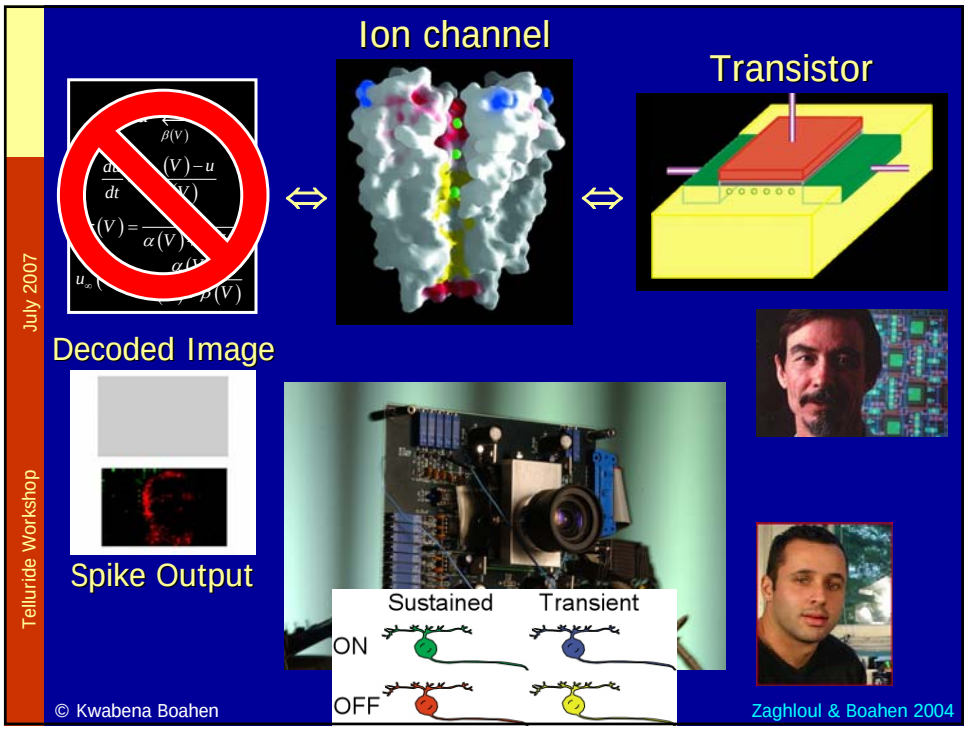
July 2007

Telluride Workshop


Cell
Compartment
Ion-channel
$1-u \underset{\beta(V)}{\overset{\alpha(V)}{\rightleftarrows}} u$
$\frac{du}{dt} = \frac{u_\infty(V)-u}{\tau(V)}$
$\tau(V) = \frac{1}{\alpha(V)+\beta(V)}$
$u_\infty(V) = \frac{\alpha(V)}{\alpha(V)+\beta(V)}$
July 2007
Telluride Workshop
(4) Movement Period
Target Acquired
Shenoy et al. 2006
© Kwabena Boahen
Blue Gene Supercomputer


Two spiral galaxies
Point mass approx.
Hubble Telescope 1999
$F_j = Gm_j \sum_i \frac{m_i}{r_{ij}^2}$
Law of gravity
July 2007
Telluride Workshop
GRAPE6 Supercomputer
Jun Makino
© Kwabena Boahen


July 2007
Telluride Workshop
2 years
5 years
Neurogrid (chips)
4×4
8×8
Neurocore (neurons)
256×256
1K×1K
Total (neurons)
1M
64M
Speed (teraflops)
280
18,200
© Kwabena Boahen


July 2007
Telluride Workshop
Ion channel
Transistor
Decoded Image
Spike Output
Sustained
Transient
ON
OFF
© Kwabena Boahen
Zaghloul & Boahen 2004

