Digital Electronics:
Fundamental Limits and Future Prospects

Konstantin K. Likharev
Stony Brook University

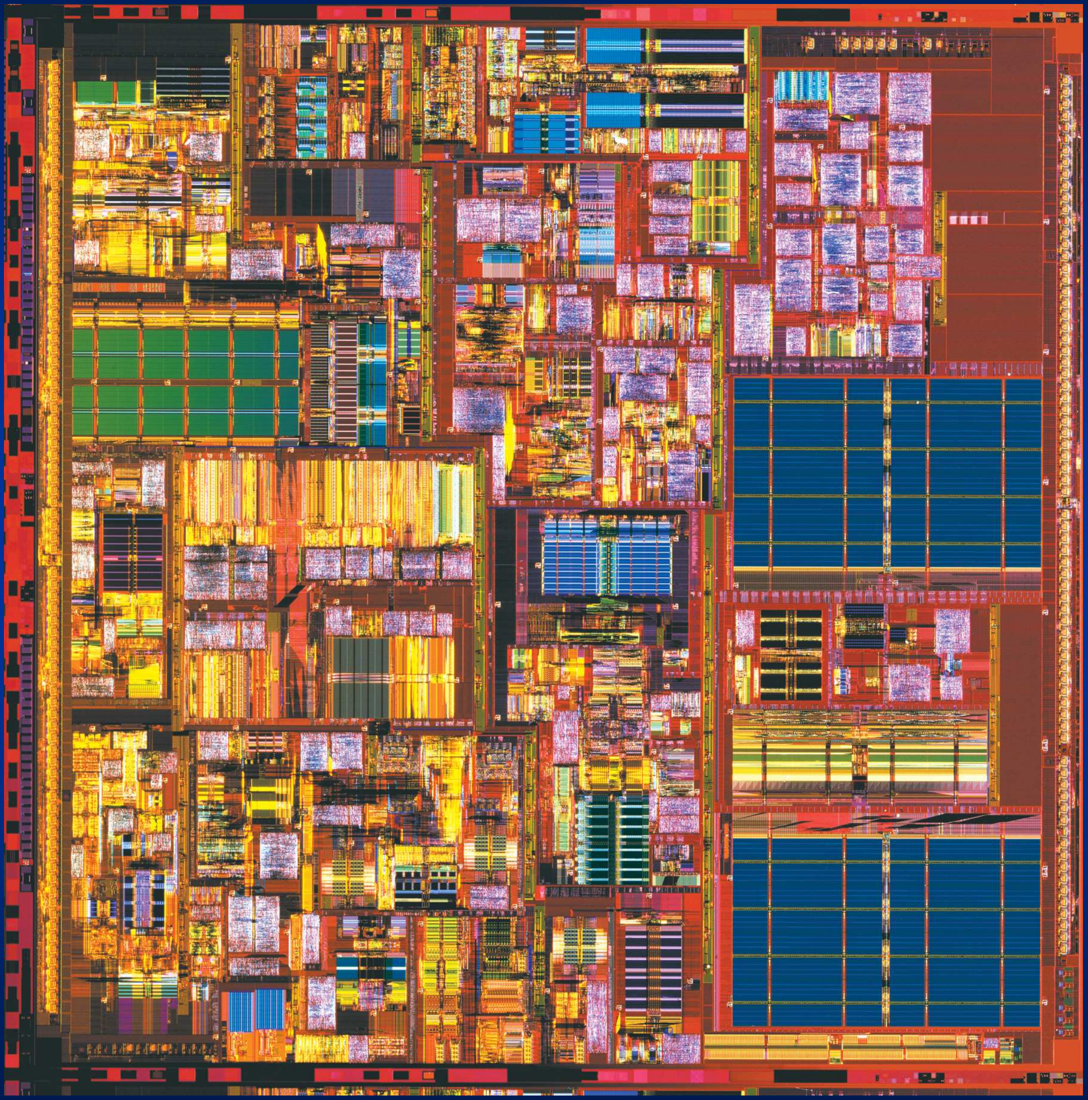
Unpublished Results:
W. Chen, E. Cimpoiasu, J. Lee, X. Liu, J. Lukens
X. Ma, A. Mayr, Ö. Türel

Discussions:
P. Adams, D. Antoniadis, J. Barhen, D. Frank, R. Landauer,
M. Lundstrom, V. Protopopescu, T. Sejnowski, P. Solomon, S. Tiwari

Support: DOE, NSF, SRC

General Reference: KKL, "Electronics below 10 nm", in: J. Greer *et al*., eds. *Nano and Giga Challenges in Microelectronics* (Elsevier, 2003), pp. 27-68
(available at http://rsfq1.physics.sunysb.edu/~likharev/nano/NanoGiga.pdf)

CMOS TECHNOLOGY



130-nm Pentium 4 ("Northwood") processor:

• 42 million transistors

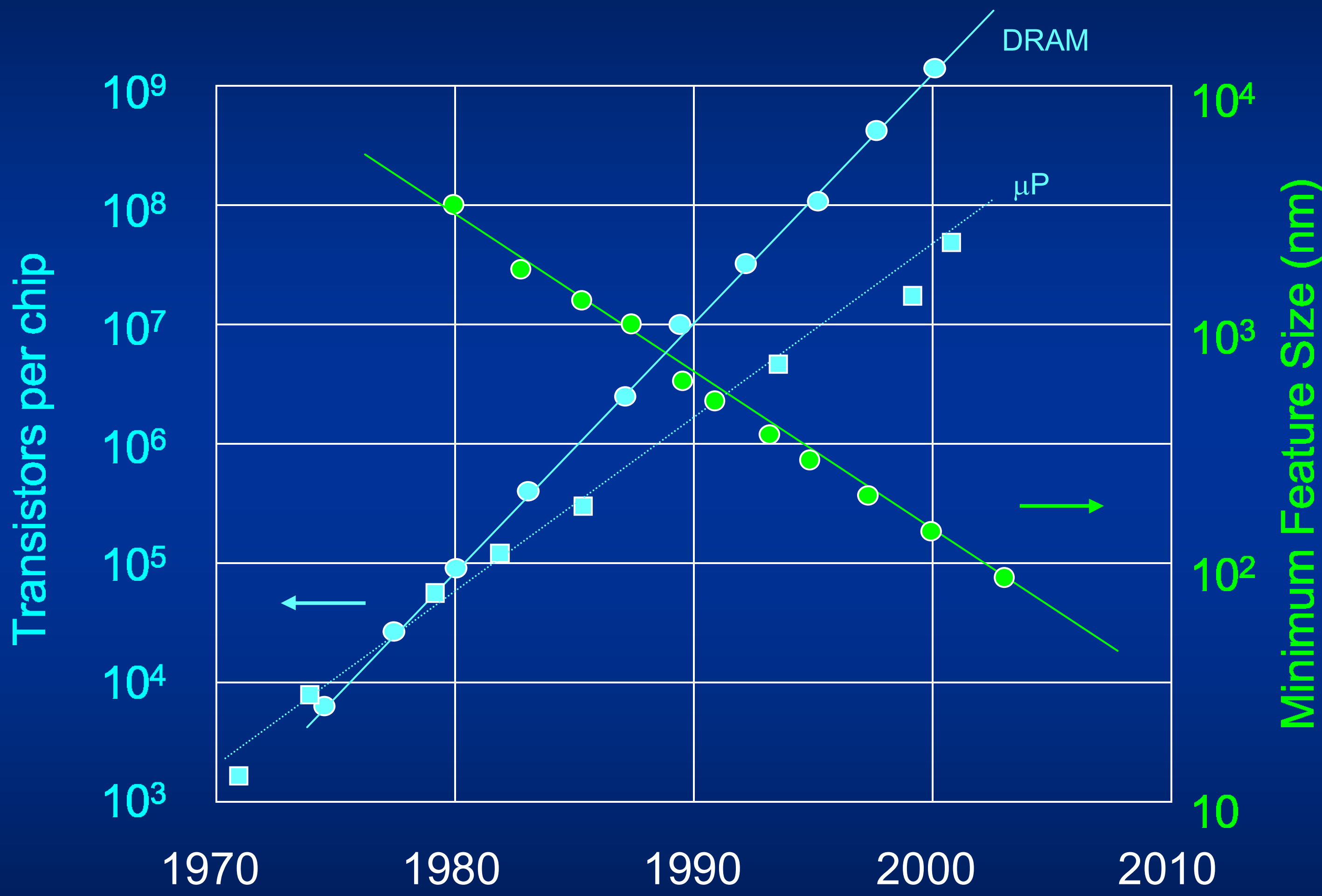
• > 3 GHz clock frequency

DRAM memories:

4 Gb chips demonstrated

(~ 10^9 transistors/cm^2)

MOORE'S LAW


DRAM
μP
Transistors per chip
10^9
10^8
10^7
10^6
10^5
10^4
10^3
Minimum Feature Size (nm)
10^4
10^3
10^2
10
1970
1980
1990
2000
2010

CONCERN # 1: POWER

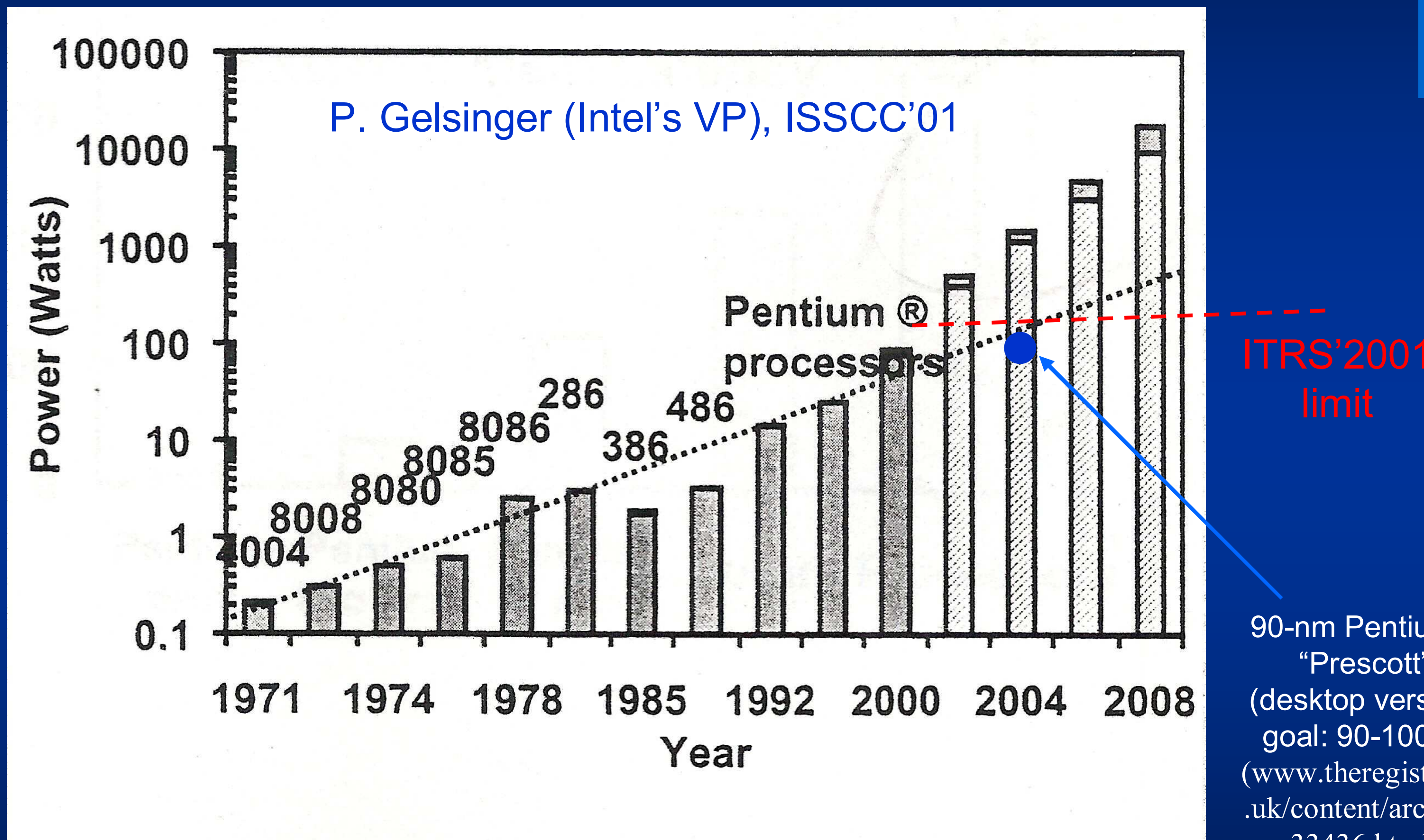


Figure 1.3.3: Lead microprocessor power increases dramatically beyond expected trend.

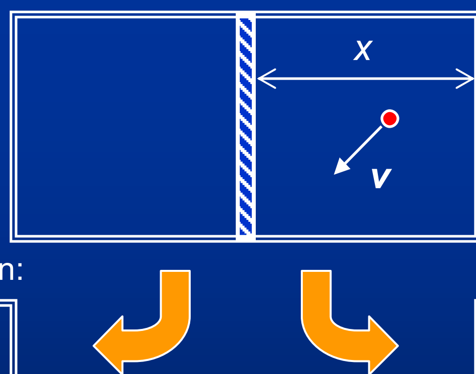
90-nm Pentium 4 "Prescott" (desktop version) goal: 90-100 W (www.theregister.co.uk/content/archive/33436.html)

FUNDAMENTAL LIMITS ON POWER CONSUMPTION?

Thermodynamic (Maxwell-demon) "limit": $E > k_B T \ln 2$

J. Maxwell, *Theory of Heat* (Green and Co., London, 1875)
L. Boltzmann, Wiener Berichte **76**, 373 (1877)
L. Szillard, Z. f. Physik **53**, 840 (1929)
L. Brillouin, *Science and Information Theory* (Acad. Press, New York, 1956)
R. Landauer, IBM J. Rev. Devel. **5**, 183 (1961)

Simple mechanical model:

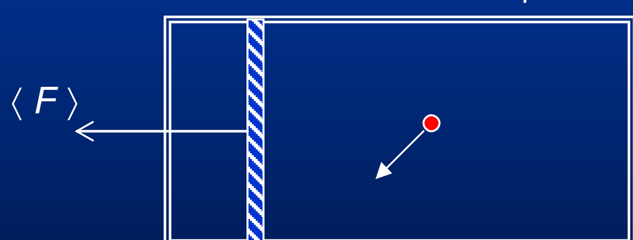


$mv^2/2 = (3/2)k_BT$, $\quad mv_x^2/2 = (1/2)k_BT$

$\Delta p = 2mv_x$, $\quad f = v_x/2x$

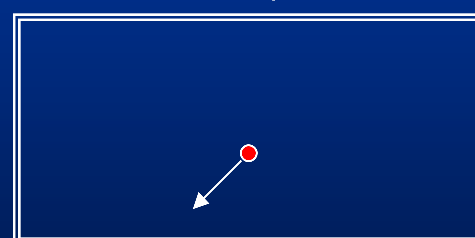
$\langle F \rangle = \Delta p \times f = mv_x^2/x = k_BT/x$

Reversible isothermal expansion:



$W = \int \langle F \rangle\, dx = k_BT \int_{x0}^{2x0} dx/x = k_BT \ln 2$

Irreversible expansion:



$W = 0$

REVERSIBLE COMPUTING

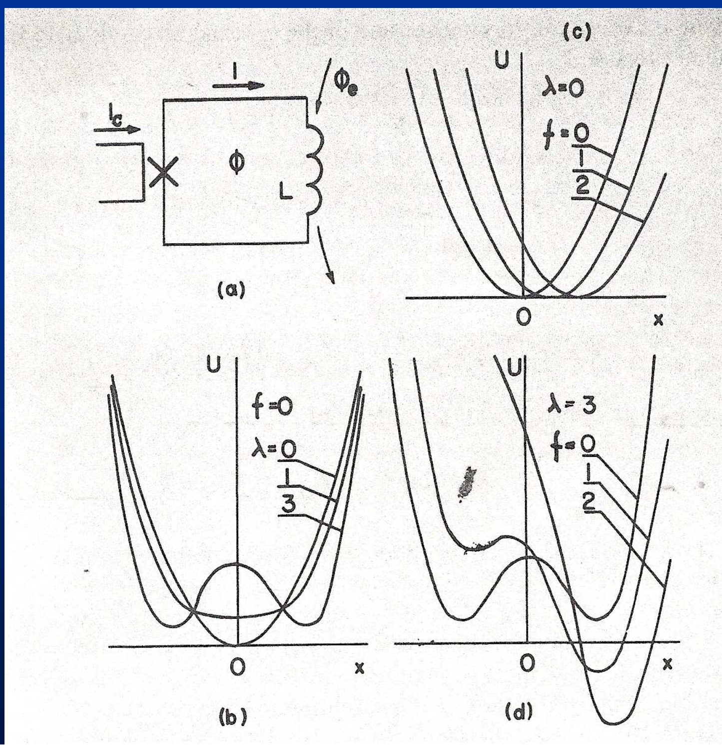
Basic Idea: C. Bennett, IBM J. Res. Devel. **17**, 525 (1973)