Limitations of neuromorphic chips as simulation platforms

Specialized circuits

- ❖ Neuronal properties are fixed

Hardwired connections

- ❖ Synaptic organization is fixed

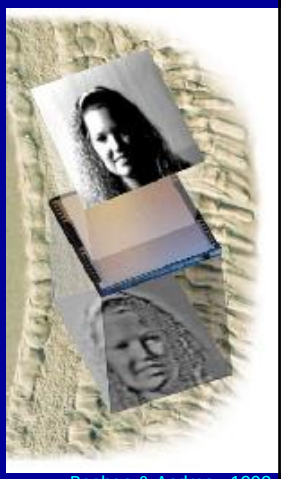
Proposed solutions

Implement Hodgkin-Huxley model

- ❖ Describes various ion-channels

Implement softwires

- ❖ User can specify connectivity

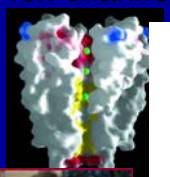


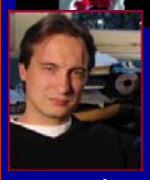
Boahen & Andreou 1992

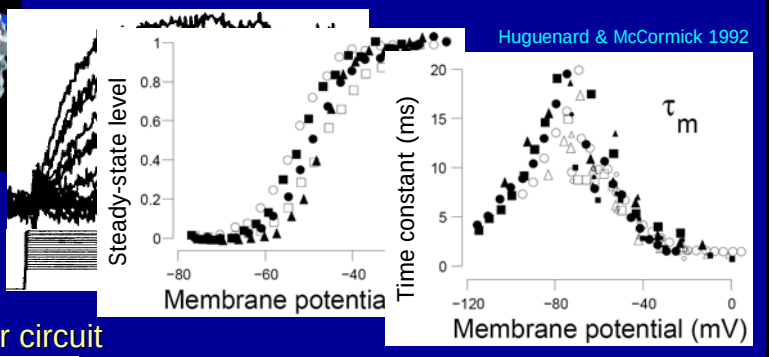
© Kwabena Boahen

July 2007 Telluride Workshop

Ion-channel population

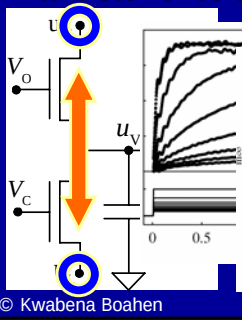


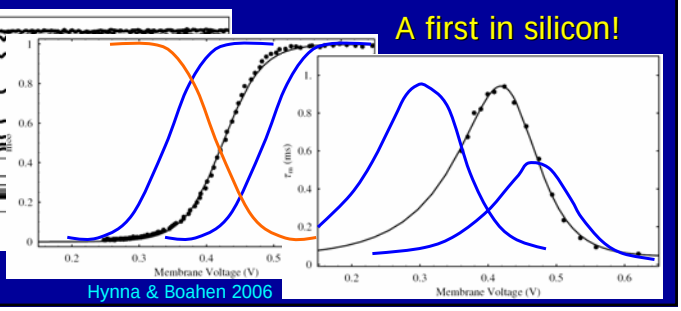




Huguenard & McCormick 1992

Transistor circuit





A first in silicon!

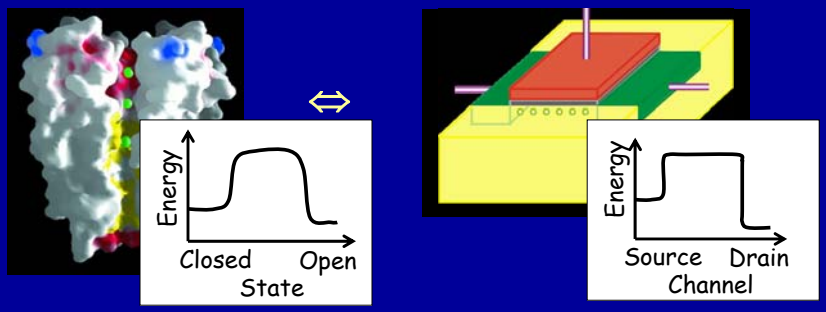
Hynna & Boahen 2006

© Kwabena Boahen

July 2007 Telluride Workshop

Telluride Workshop July 2007

Ion channel ⇔ Transistor



Transistors and ion-channels are analogs

- Gating-particles and electrons both overcome their energy barriers at exponential rates.

This observation yields a compact silicon model

- We modulate the transistor's barrier-height the same way the ion-channel's is modulated.

Telluride Workshop July 2007

First-order kinetics

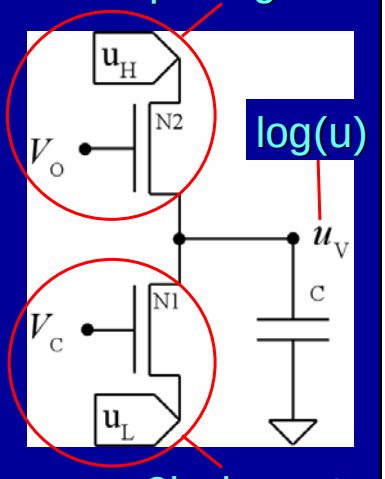
$$1-u \underset{\beta(V)}{\overset{\alpha(V)}{\rightleftarrows}} u$$