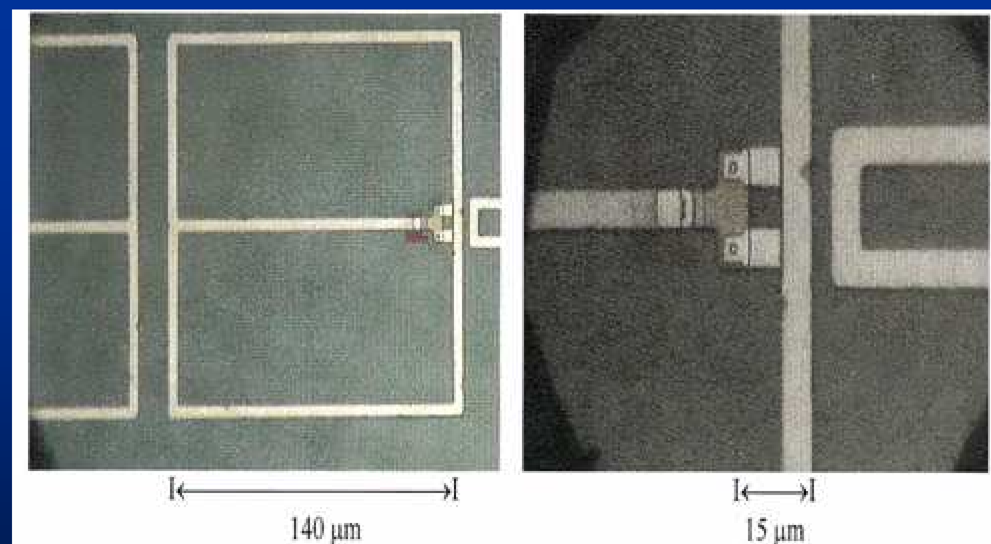
Requirements:

- physically reversible devices
- informationally reversible architecture

Device Example: Parametric Quantron ("Flux Quantum Parametron")

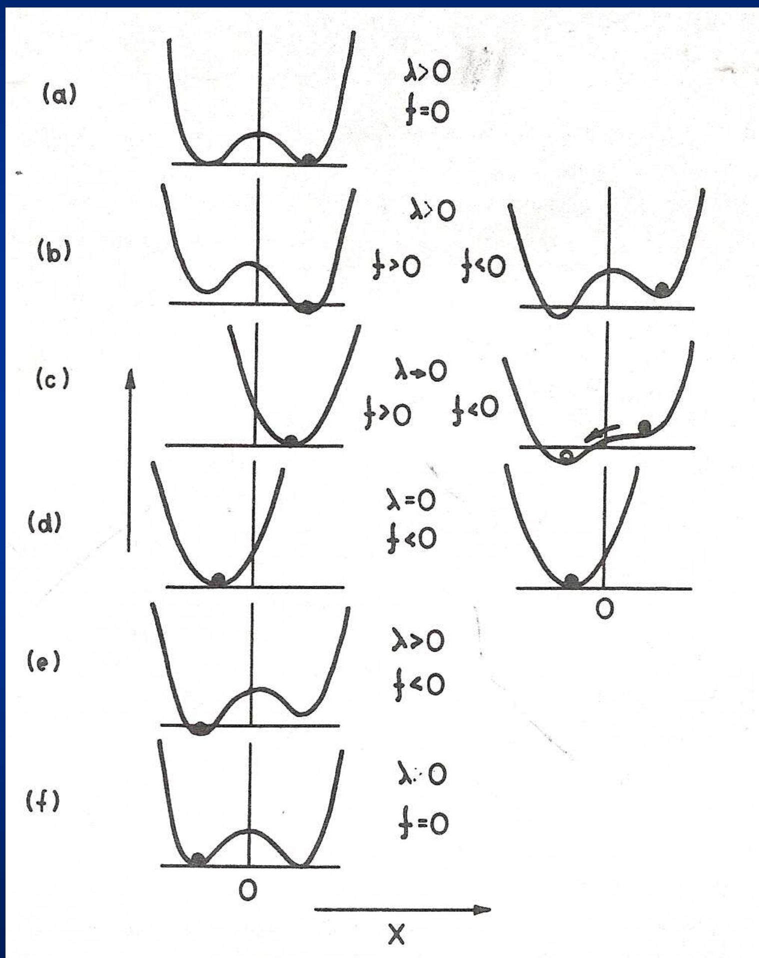
KKL, IEEE Trans. Magn. **15**, 240 (1977)

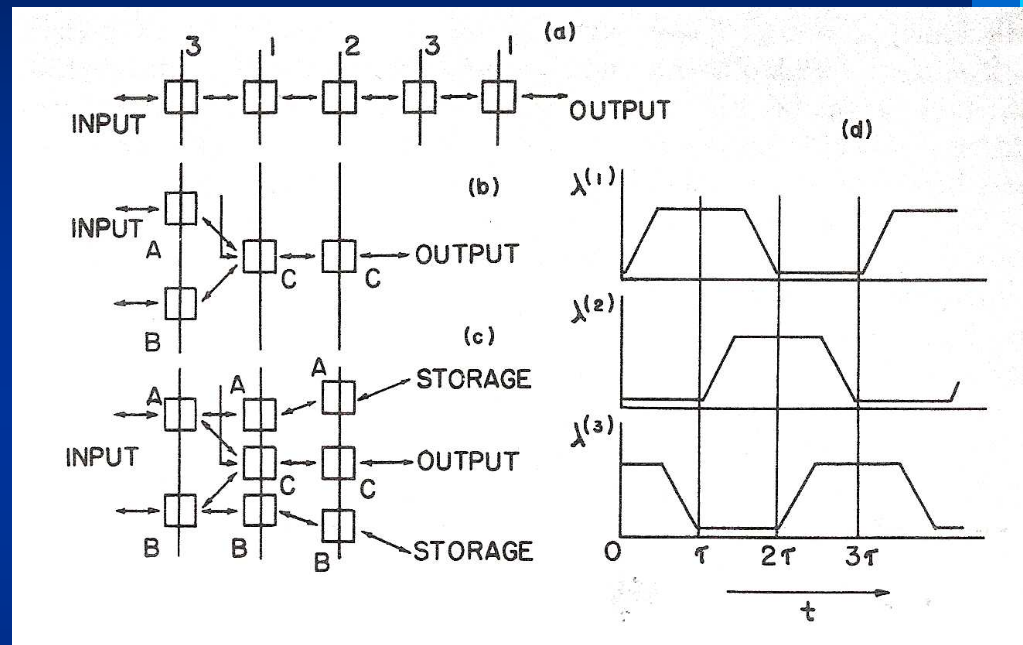




Picture courtesy: J. Lukens (SBU)

REVERSIBLE COMPUTING WITH PARAMETRIC QUANTRONS





C. Bennett and R. Landauer, private communication (1976)

KKL, Int. J. Theor. Phys. **21**, 311 (1982)

PARAMETERIC QUANTRON: ANALYSIS RESULTS

$$E > \left\{ \begin{array}{c} \dfrac{k_B T}{\omega_c \tau} \\ \\ \dfrac{\hbar\omega}{\omega_c \tau} \end{array} \right\} \times \ln \frac{1}{p(\omega\tau)}, \qquad \text{for } \omega_c\tau \text{ ◷ } 1,\ p \text{ ◷ } 1,\ \omega \text{ ⮪ } \omega_c$$

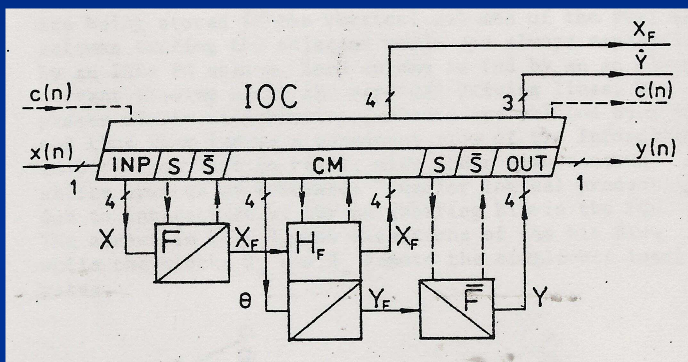
Thus, E may be much less than both $k_B T$ and $\hbar/\tau$,
i.e. not only the “thermodynamic limit”,
but also the “quantum-mechanical limit” may be also overcome

General note: $\Delta E \times \Delta t$ ◷ $\hbar$ is much less general than $\Delta x \times \Delta p > \hbar/2$
(e.g., QND measurements)

IS REVERSIBLE COMPUTATION PRACTICABLE?

Answer: Yes and Not

e.g., KKL, S. V. Rylov, and V. K. Semenov, IEEE Trans. Magn. **21**, 947 (1985)



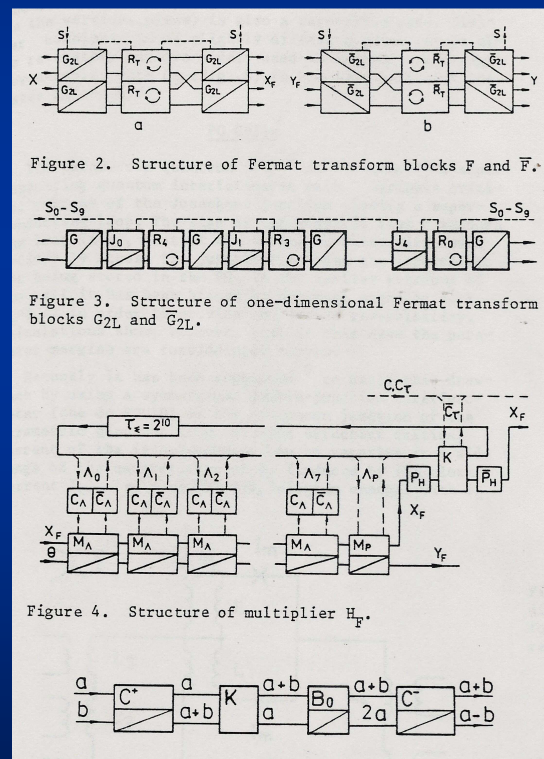


Figure 2. Structure of Fermat transform blocks F and F̄.

Figure 3. Structure of one-dimensional Fermat transform blocks G_{2L} and $\bar{G}_{2L}$.

Figure 4. Structure of multiplier H_F.

Figure 5. Structure of the basic block G (the Fermat number transformer).

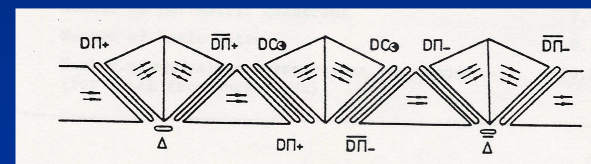


Figure 6. Structure of the Fermat adder C^+.

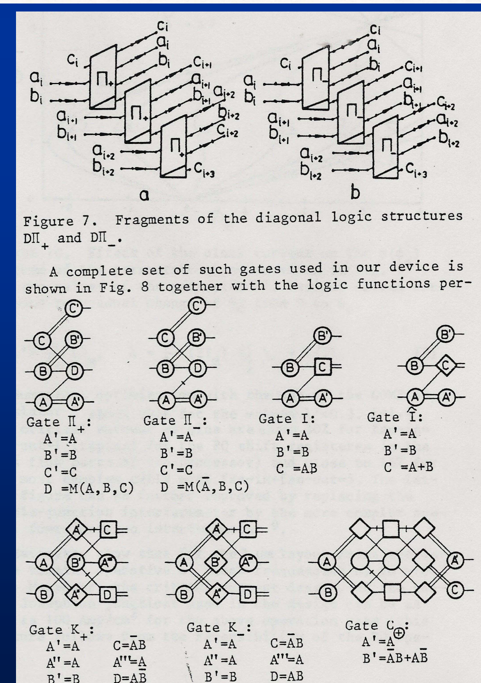


Figure 7. Fragments of the diagonal logic structures $D\Pi_+$ and $D\Pi_-$.

A complete set of such gates used in our device is shown in Fig. 8 together with the logic functions per-

Figure 8. Complete set of single-bit logic gates and the functions they perform when bits are flowing from left to right. M denotes the 2/3 majority function.

Fast convolver

$$y(n) = \Sigma_k x(n) \times h(n-k)$$

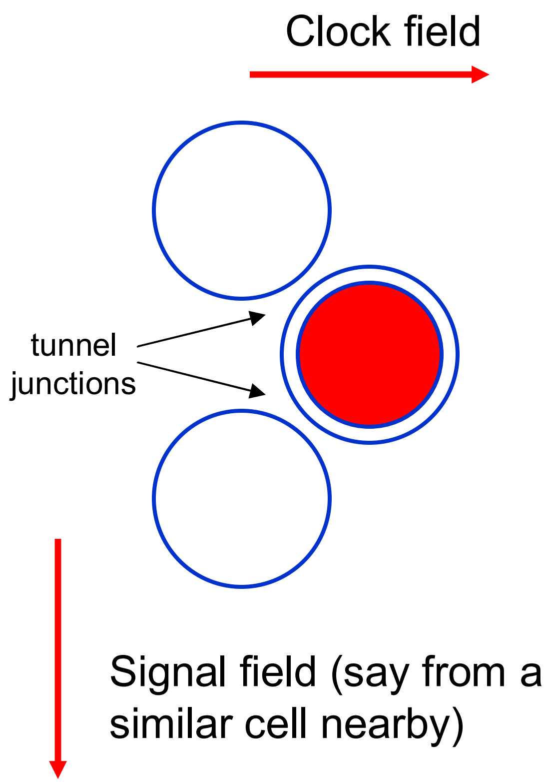
Bottom line for 8 bits, 1024 points:

- 30 nW @ 1 GHz & 4.2 K
- 9.2×10^6 PQs

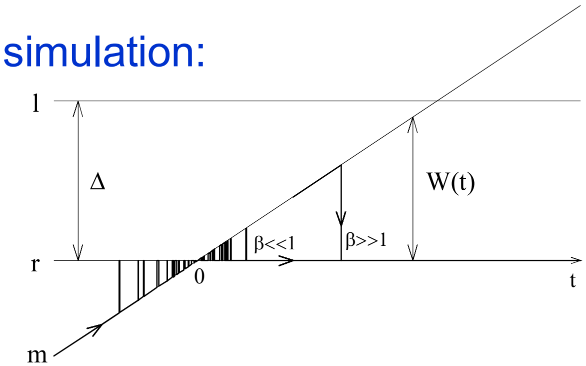
(too many, but may be dramatically reduced at partial Irreversibility)

SET PARAMETRON

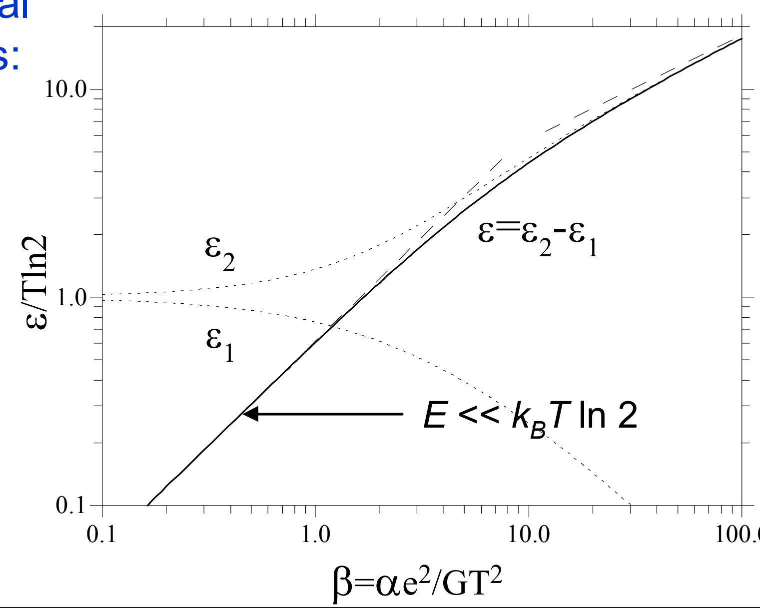
KKL and A. N. Korotkov, Science **273**, 763 (1996)



Monte Carlo simulation:



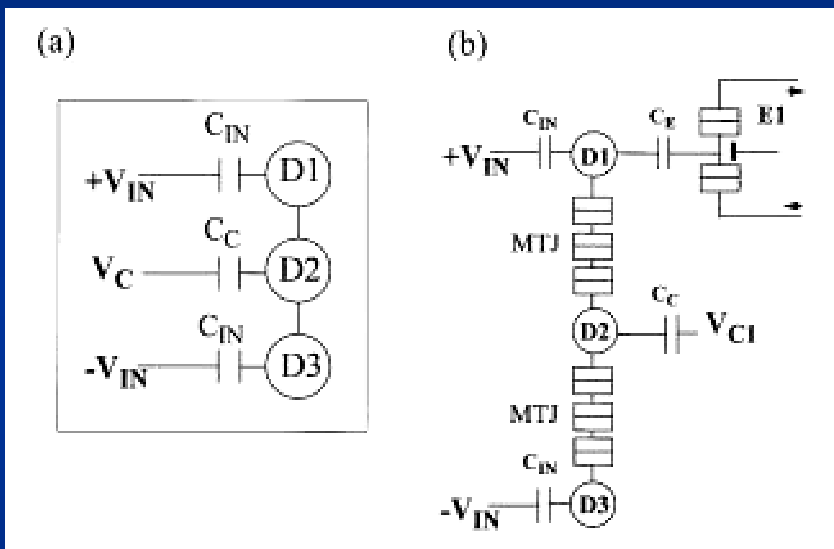
Statistical analysis:



Different from the Parametric Quantron:
 (i) Discrete state variable
 (ii) Statistical rather than dynamic system

SET PARAMETRON: EXPERIMENTAL IMPLEMENTATION

A. O. Orlov *et al*., Appl. Phys. Lett. **78**, 1625 (2001)
E. G. Emiroglu *et al*., J. Vac. Sci. Technol. B **20,** 2806 (2002).