Differential equation

$$\frac{du}{dt} = \alpha(V)(1-u) - \beta(V)u$$

$$= -\frac{1}{\tau(V)}\left(u - u_\infty(V)\right)$$

Opening rate



Closing rate

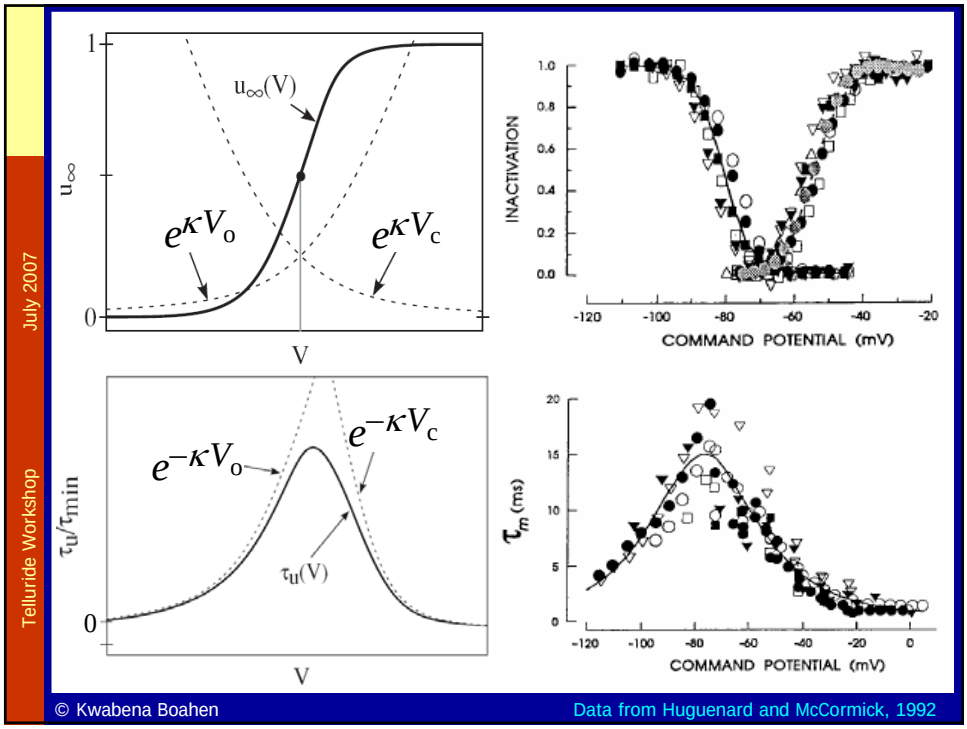
Time-constant

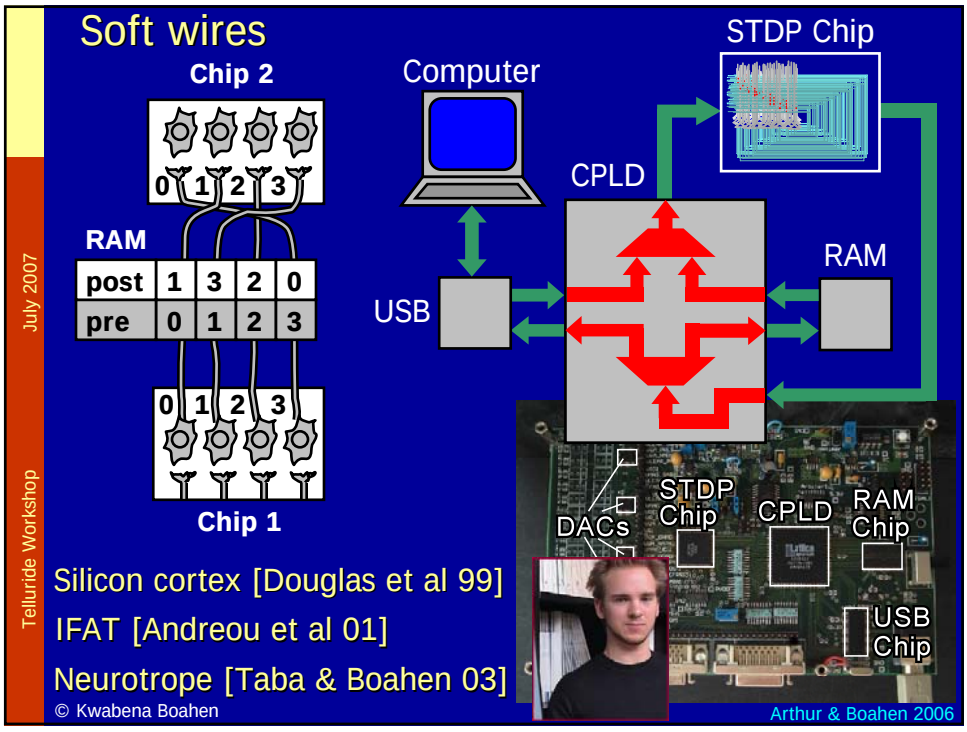
$$\tau(V) = \frac{1}{\alpha(V)+\beta(V)}$$