"clocked QCA [Quantum-Dot Cellular Automata] half cell"

QCA species:

(i) "Ground-state computing"

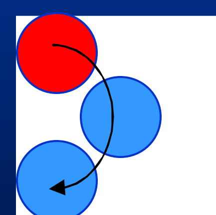
C. Lent *et al*., Nanotechnology **4**, 49 (1993)

(cannot work, now abandoned)

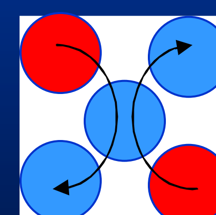
(ii) "Clocked" QCA

C. Lent and P. Tougaw, Proc. IEEE **85**, 541 (1997)

(a version of the SET parametron)

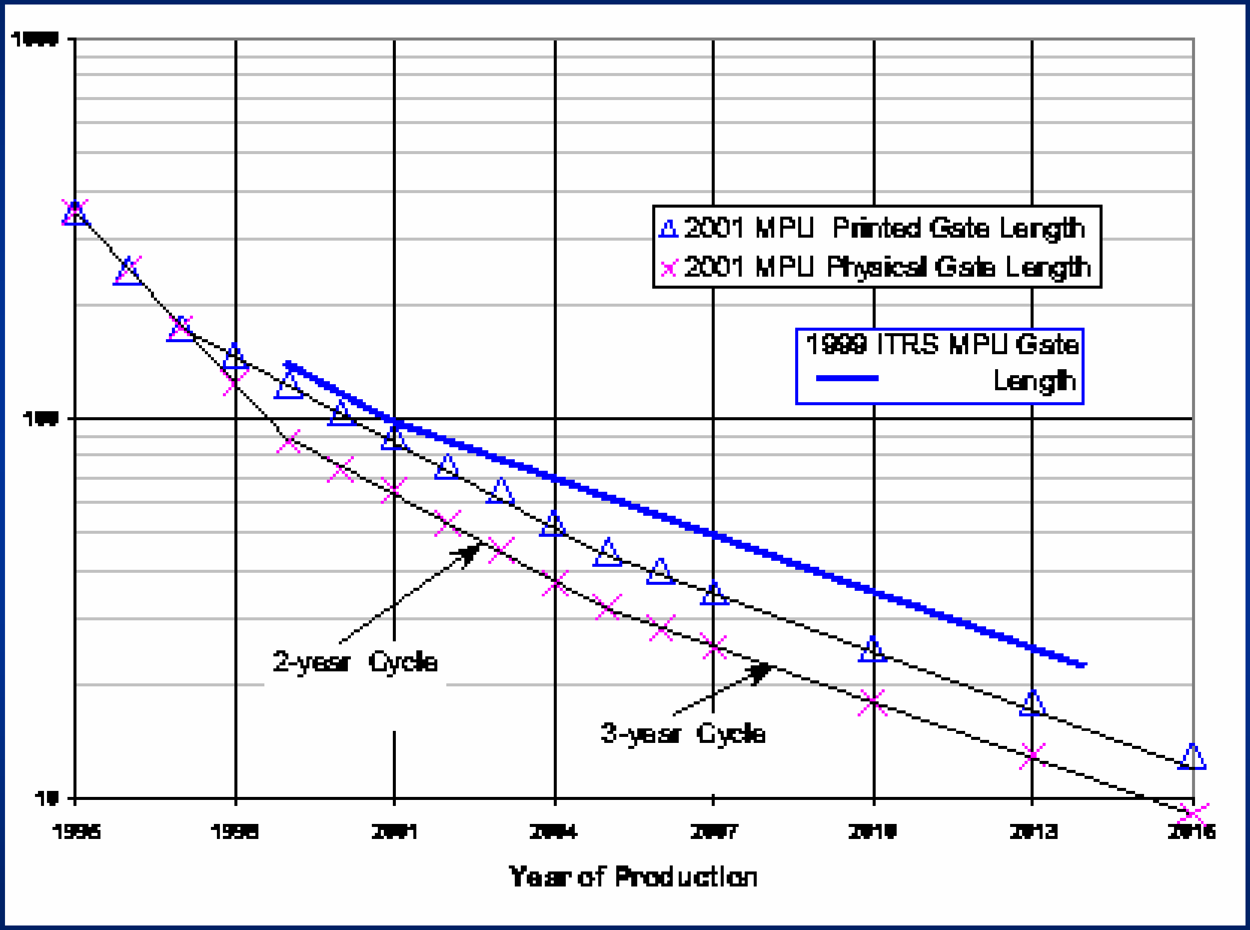


Stony Brook version



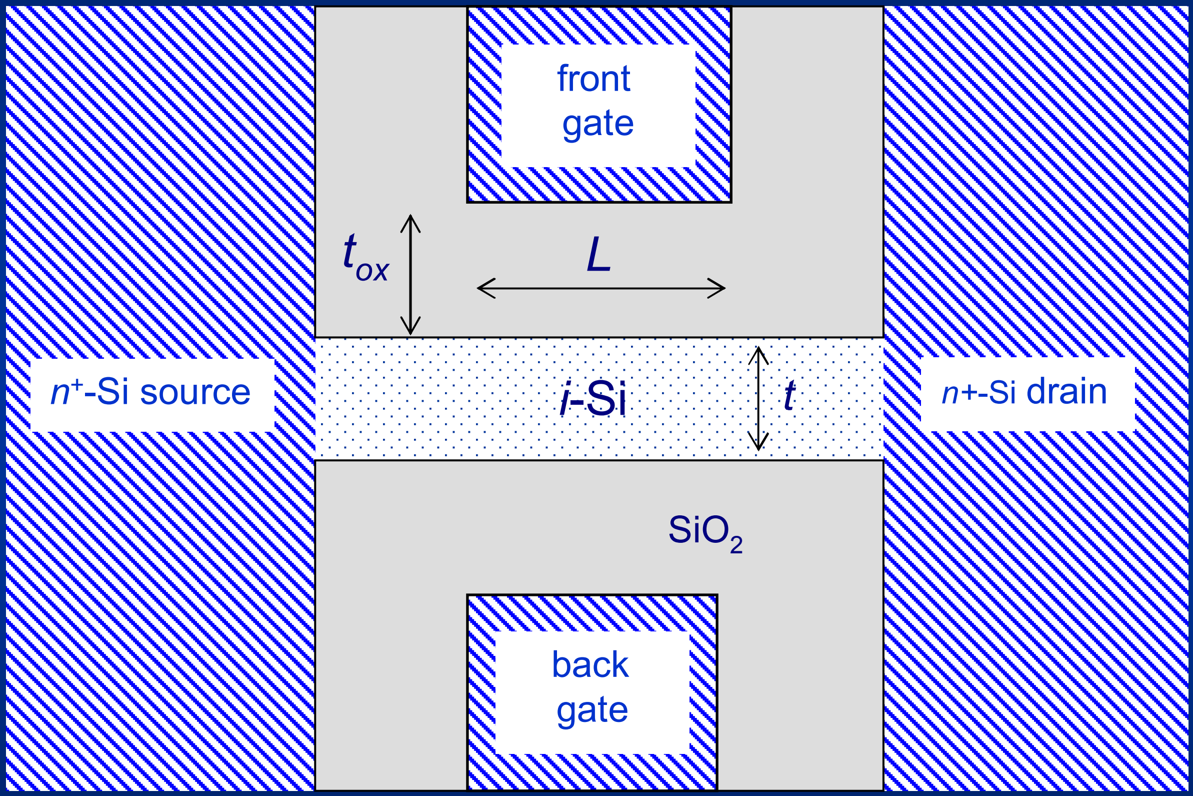
Notre Dame version

CONCERN #2: SIZE/FABRICATION



INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS
2001 EDITION

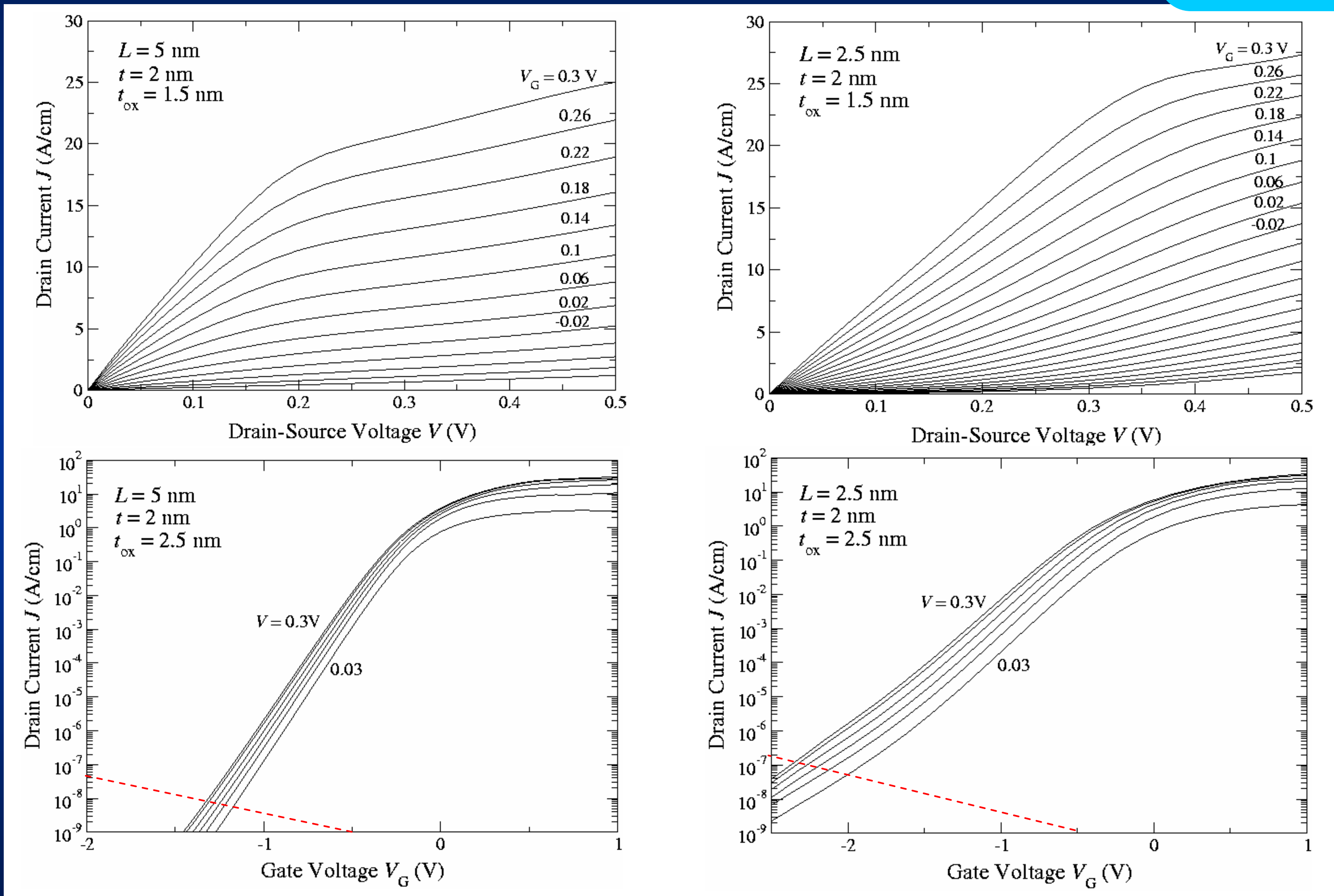
DOUBLE-GATE MOSFETs: A SIMPLE MODEL



Pseudo-Hartree approach: Schrödinger + Poisson equations

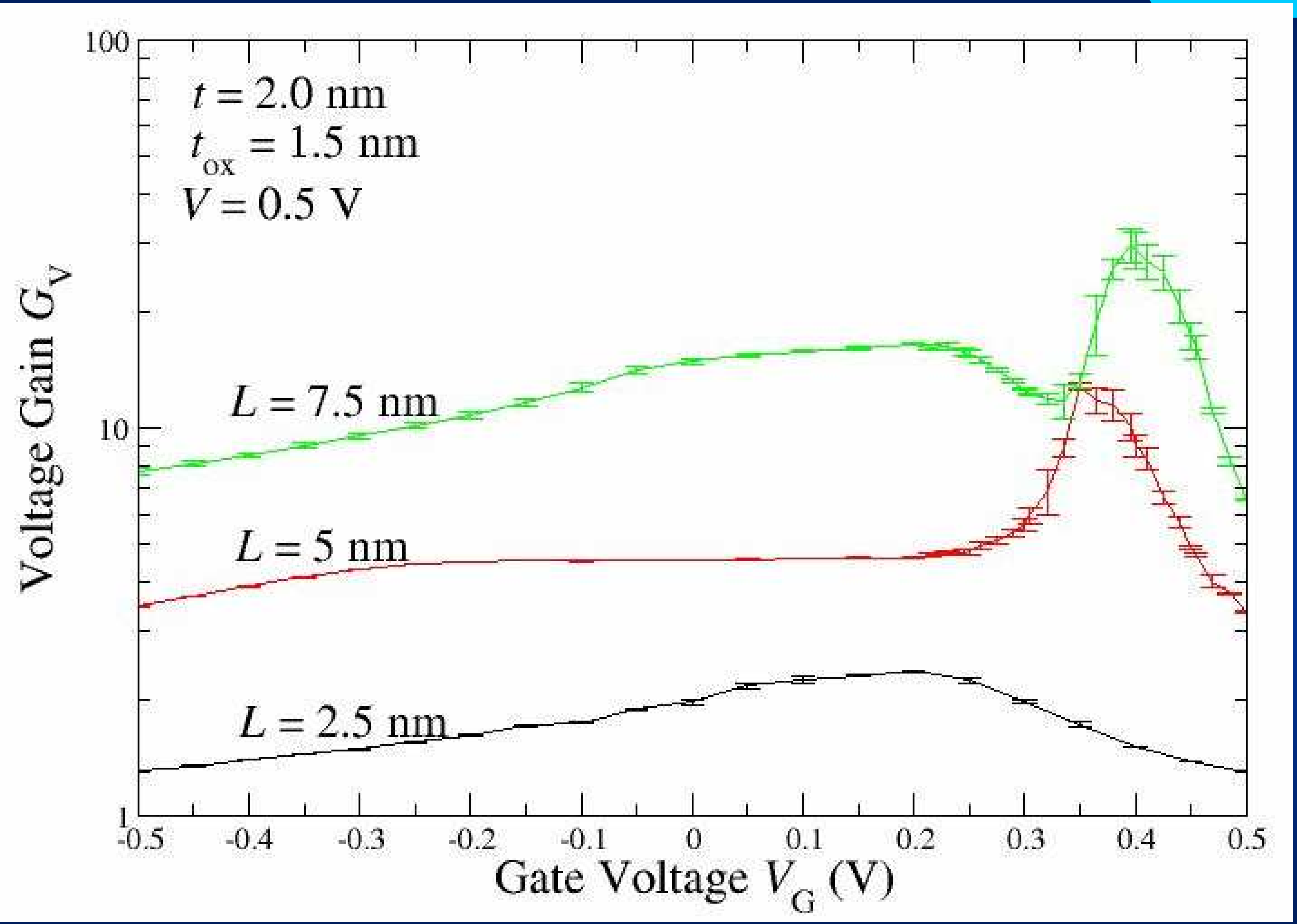
V. Sverdlov, T. Walls, and KKL, IEEE T-ED **50**, 1926 (2003)

SUB-10-NM MOSFETs: RESULTS



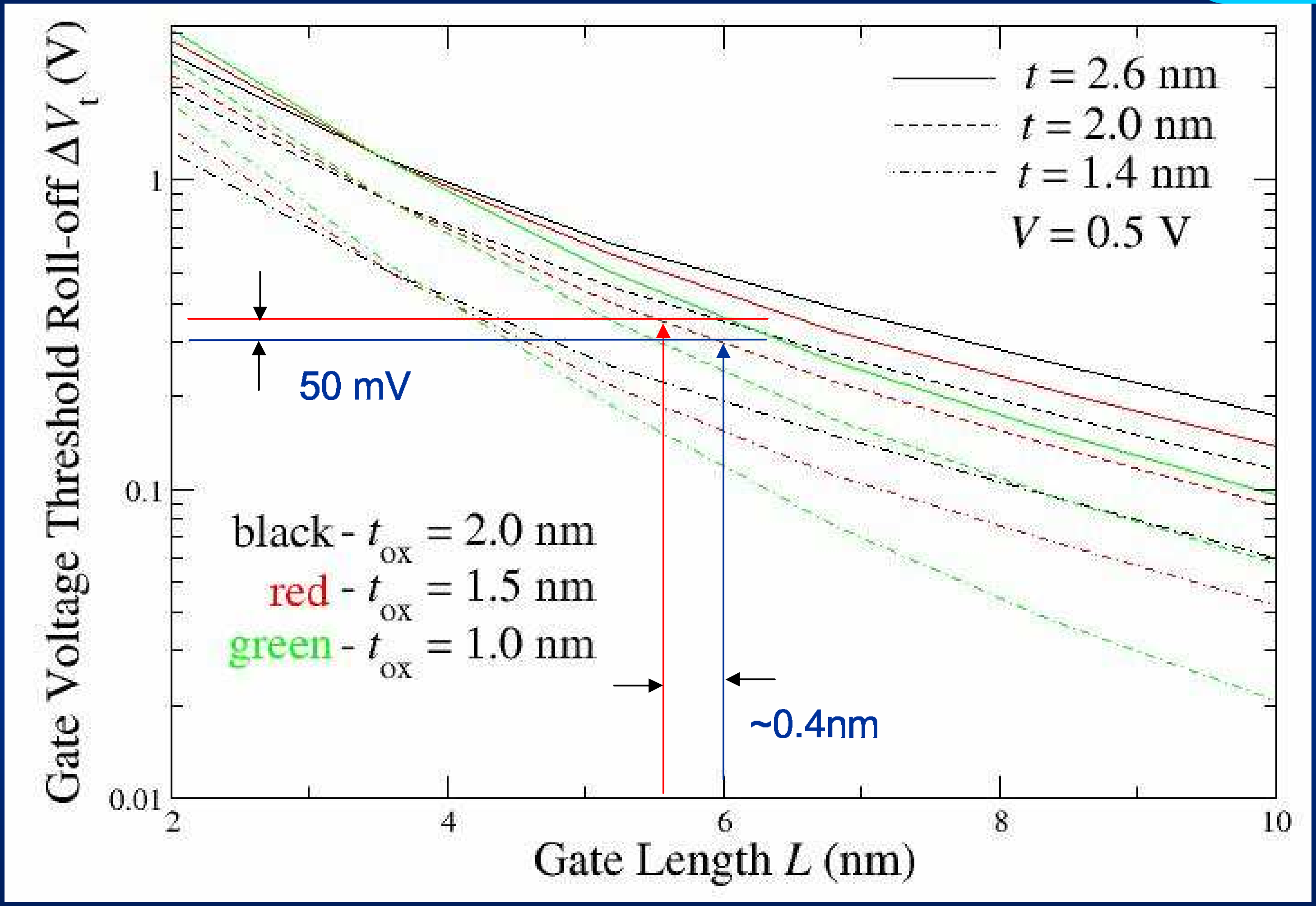
$N_D = 3\times10^{20}$ cm^{-3}

VOLTAGE GAIN



V. Sverdlov, T. Walls, and KKL, IEEE T-ED **50**, 1926 (2003)

PROBLEM: FAB SENSITIVITY



V. Sverdlov, T. Walls, and KKL, IEEE T-ED **50**, 1926 (2003)

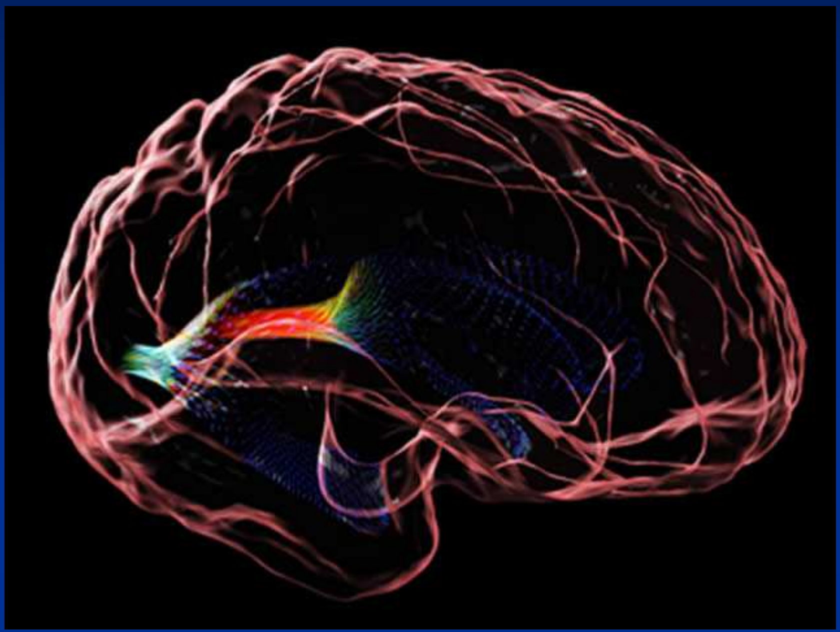
$N_D = 3\times10^{20}$ cm^{-3}

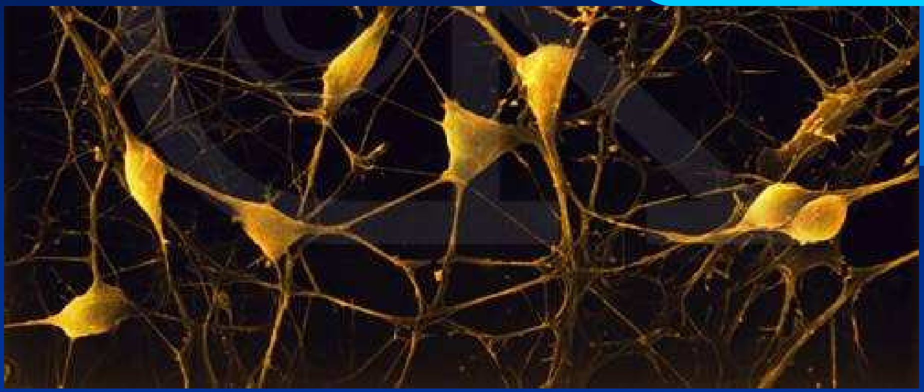
ULTIMATE CMOS PROSPECTS RANGE

	Pessimistic	Optimistic
Minimum half-pitch *F*	45 nm (Yr. 2010)	20 nm (Yr. 2016)
Physical gate length *L*	18 nm	9 nm
Transistor density *n*:	5×10^{9} cm^{-2}	3×10^{10} cm^{-2}
$(k_B T \ln 2)\times f\times n$	0.1 W/cm^2	2 W/cm^2

(much below the real power consumption!)

CORTICAL CIRCUITRY





Areal density:

Cells: ~ 1.5×10^{7} cm$^{-2$}

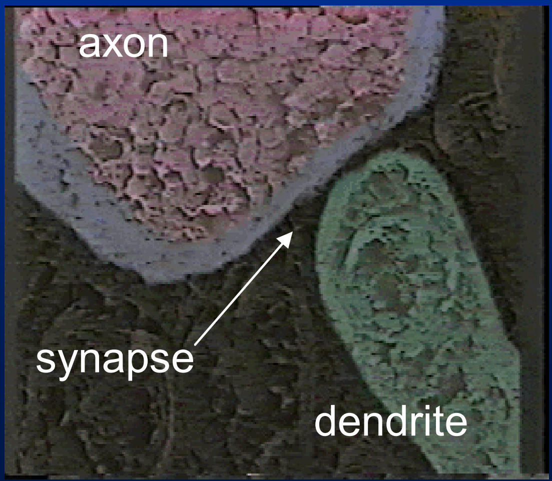
Synapses: ~ 1.0×10^{11} cm^{-2}

Brain:

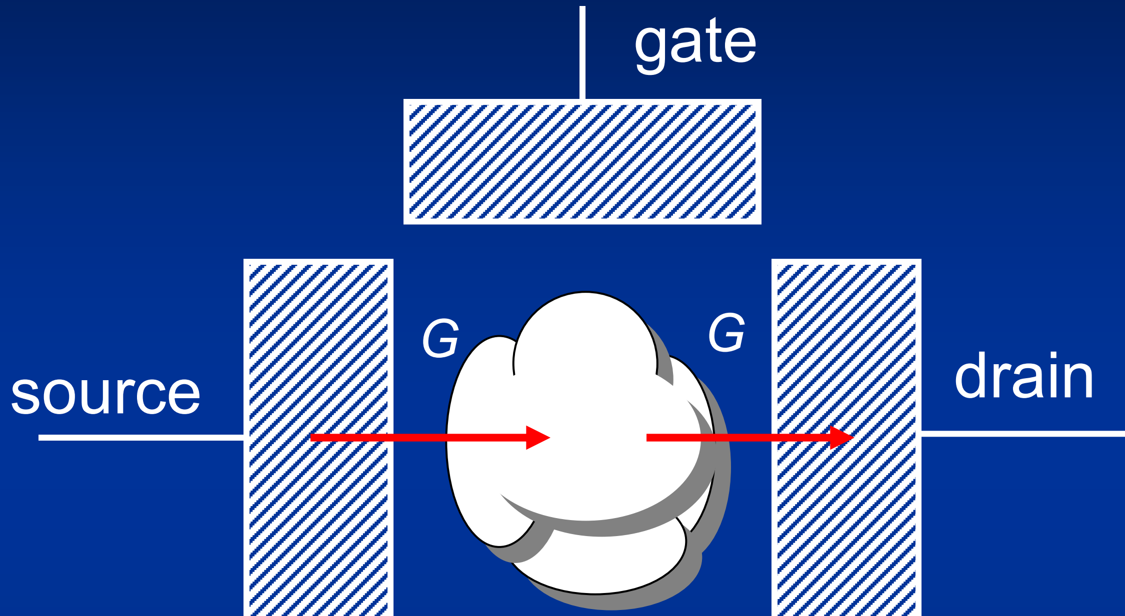
~ 2×10^{10} neural cells

~ few$\times10^{14}$ synapses

Each synapse is an "active device" (5-10 transistors)



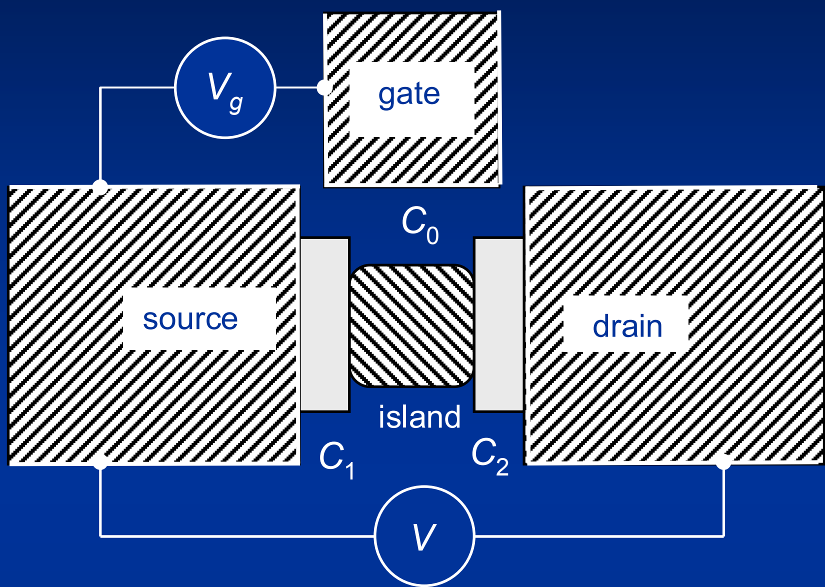
TRANSISTORS: SET vs FET



Choice: $G \lessgtr e^2/\hbar$

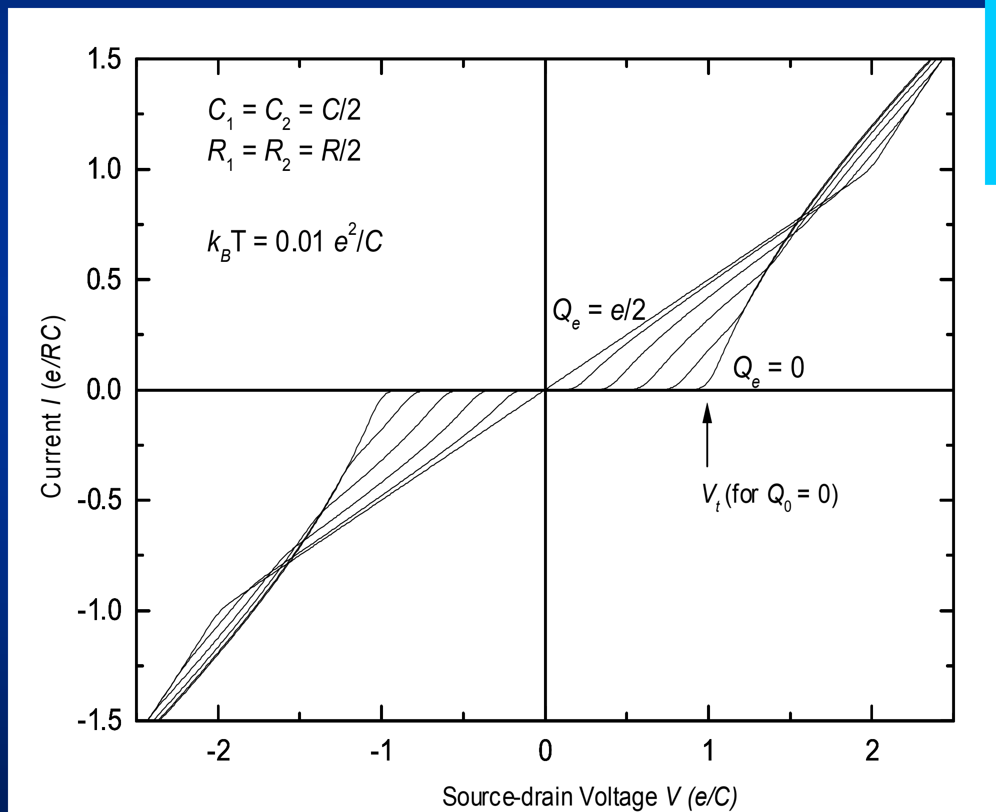
FET ↔ SET

SINGLE-ELECTRON TRANSISTOR

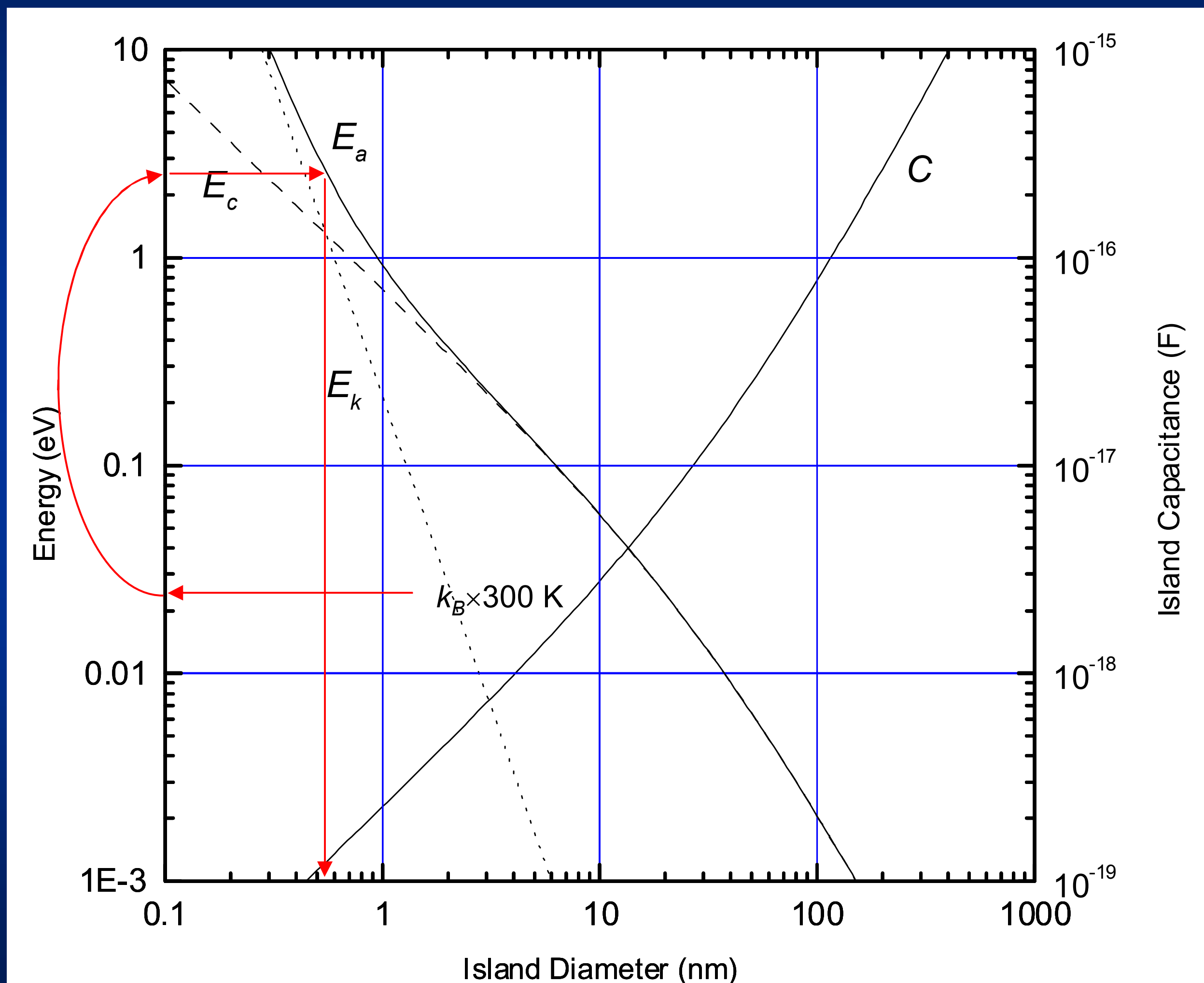


Averin and Likharev, 1985 (theory)

Fulton and Dolan, 1987 (experiment)



SET PROBLEM #1: FABRICATION



E_c ⏲ $10^2\ k_BT$

MINIMUM FEATURE SIZE PHYSICS

Field-Effect Transistors:

(i) $D_Q \sim 1$ at $L \sim$ ◎$/(mE_g)^{1/2} \sim 2$ nm (for Si)

(ii) $D_Q \sim D_T$ at $L \sim$ ◎$/(2mk_BT)^{1/2} \sim 8$ nm (T = 300 K)

Single-Electron Transistors and Other Devices:

◎$^2/2ma^2 > \ln(1/p)\, k_BT$, *i.e.* a 🕘 ◎$/(2mE)^{1/2}$,

where $E \approx \ln(1/p)\, k_BT \sim 1$ eV; as a result, a 🕘 1 nm

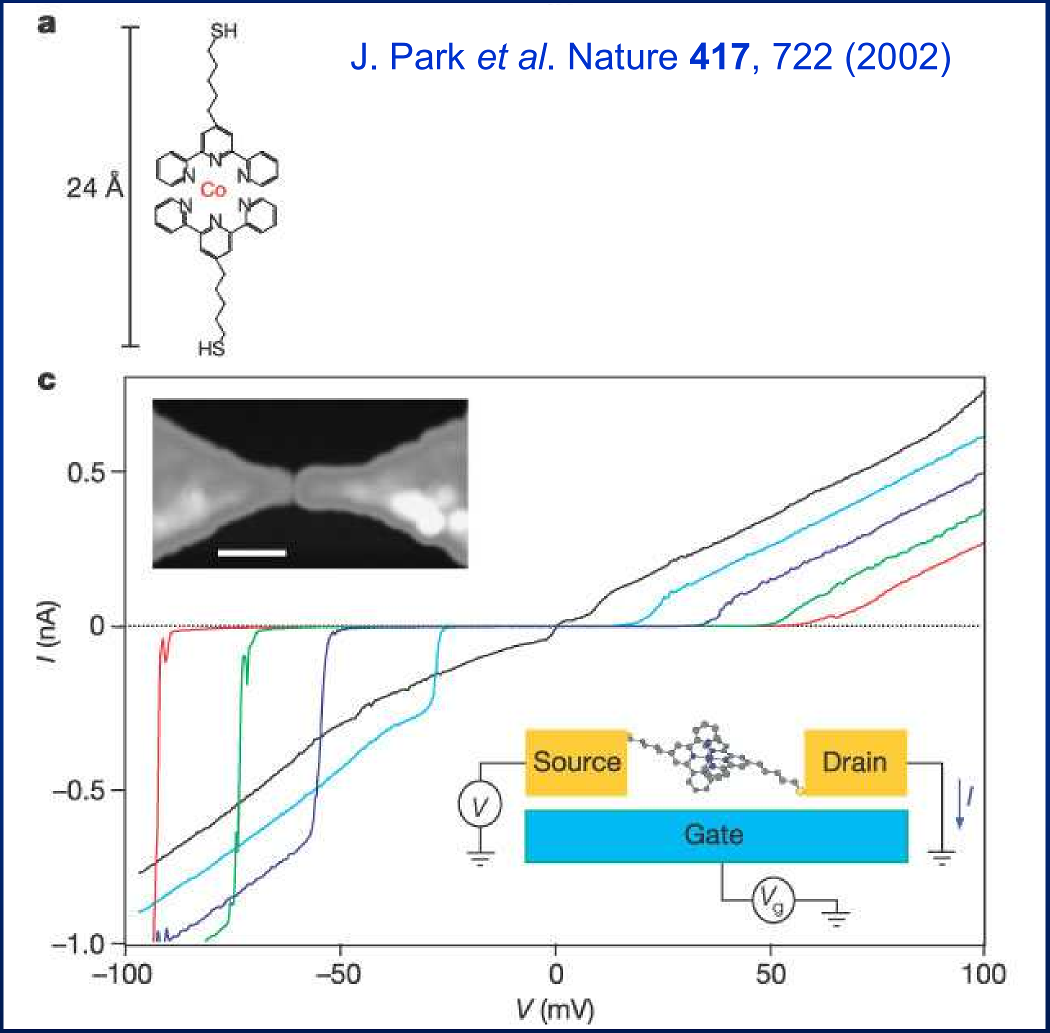
Quantum Interference Transistors:

a 🕘 $\lambda_B/4 \approx$ ◎$/4[2m(E\text{-}U)]^{1/2}$, with $(E - U)$ 🕘 $3\, k_BT$,

giving a 🕘 2 nm

A (quasi-) universal crossover length - hard to reach!
Possible remedy: use natural standards of 1-nm lengths

SINGLE-ELECTRON SINGLE-MOLECULE TRANSISTORS

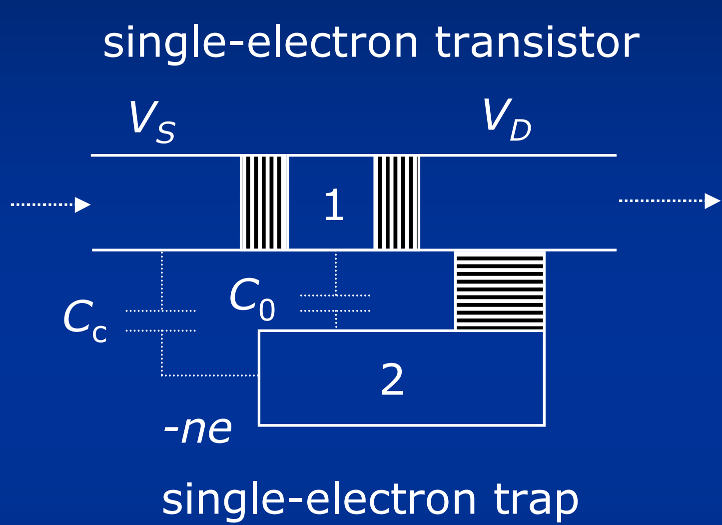


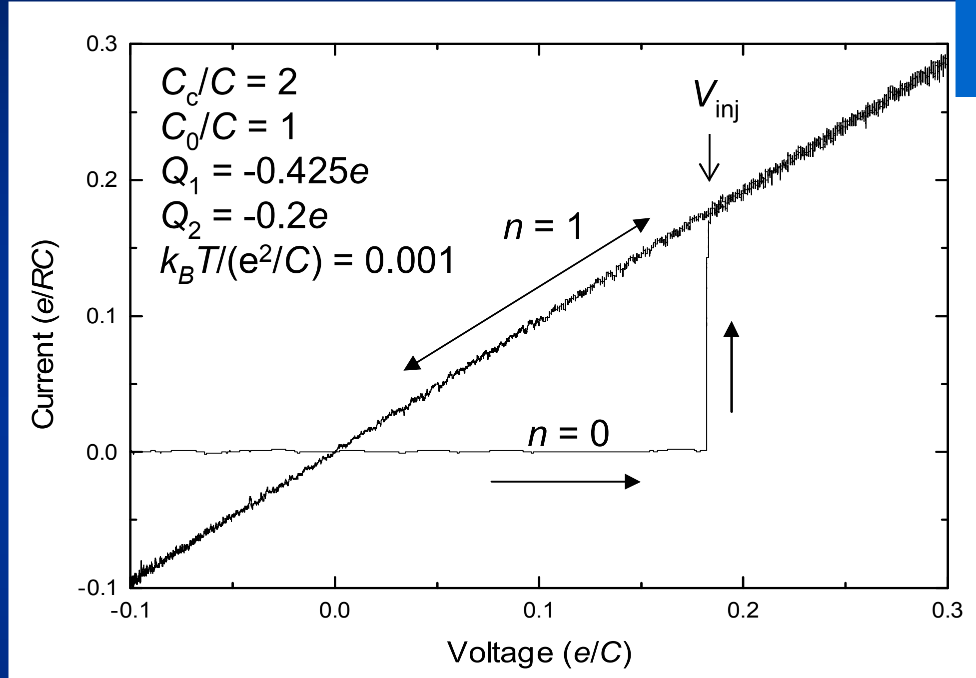
J. Park *et al*. Nature **417**, 722 (2002)

see also:

- E. S. Soldatov *et al.* JETP Lett. **64**, 556 (1996)
- H. Park *et al*. Nature **407**, 57 (2000)
- N. Zhitenev, H. Meng, and Z. Bao PRB **88**, 226801 (2002)

SINGLE-ELECTRON LATCHING SWITCH





"quasi-fuzzy"dynamics:

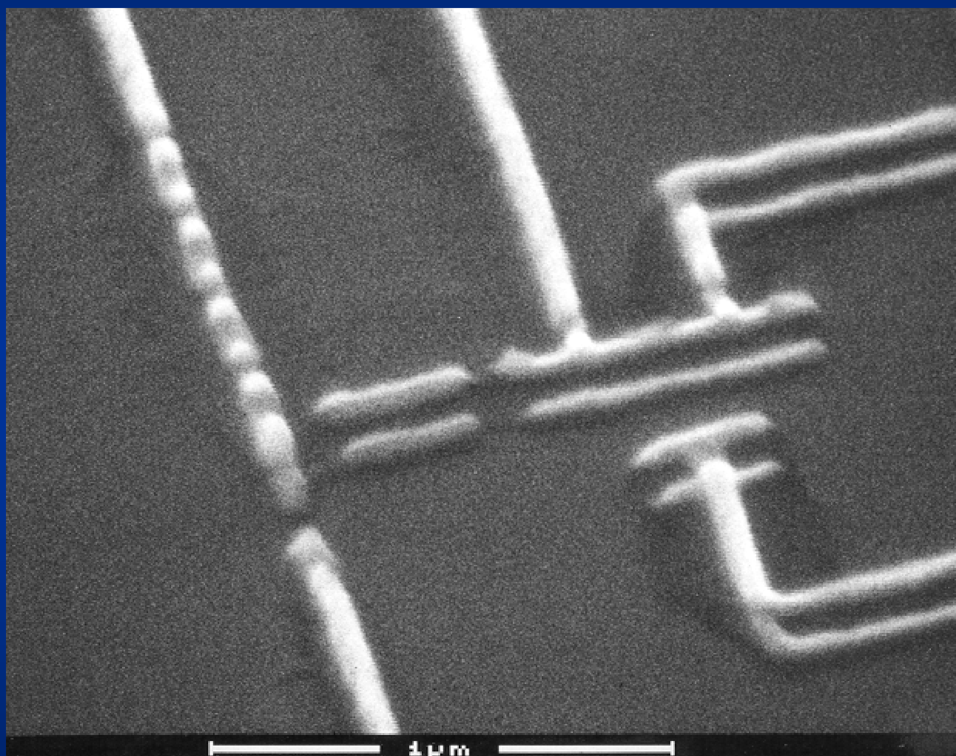
$dp/dt = \Gamma_{\uparrow}(1-p) - \Gamma_{\downarrow}p,$

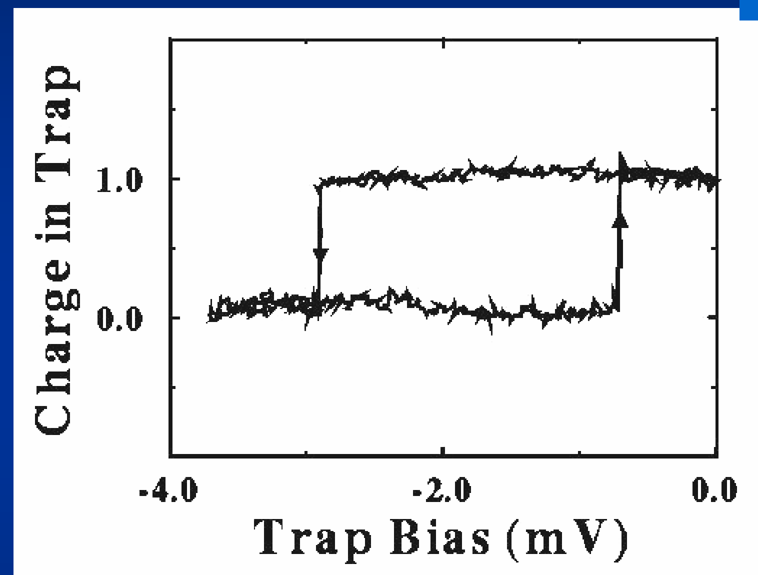
$\Gamma_{\uparrow\downarrow} = \Gamma_0 \exp\{\pm e(V-S)/k_B T_{ef}\},$

S. Fölling, Ö. Türel, and K.L., 2001

SINGLE-ELECTRON LATCHING SWITCH:

low-*T* prototype



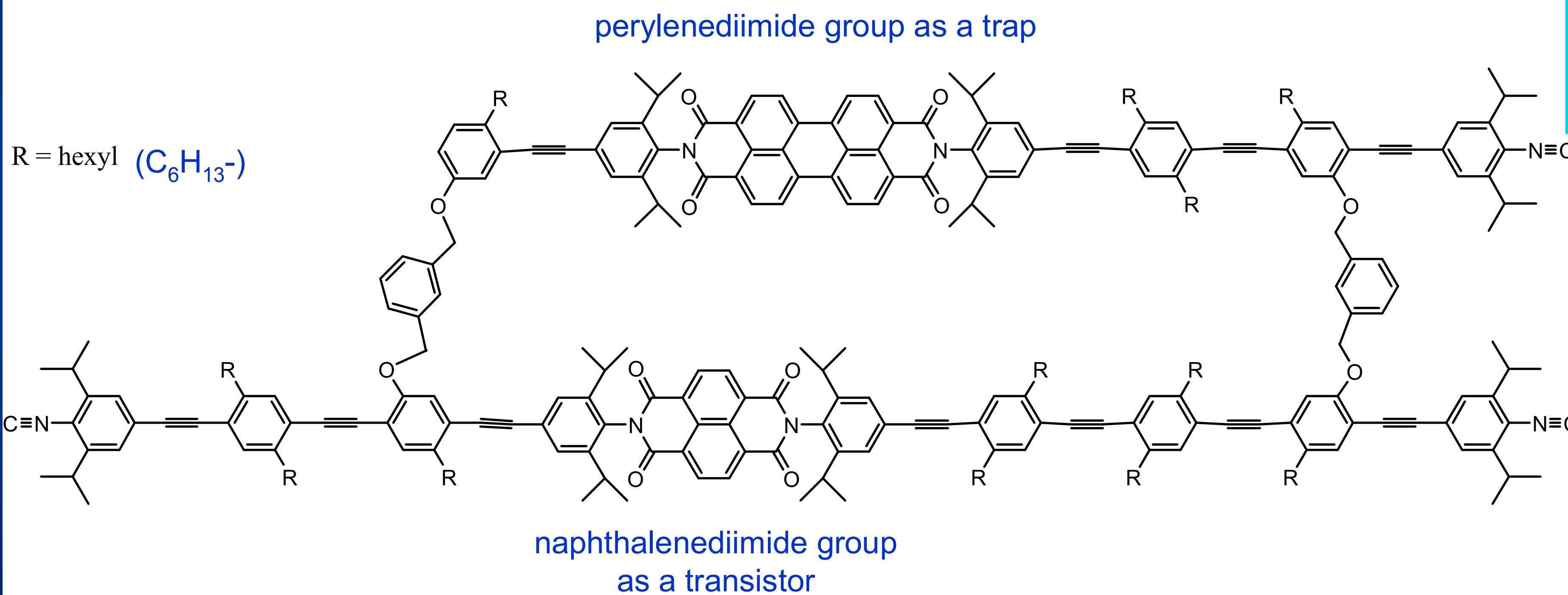


P. Dresselhaus *et al.*, 1994

- Al/AlO$_x$/Al structure
- stored an electron for > 12 hrs (at T < 1 K)

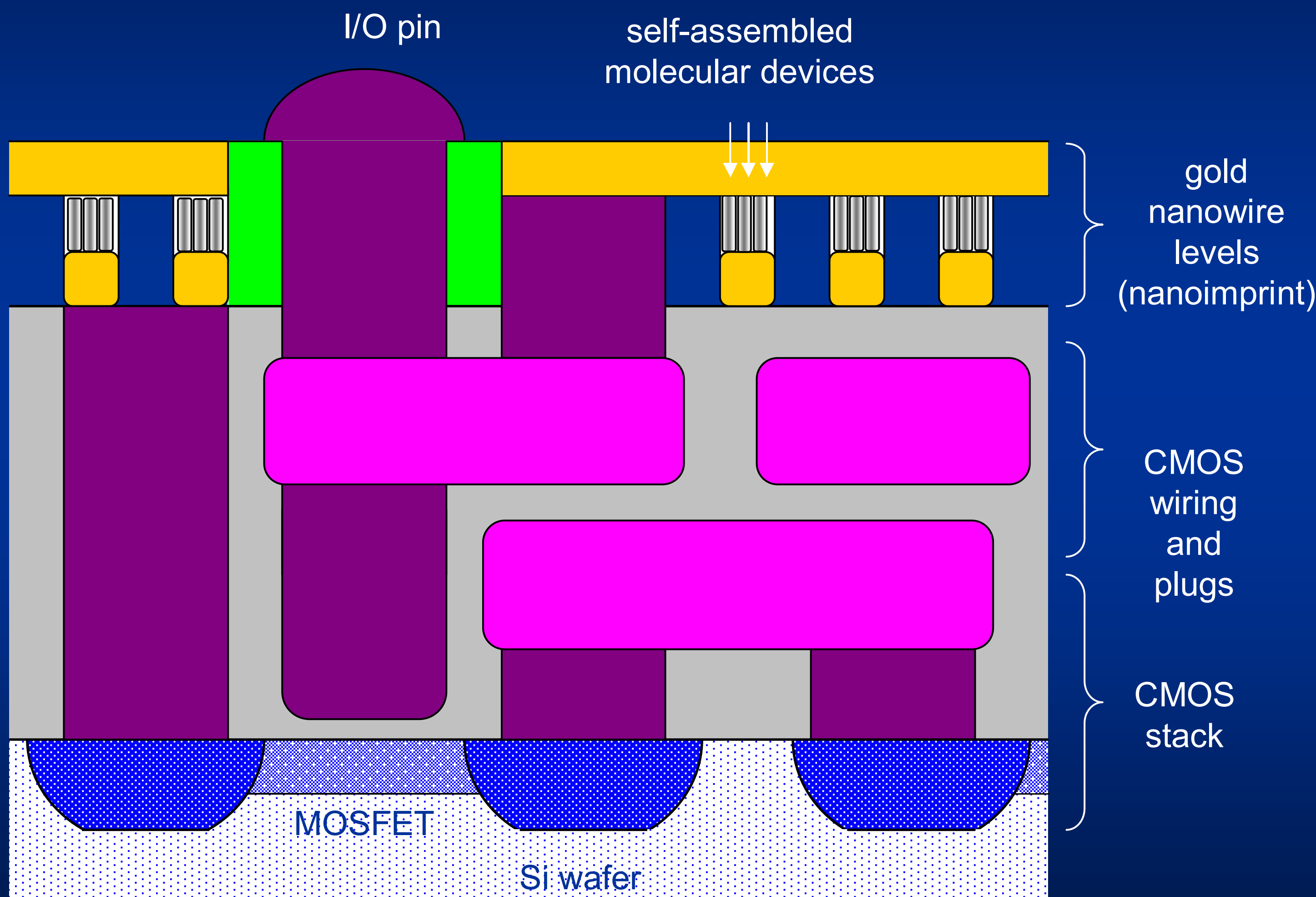
SINGLE-ELECTRON LATCHING SWITCH:

possible molecular implementation

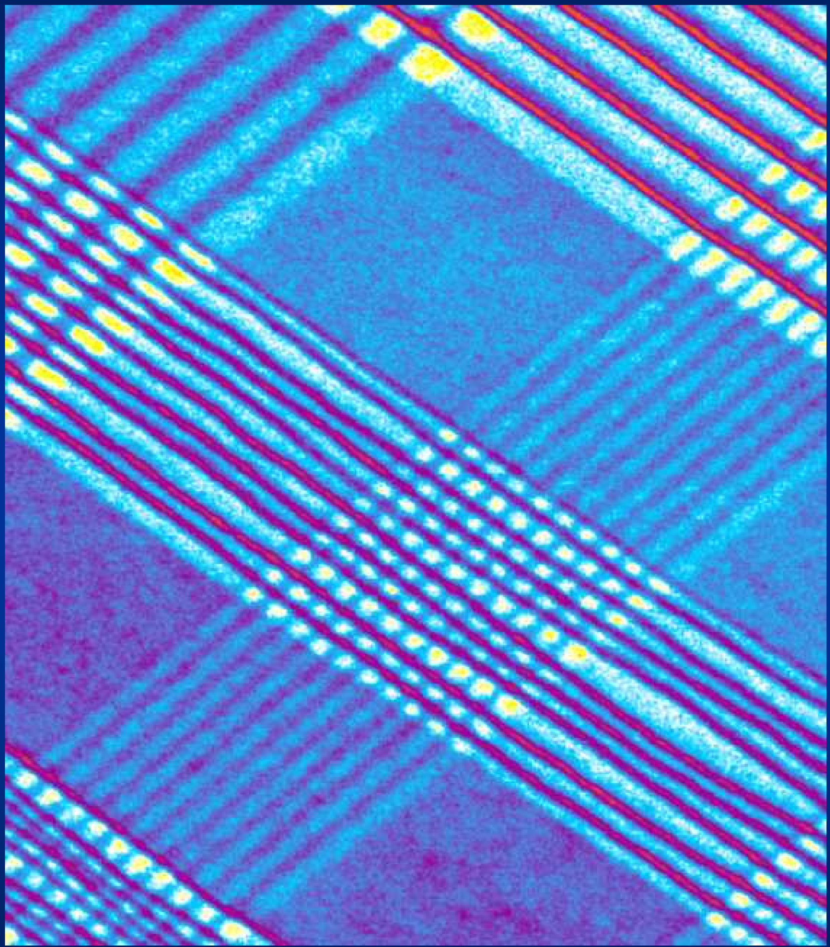


Courtesy A. Mayr (SBU/Chemistry)