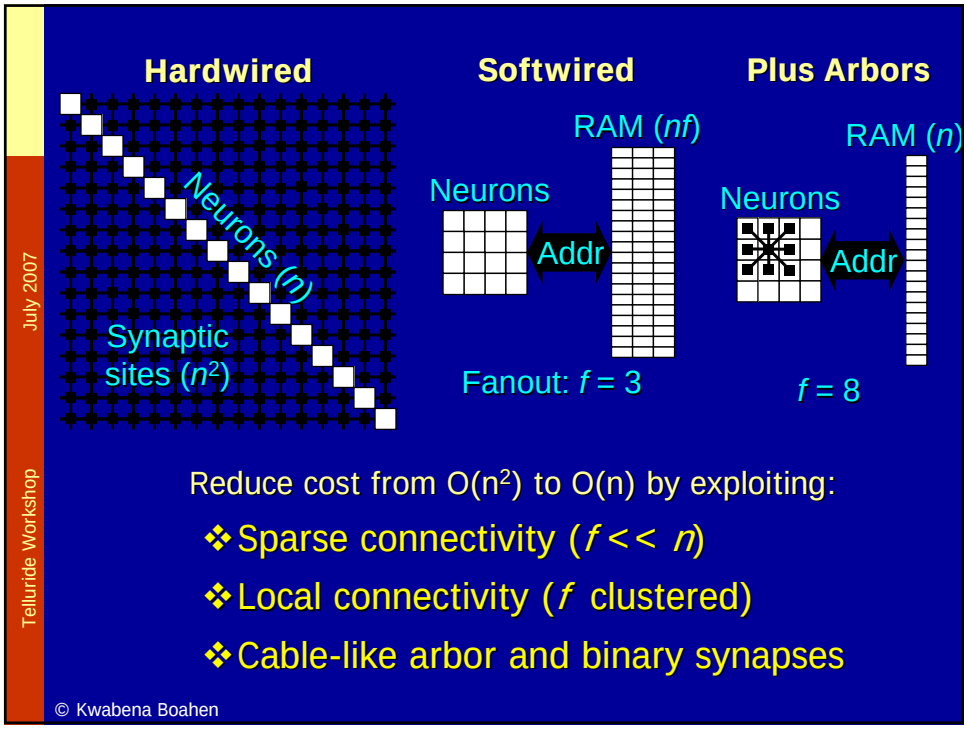
Steady-state

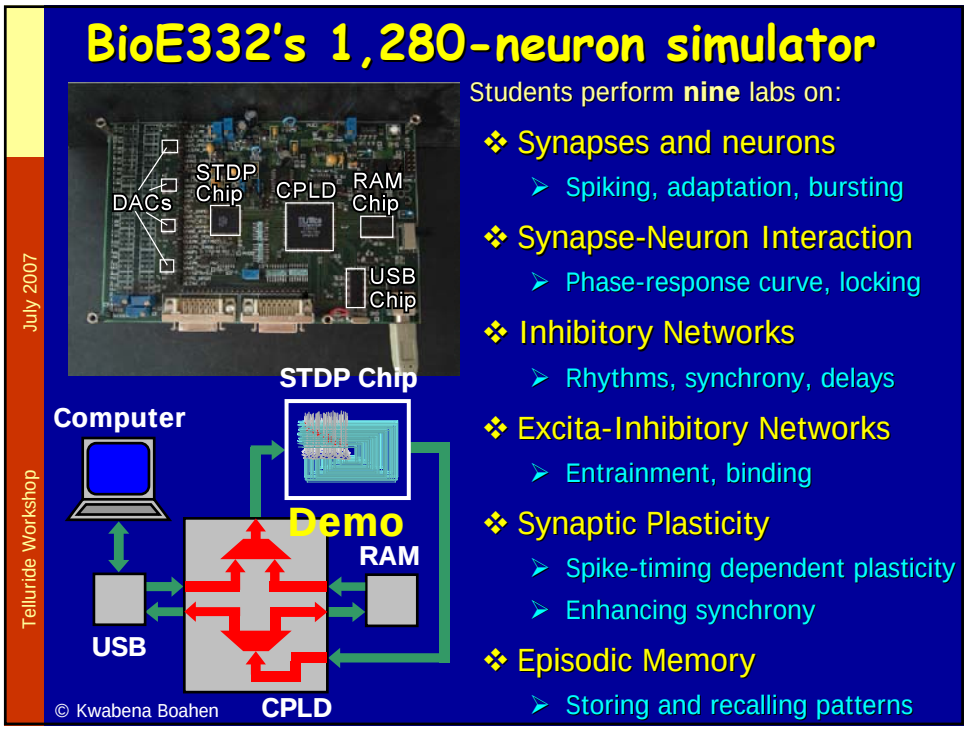
$$u_\infty(V) = \frac{\alpha(V)}{\alpha(V)+\beta(V)}$$

© Kwabena Boahen

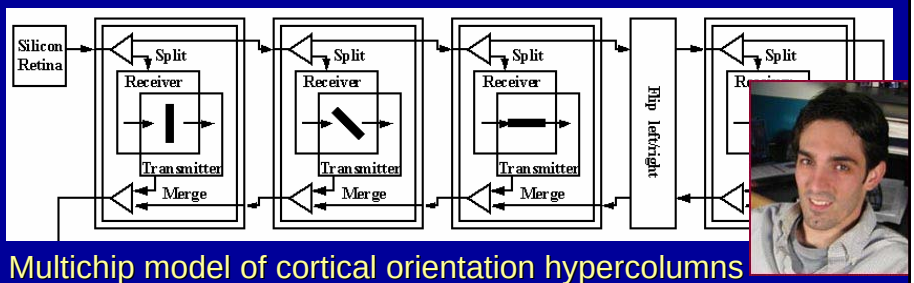

Telluride Workshop July 2007
1
0
u_∞
$u_\infty(V)$
$e^{\kappa V_o}$
$e^{\kappa V_c}$
V
INACTIVATION
COMMAND POTENTIAL (mV)
τ_u/τ_{min}
$e^{-\kappa V_o}$
$e^{-\kappa V_c}$
$\tau_u(V)$
τ_m (ms)
© Kwabena Boahen
Data from Huguenard and McCormick, 1992


Telluride Workshop July 2007
Soft wires
Chip 2
0 1 2 3
RAM
post 1 3 2 0
pre 0 1 2 3
0 1 2 3
Chip 1
Computer
USB
CPLD
STDP Chip
RAM
DACs
STDP Chip
CPLD
RAM Chip
USB Chip
Silicon cortex [Douglas et al 99]
IFAT [Andreou et al 01]
Neurotrope [Taba & Boahen 03]
© Kwabena Boahen
Arthur & Boahen 2006

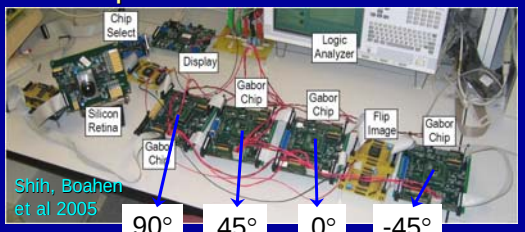

Hardwired
Softwired
Plus Arbors
Neurons (n)
Synaptic sites (n^2)
RAM (nf)
Neurons
Addr
Fanout: f = 3
RAM (n)
Neurons
Addr
f = 8
Reduce cost from O(n^2) to O(n) by exploiting:
❖ Sparse connectivity ($f << n$)
❖ Local connectivity (f clustered)
❖ Cable-like arbor and binary synapses
© Kwabena Boahen
July 2007
Telluride Workshop

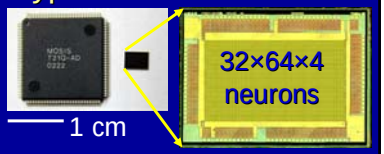

BioE332's 1,280-neuron simulator
DACs
STDP Chip
CPLD
RAM Chip
USB Chip
STDP Chip
Computer
Demo
RAM
USB
CPLD
© Kwabena Boahen
Students perform nine labs on:
❖ Synapses and neurons
➢ Spiking, adaptation, bursting
❖ Synapse-Neuron Interaction
➢ Phase-response curve, locking
❖ Inhibitory Networks
➢ Rhythms, synchrony, delays
❖ Excita-Inhibitory Networks
➢ Entrainment, binding
❖ Synaptic Plasticity
➢ Spike-timing dependent plasticity
➢ Enhancing synchrony
❖ Episodic Memory
➢ Storing and recalling patterns
July 2007
Telluride Workshop

1-D grid: 32,768 neurons on 4 chips



Multichip model of cortical orientation hypercolumns

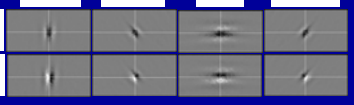




8,192 neurons
TSMC 0.25 um
9.8 mm²; 3 mW

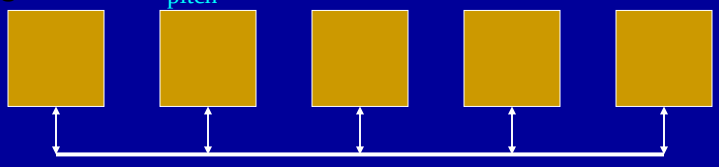
90° 45° 0° -45°

Even
Odd



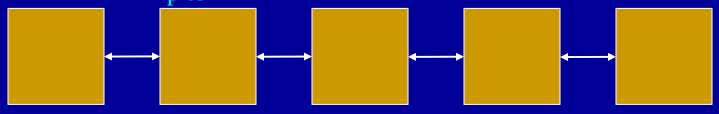
Steerable orientation
Even/odd, on/off

Bus $\leftarrow d_{\text{pitch}} \rightarrow$



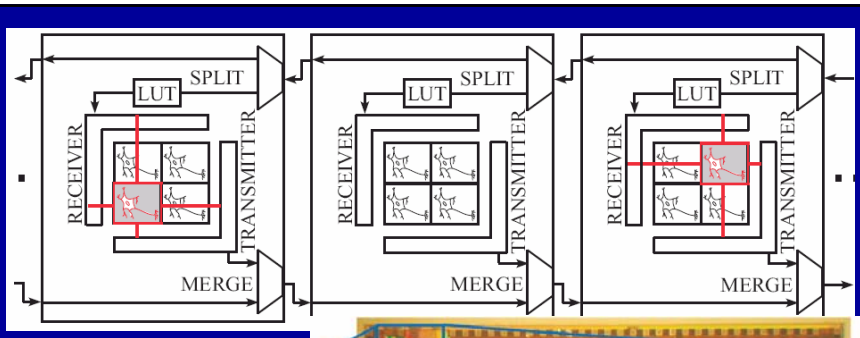
$$T_{\text{bus}} = (2\,\text{ns/inch}) d_{\text{pitch}} n_{\text{chip}}$$

Grid $\leftarrow d_{\text{pitch}} \rightarrow$



$$T_{\text{link}} = (2\,\text{ns/inch}) d_{\text{pitch}}$$

- Bus' capacity drops as more chips are added
- Grid's capacity remains the same (expandable)