CMOL CONCEPT

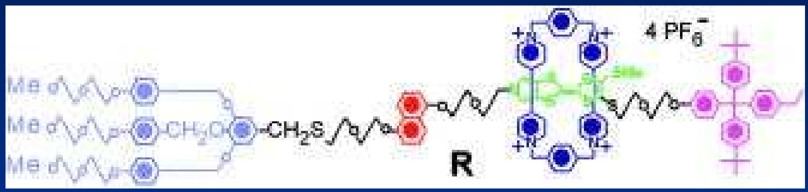


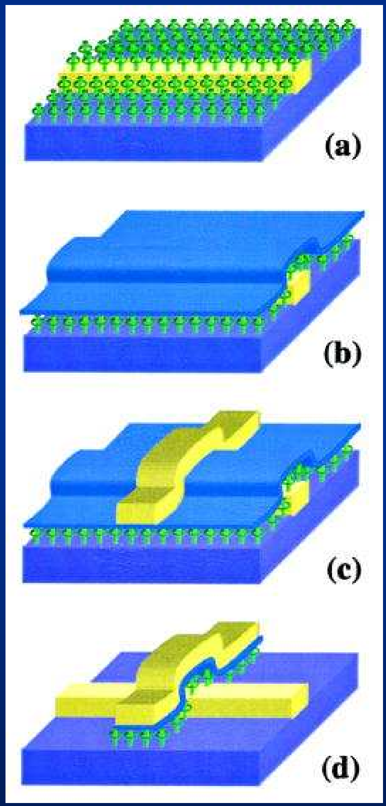
TOWARD CMOL-TYPE MEMORIES

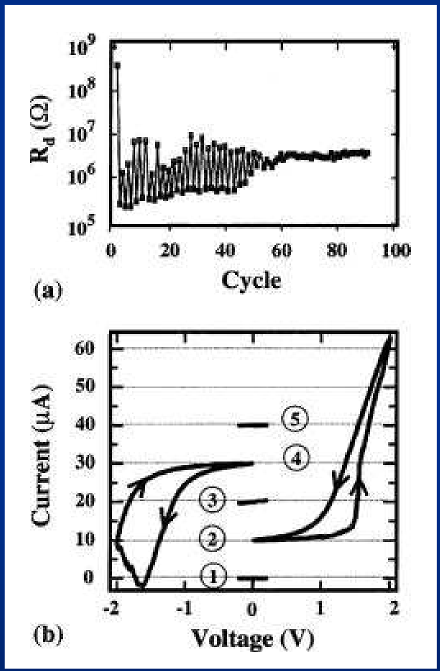


J. Heath and M. Ratner, Phys. Today, May 2003

(picture F. Krausz, HPL)







Y. Chen *et al.* APL **82**, 1610 (2003)

LAST NEWS

Intel enlists Nanosys to probe nano-memory research

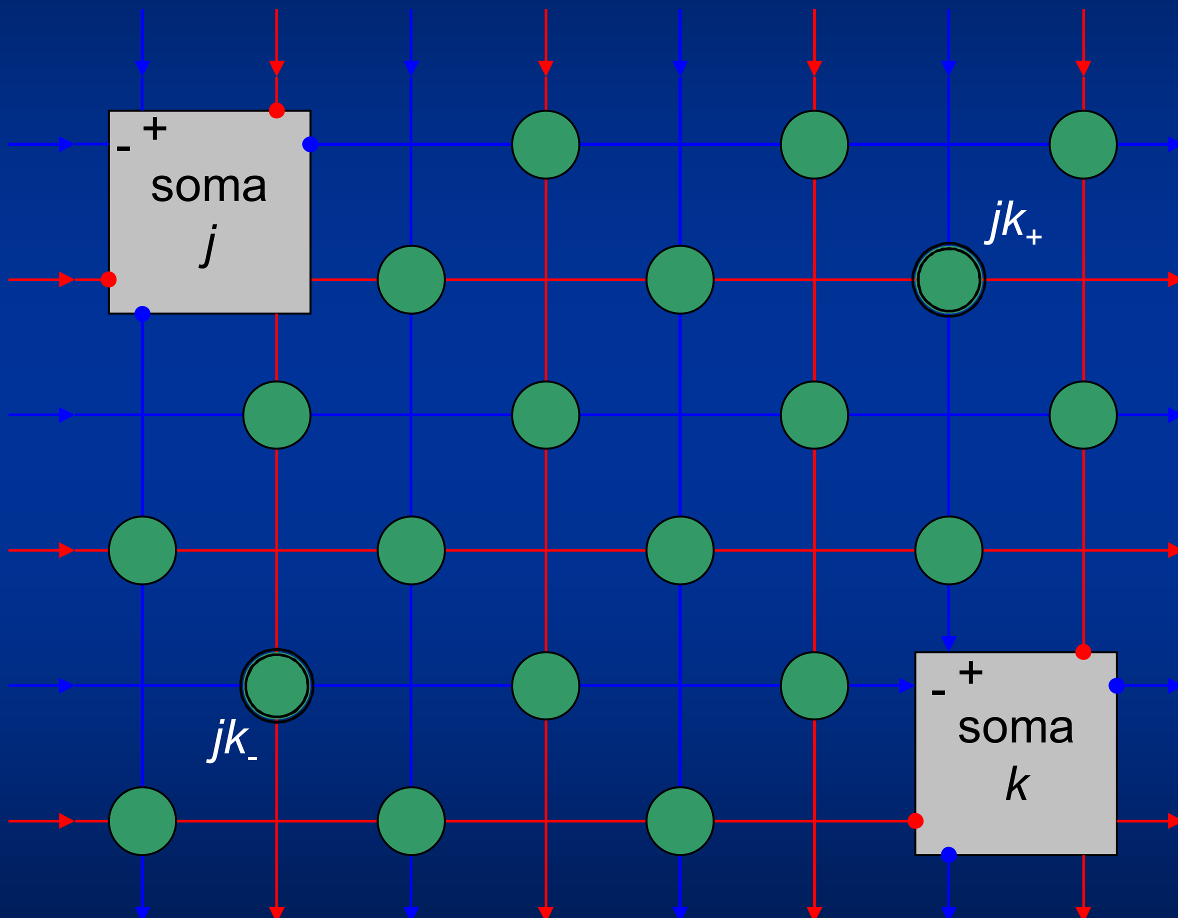
Silicon Strategies
January 15, 2004 (9:42 a.m. ET)

PALO ALTO, Calif. — Nanosys Inc. said Wednesday (Jan. 14) it will work with Intel Corp. to investigate using nanometer technology for future memory systems.

Intel will support nano-related technology efforts at Nanosys for possible use in memory products. According to the agreement, Nanosys and Intel will work together exclusively on certain areas of memory-related technologies for a specified period of time.

CROSSNET: GENERAL STRUCTURE

(feedforward option)

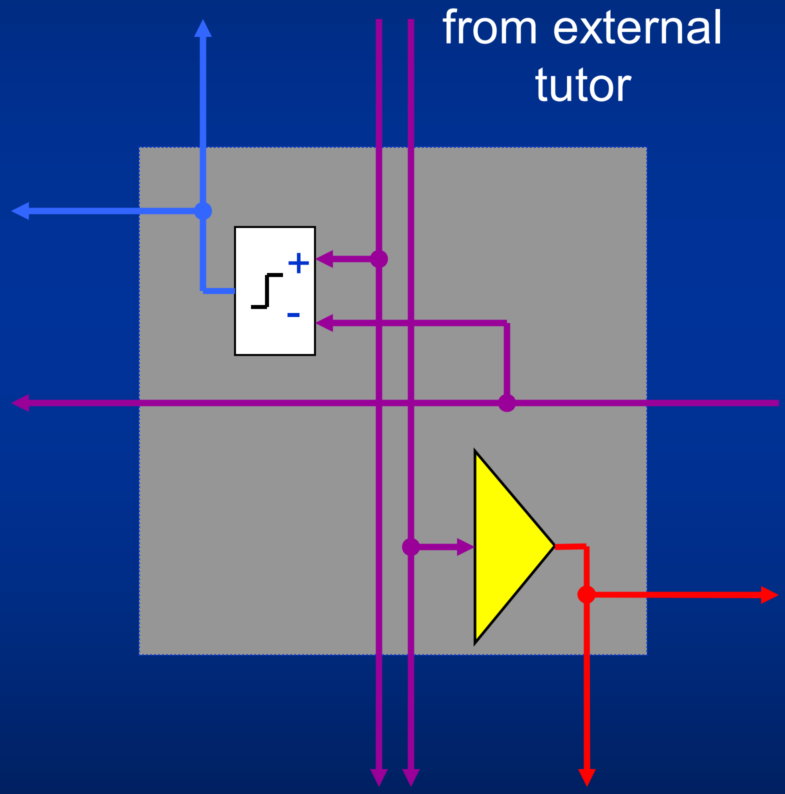


$w_{jk} = \{-1, 0, +1\}$

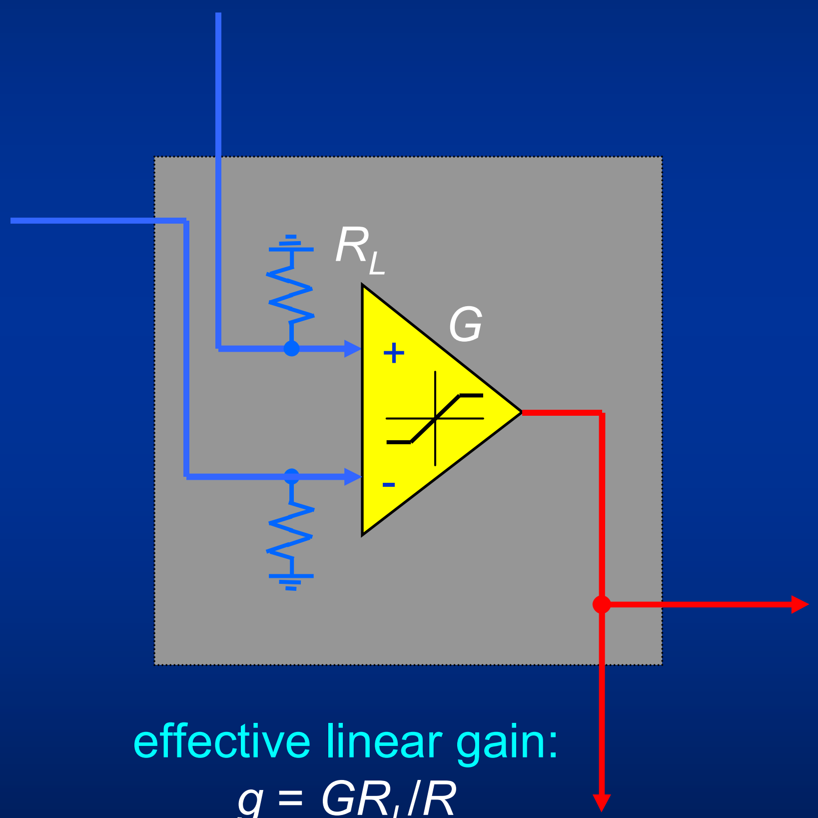
"GRAY CELL" (SOMA)

(fire-rate model, feedforward network)

Training mode:



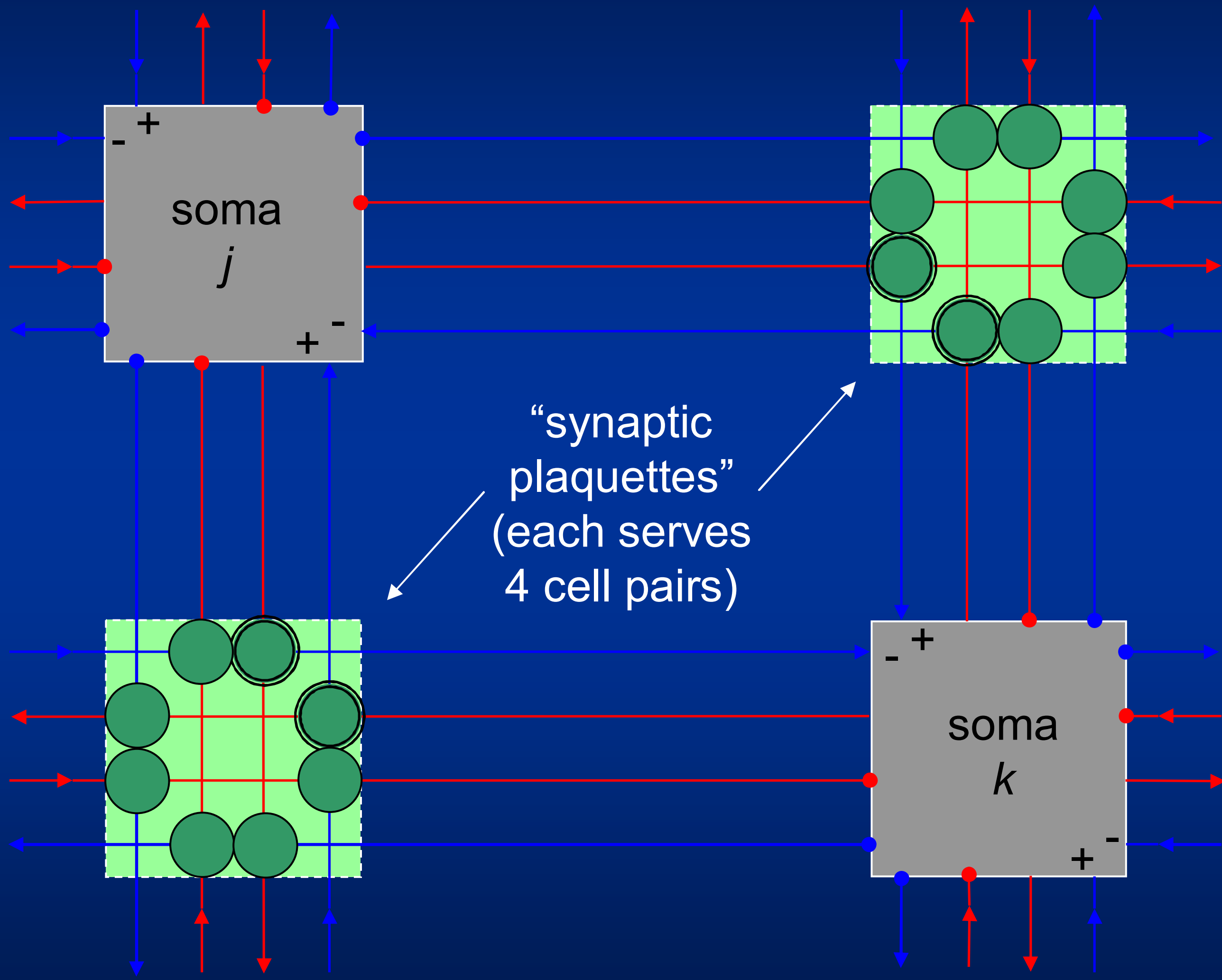
Working mode:



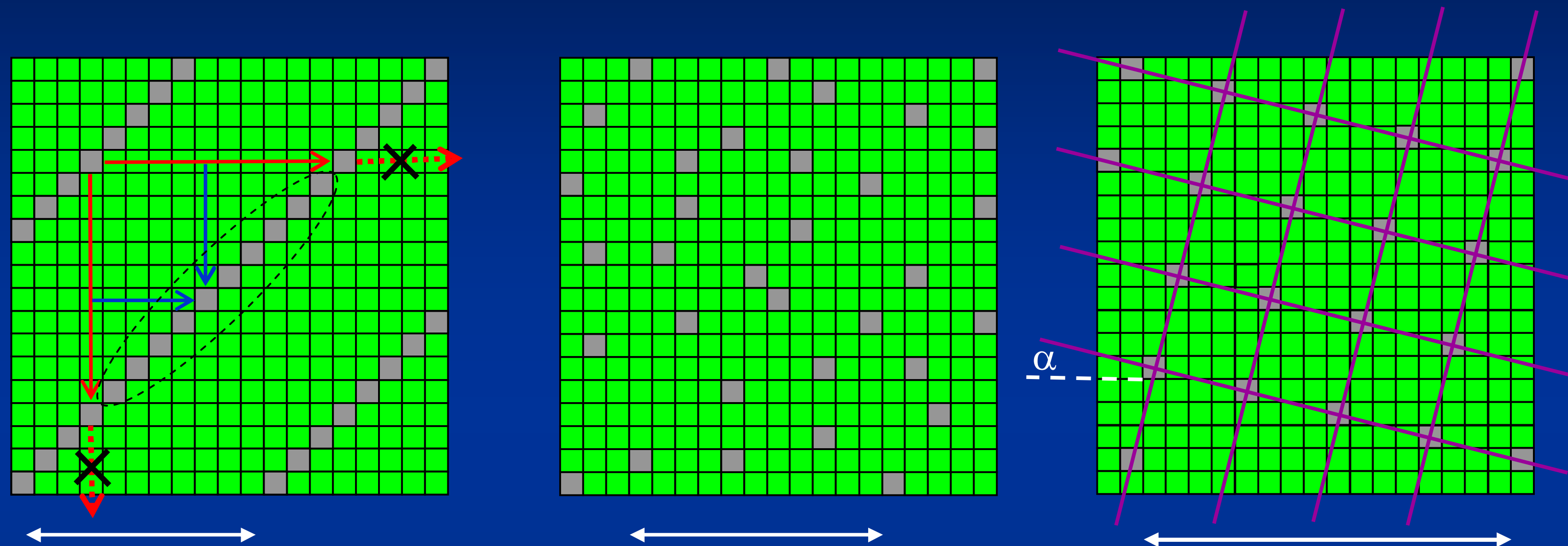
effective linear gain:

$g = GR_L/R$

RECURRENT CROSSNET



CROSSNET SPECIES

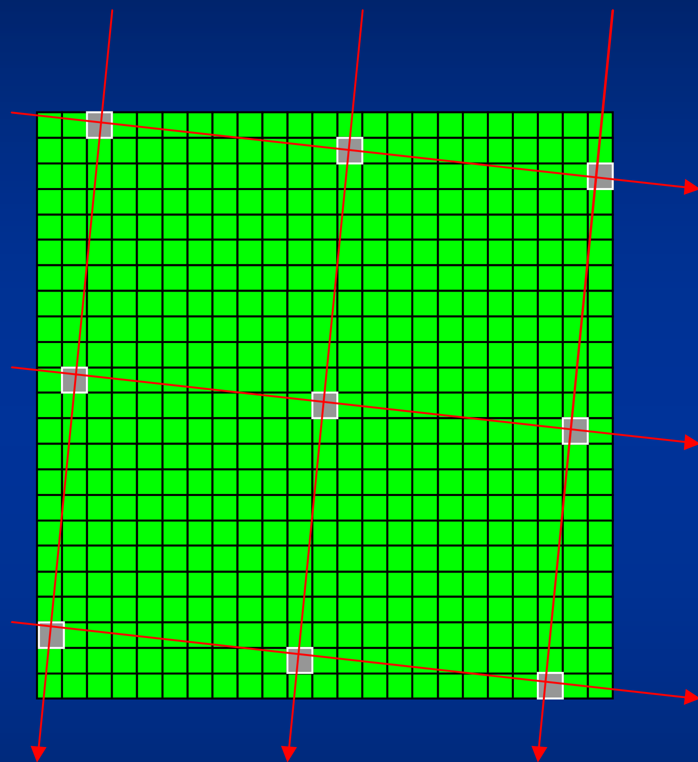


$\Delta x = \Delta y = \text{const} = M$ | $\langle \Delta x \rangle = \langle \Delta y \rangle = M$ | $\Delta x = \Delta y = \text{const} = M = 1/\tan^2\alpha$

FlossBar | RandBar | InBar

Maximum Connectivity: 4*M* (for RandBar, on the average)

HOPFIELD MODE
(Recurrent InBar)



1. Training

- model images in:

$V_j = V_0 \text{sgn} \circledast_p x_j^{(p)} x_k^{(p)}$
$V_k = V_0$
$(V_0 \odot S, T)$
$p = 1, 2, \ldots P$

- synapses adapt:

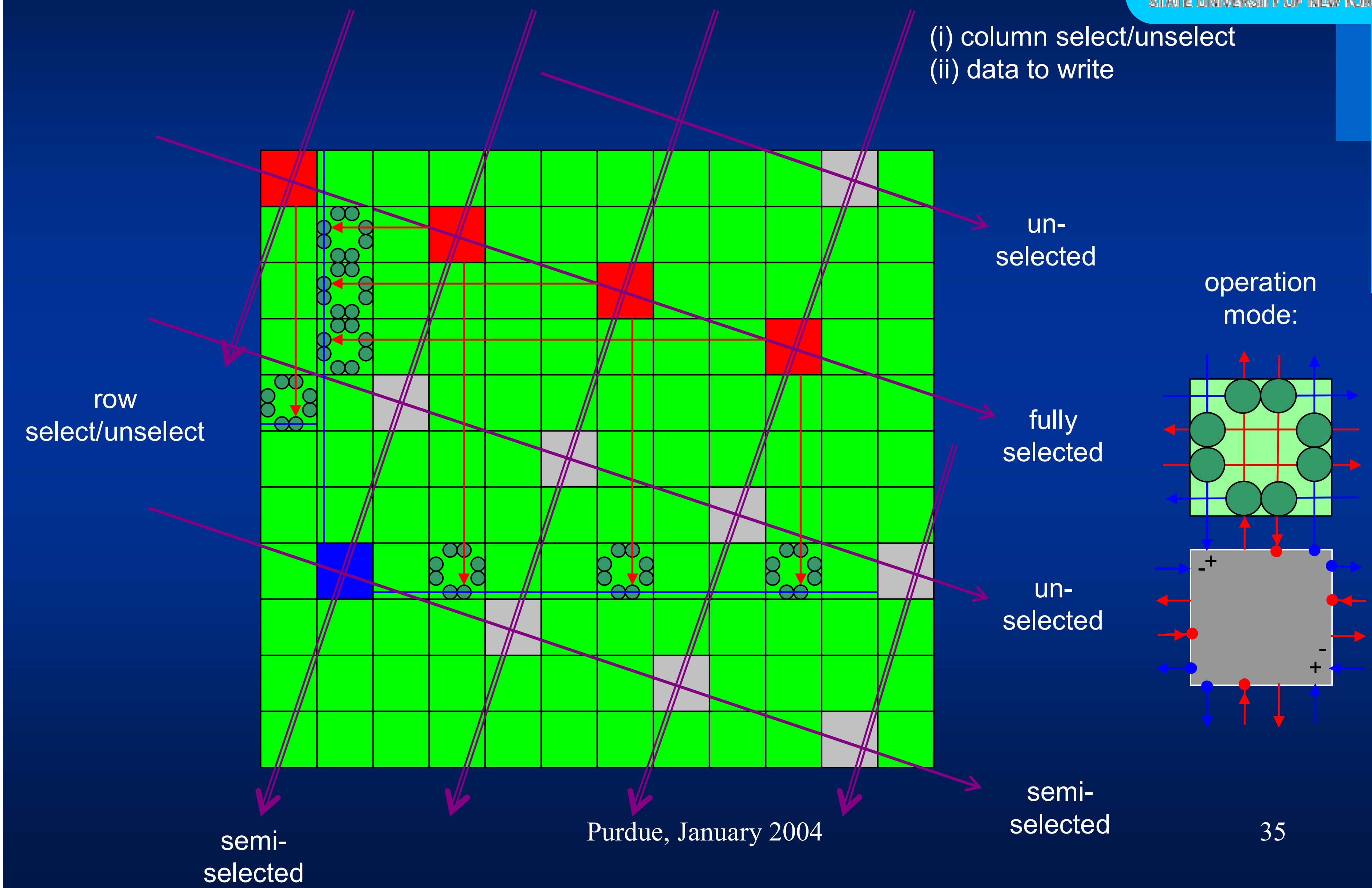
$\langle w_{jk} \rangle \to \text{sgn} \circledast_p x_j^{(p)} x_k^{(p)}$

2. Operation

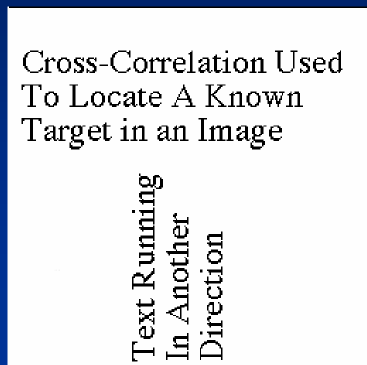
- scrambled inputs in (as initial conditions):
$x_j(0) = x_j^{(p)} + n_j;$

- recognized images out (after a short transient):
$x_j(t) \to x_j^{(p)}$

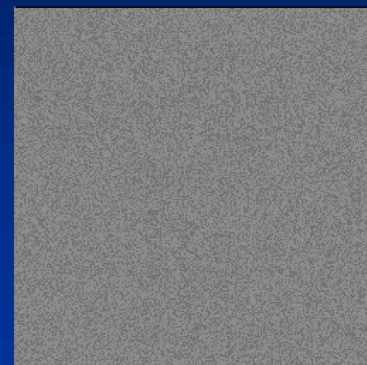
HOPFIELD-MODE TRAINING



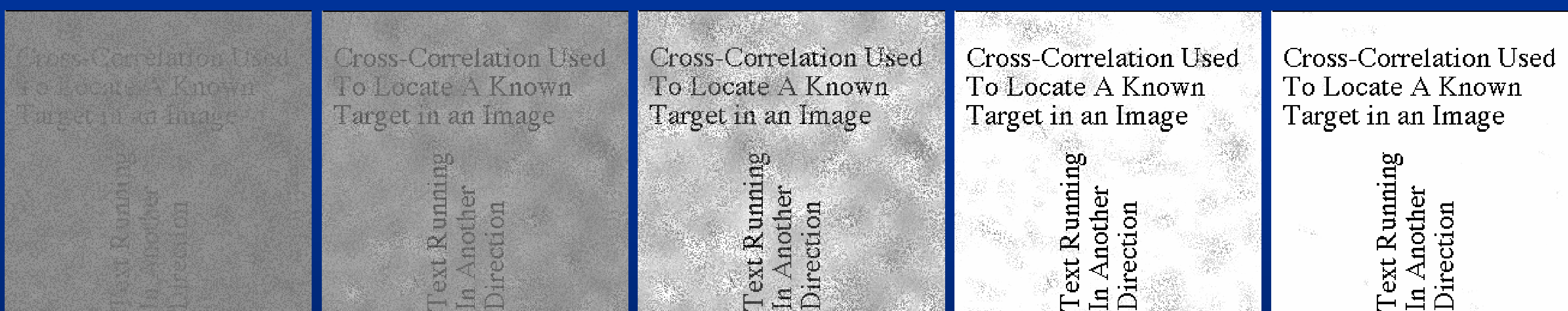
HOPFIELD-MODE IMAGE RECOGNITION: DYNAMICS



original B/W image (1 of 3 taught)



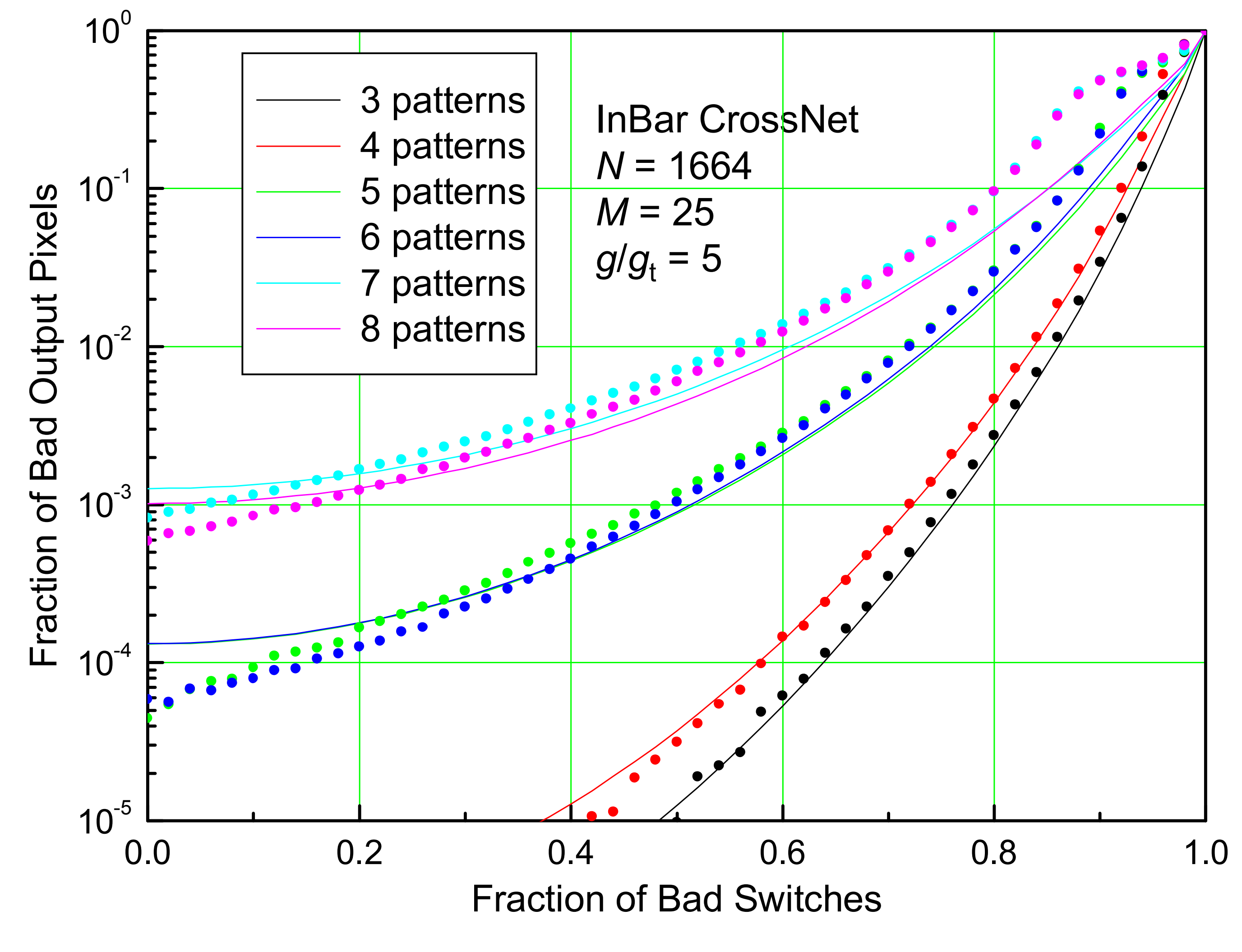
random 40% bits flipped (t = 0)



$t/\tau_0 = 1$ 2   3   4   5

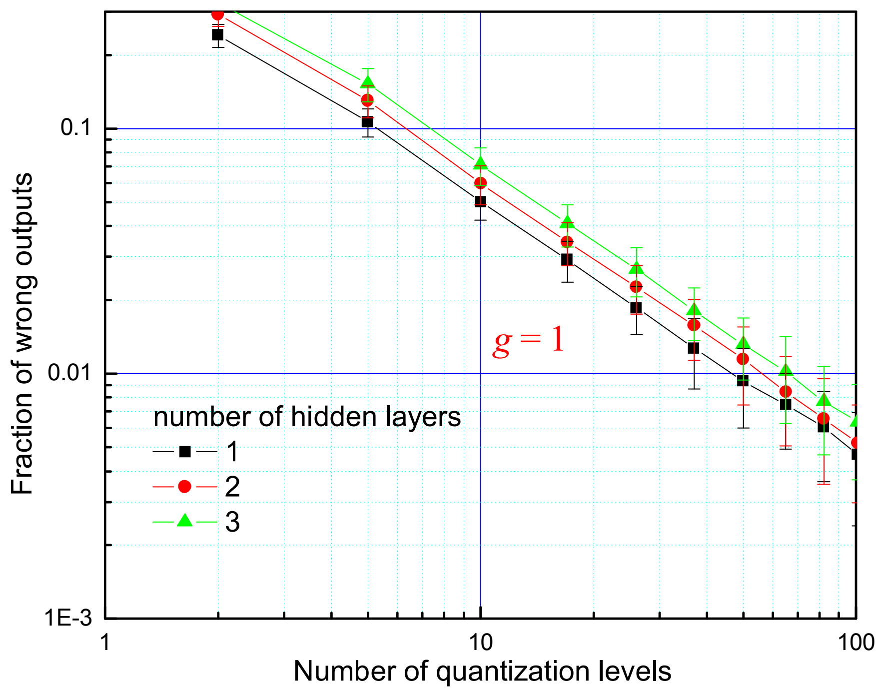
where $\tau_0 \equiv MR_L C_0$

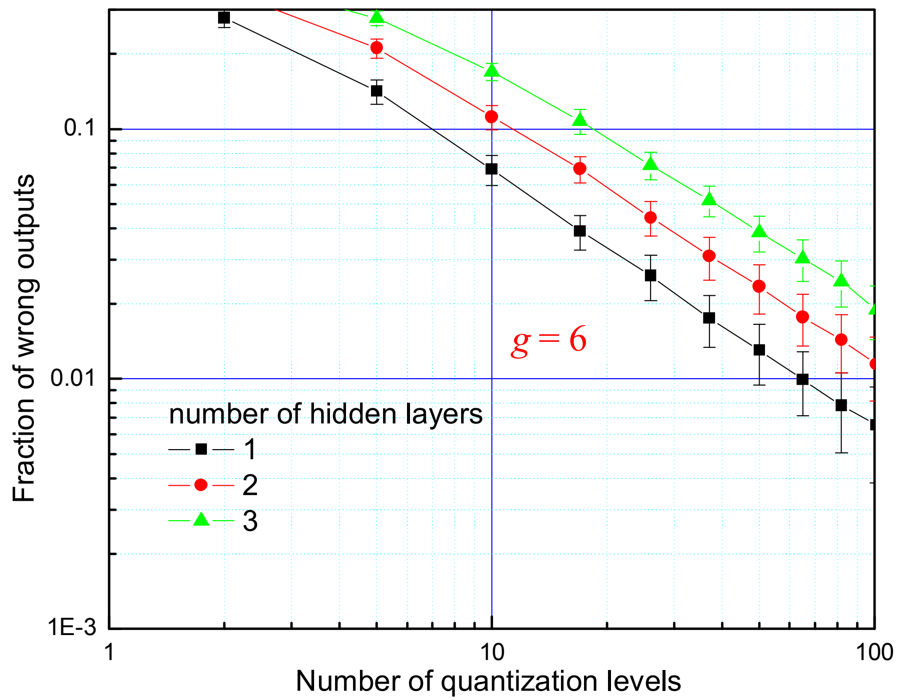
HOPFIELD-MODE OPERATION: DEFECT TOLERANCE