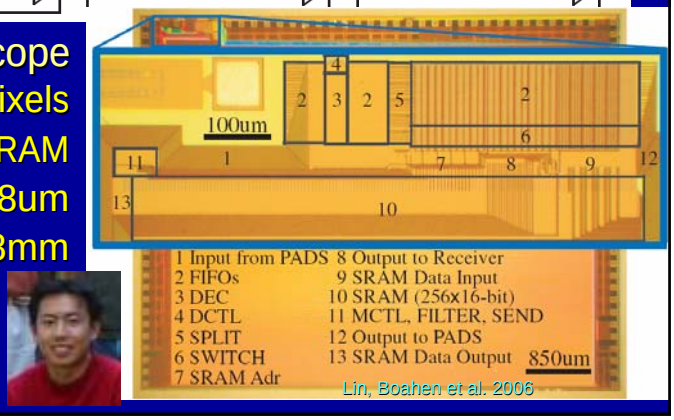
Neuroscope

320×240×4pixels

256×16-bit SRAM

TSMC 0.18um

6.2×4.8mm

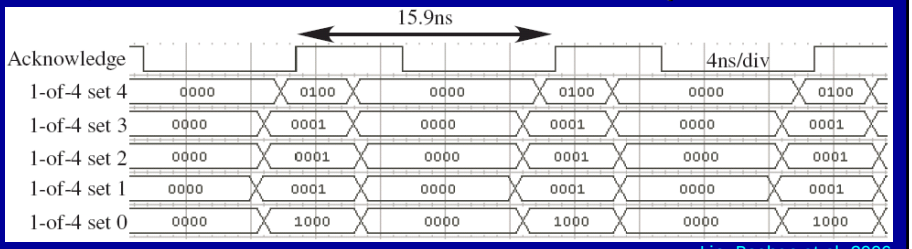


Lin, Boahen et al. 2006

July 2007 Telluride Workshop

Delay-insensitive link

63M spikes/sec!



Lin, Boahen et al. 2006

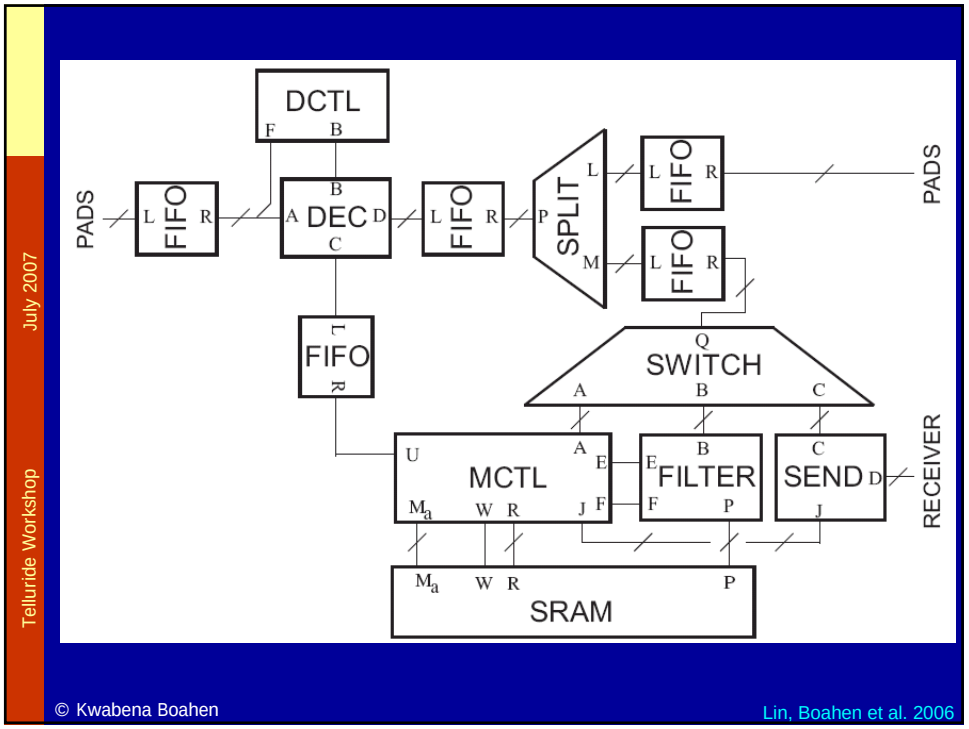
Word-serial format

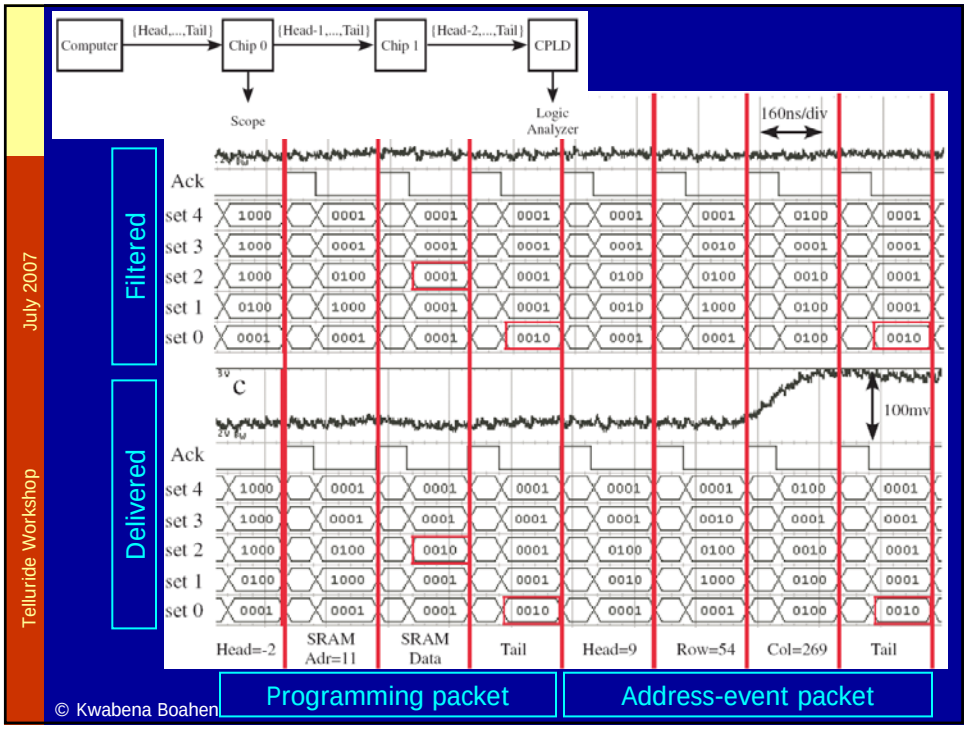
Address-event Packet

	9	8	7	6	5	4	3	2	1	0
Head	source chip-address								x	0
Row	address									0
Col	address									0
Tail	x	x	x	x	x	x	x	x	x	1

Programming Packet

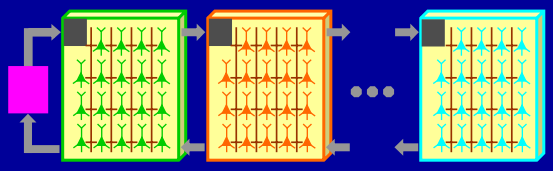
	9	8	7	6	5	4	3	2	1	0
Head	target chip-address								x	0
SRAM address	source chip-address								x	0
SRAM data	x	x	x	x	x	K	AP		x	0
Tail	x	x	x	x	x	x	x	x	x	1


July 2007
Telluride Workshop
DCTL
F B
B
A DEC D
C
PADS
L FIFO R
L FIFO R
P SPLIT L
M
L FIFO R
L FIFO R
PADS
L FIFO R
Q
SWITCH
A B C
U
A
MCTL
E
Ma W R J F
E
B
FILTER
F P
C
SEND D
J
RECEIVER
Ma W R
P
SRAM
© Kwabena Boahen
Lin, Boahen et al. 2006


July 2007
Telluride Workshop
Computer
{Head,...,Tail}
Chip 0
{Head-1,...,Tail}
Chip 1
{Head-2,...,Tail}
CPLD
Scope
Logic Analyzer
160ns/div
Filtered
Delivered
Ack
set 4
set 3
set 2
set 1
set 0
C
100mv
Head=-2
SRAM Adr=11
SRAM Data
Tail
Head=9
Row=54
Col=269
Tail
Programming packet
Address-event packet
© Kwabena Boahen

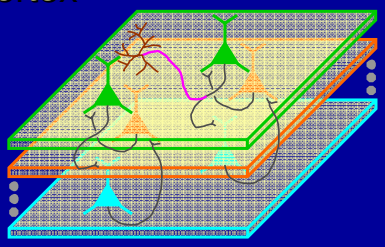
Telluride Workshop July 2007

Neurogrid



RAM RAM Resistors

Cortex



Vertical Horizontal Arbor

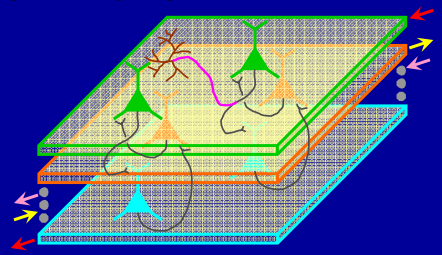
© Kwabena Boahen

Connecting 6000 synapses

- **On-chip RAM:** Vertical projections (f = 6)
- **Off-chip RAM**: Horizontal projections (f = 10)
- **Resistive mesh**: Adjustable electro-tonic spread (f ~100)

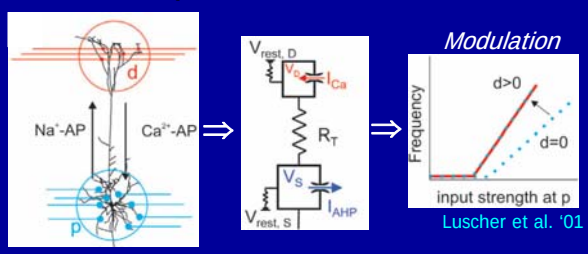
Telluride Workshop July 2007

Layered projections



Forward Backward Lateral

Two-compartment abstraction

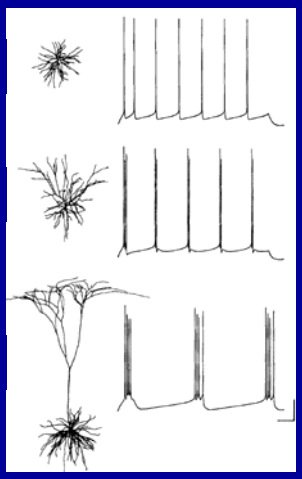


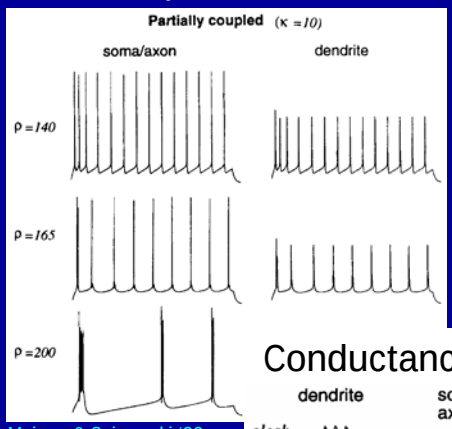
Luscher et al. '01

© Kwabena Boahen

- Projections target apical or basal dendrites
  - Terminate in distinct cortical layers
- Apical input modulates pyramidal cell's response to basal input
  - Back-propagating spikes recruit Ca current if dendrite is depolarized

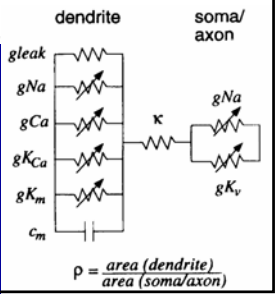
Multi- versus two-compartment





Mainen & Sejnowski '96

Conductances

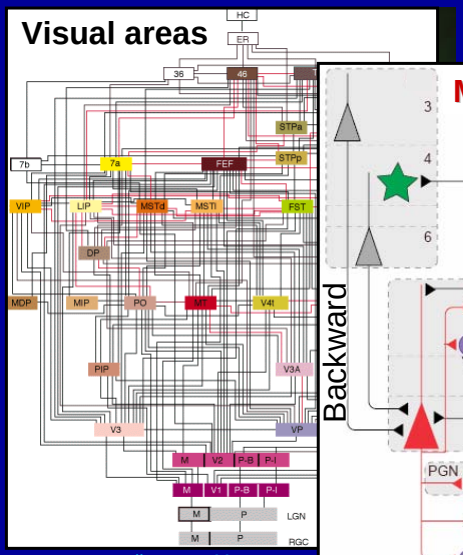


- Same ion-channel types & densities
- Only morphology differed
  - Two-compartment captured this

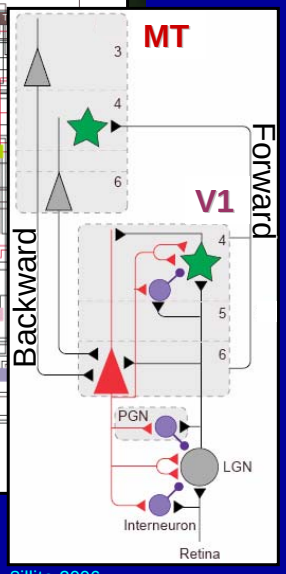
© Kwabena Boahen

July 2007 Telluride Workshop

Visual areas



Van Essen & Fellerman 1991



Sillito 2006

© Kwabena Boahen

July 2007 Telluride Workshop

Summary: We have developed three enabling technologies for Neurogrid

Analog VLSI for real-time simulation

- Models ion-channel populations

Digital VLSI for programmability

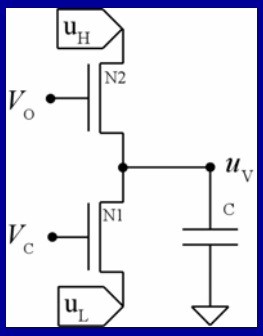
- Specifies synaptic connections

Grid network for expandability

- Relays spikes from chip to chip

Putting supercomputers on neuroscientists' desks with real-time cortex-scale simulation is feasible in 5yrs.

Circuit analysis



Assuming N1 is in saturation:

$$C\frac{du_V}{dt} = I_{ds0}\,e^{\kappa V_O}\left(e^{-u_V} - e^{-u_H}\right) - I_{ds0}\,e^{\kappa V_C}e^{-u_L}$$

$$u = e^{u_V - u_H}$$

$$\frac{du}{dt} = \frac{I_{ds0}}{C\,U_T}e^{\kappa V_O - u_H}\left(1-u\right) - \frac{I_{ds0}}{C\,U_T}e^{\kappa V_C - u_L}\,u$$

$$\Rightarrow \alpha\left(V_O\right) = \frac{I_{ds0}}{C\,U_T}e^{\kappa V_O - u_H} \text{ and } \beta\left(V_C\right) = \frac{I_{ds0}}{C\,U_T}e^{\kappa V_C - u_L}$$

Rates are exponentially dependent if V_O and V_C vary linearly with V_{mem}

- Slopes should be complementary
  - V_O increases while V_C decreases for activation variable
  - The reverse is true for an inactivation variable