FEEDFORWARD NETWORKS: SYNAPSE DISCRETENESS EFFECT

Multi-layer perceptrons
100 cells per layer
(results averaged over 100 random weight vectors)

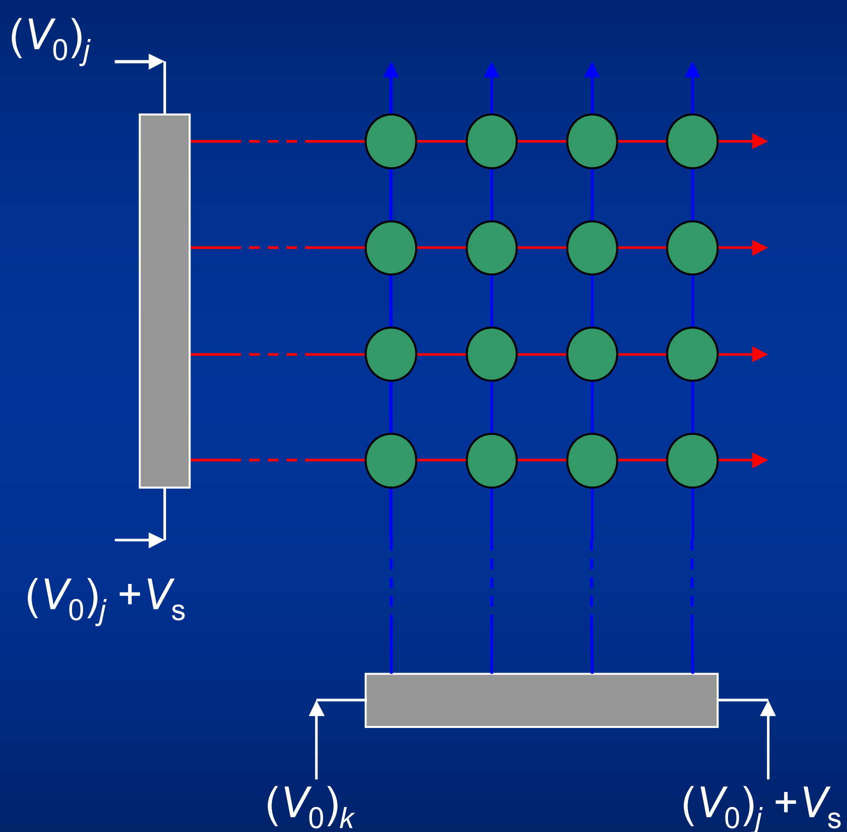




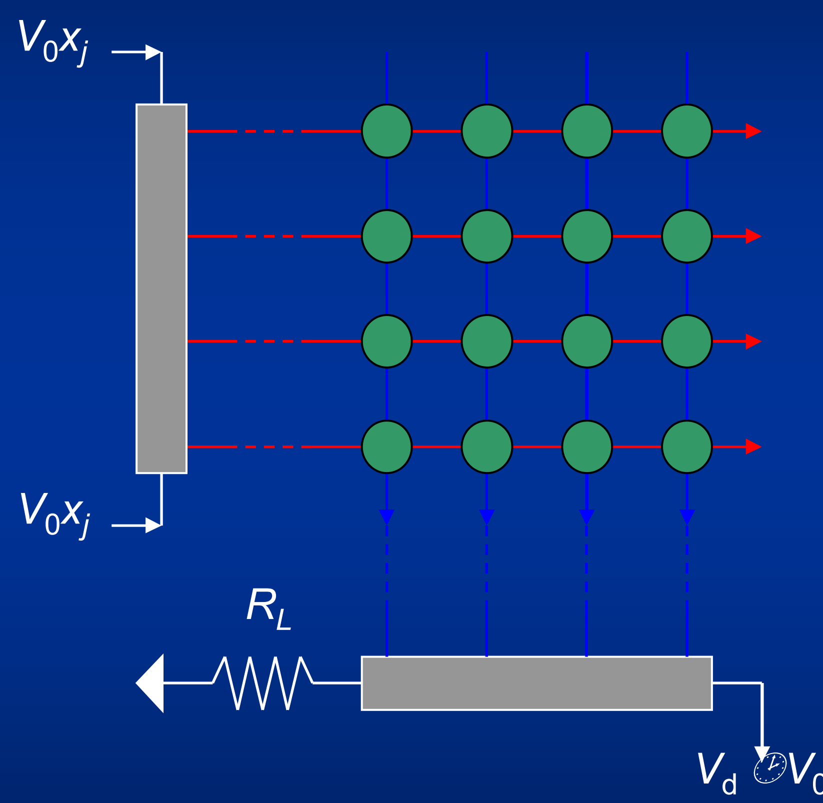
J. Lee, unpublished (2003)

MULTI-VALUED SYNAPSES

Training mode:



Working mode:



Number of levels: $L = 2n^2+1$

$I_{out} = (V_j/R)\Sigma_i n_i$

N-LAYER PERCEPTRON
(feedforward FlossBar with multi-valued synapses)

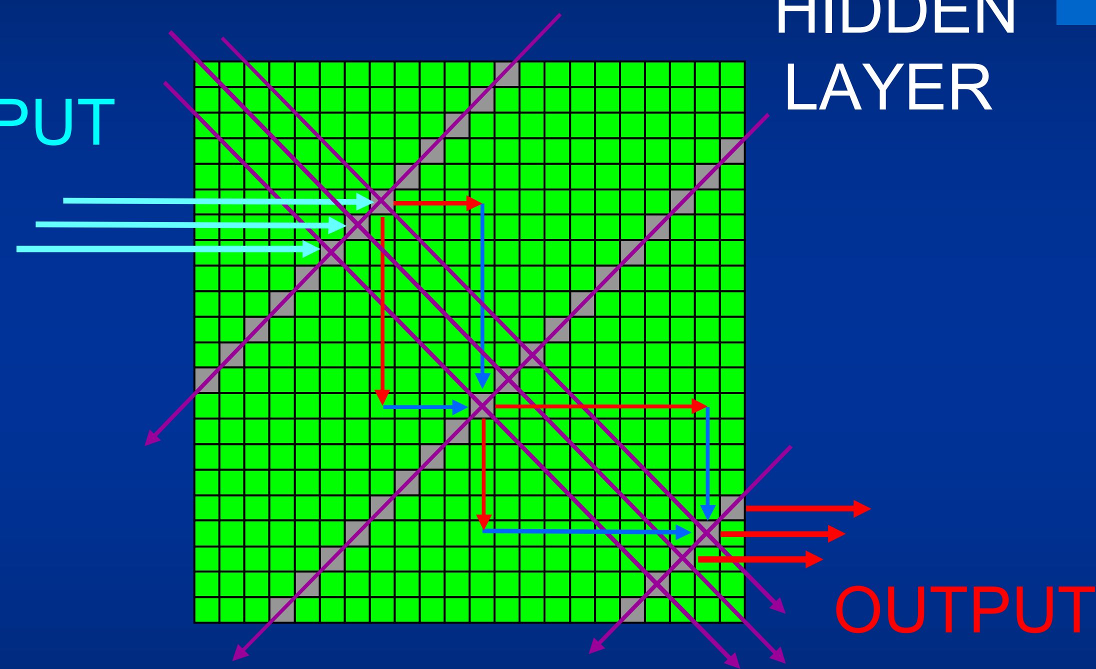
1. Training:

- train a precursor network with continuous weights W_{jk} (e.g., with backprop)

- write W_{jk} into proper somas

- somas enforce discrete weights:

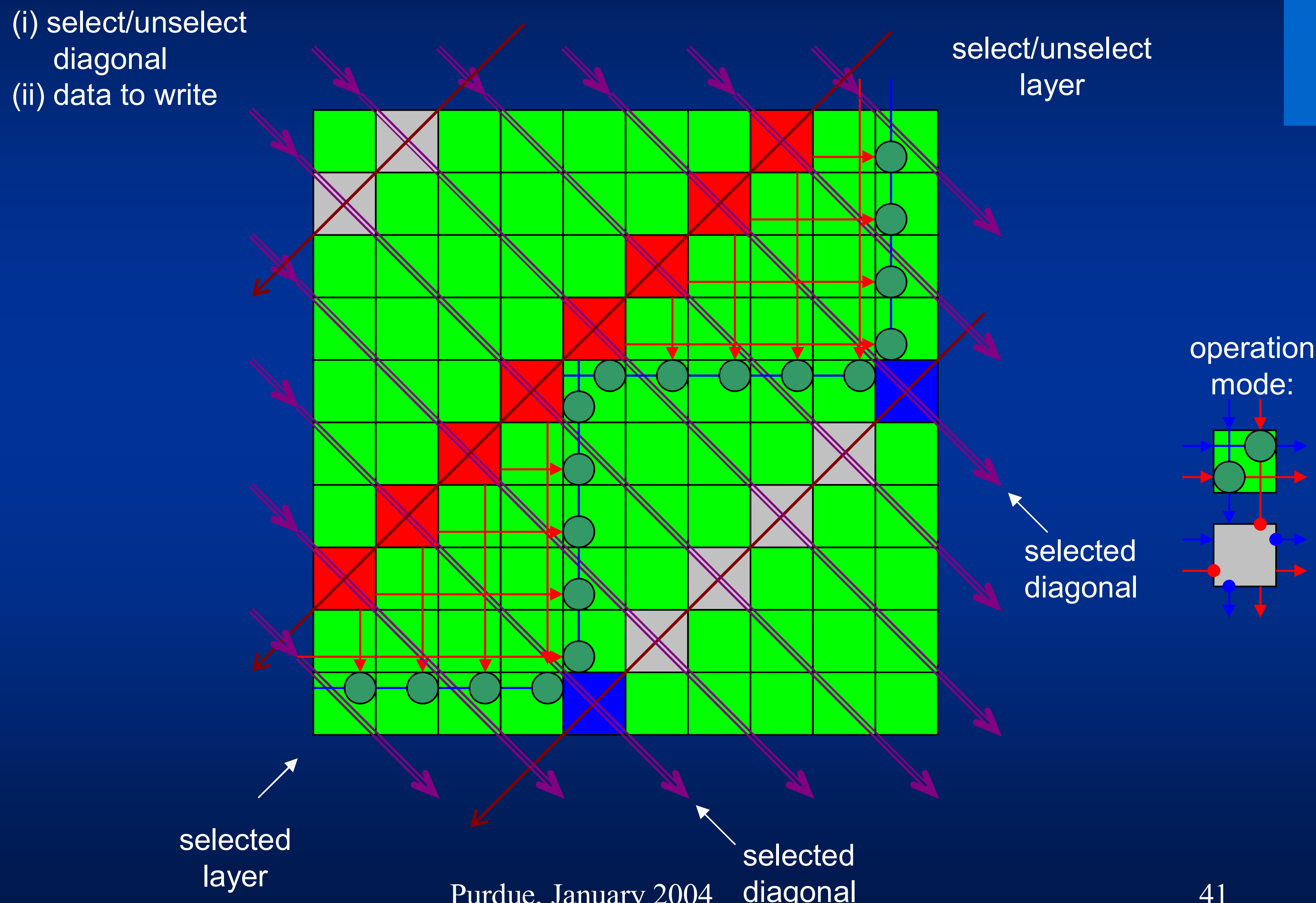
$$w_{jk} \rightarrow \text{floor}\,(LW_{jk})/L$$



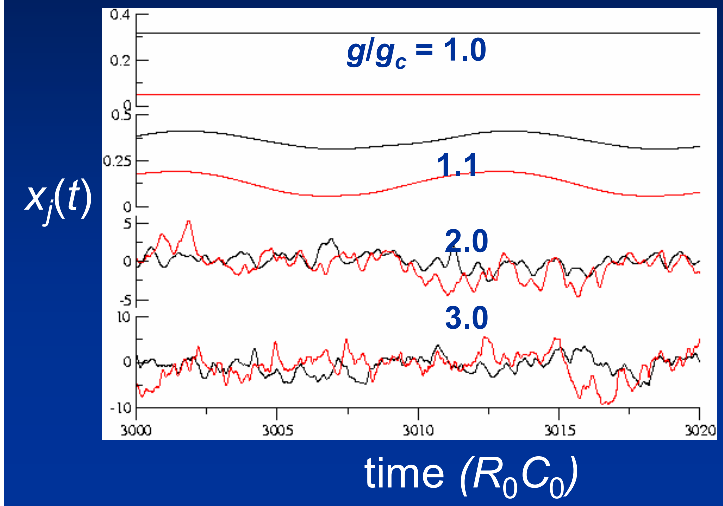
2. Operation (e.g., classification):

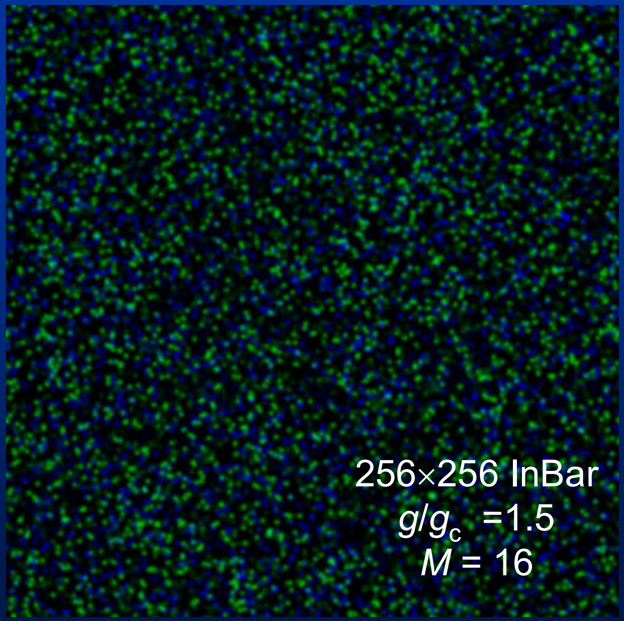
- as usual

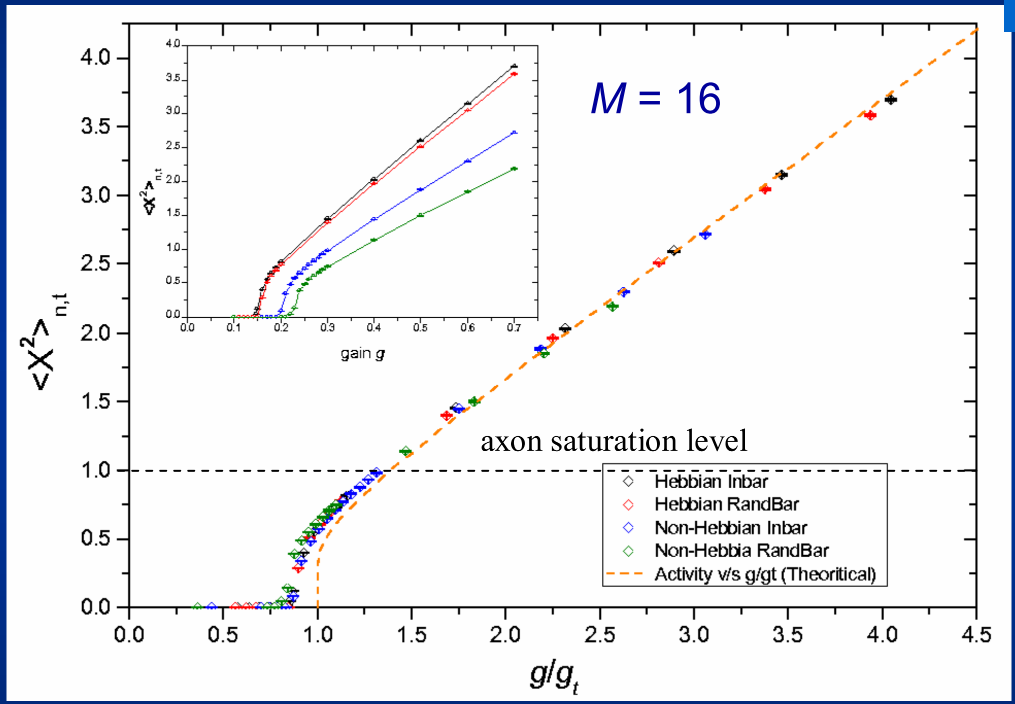
N-LAYER PERCEPTRON TRAINING



RECURRENT CROSSNETS: CHAOTIC DYNAMICS







Effective gain g

O. Turel, I. Muckra & K.L., 2003

GLOBAL REINFORCEMENT TRAINING (PLANS ONLY)

- Self-evolution:

  $x_j = x_j(t)$

- Inputs:

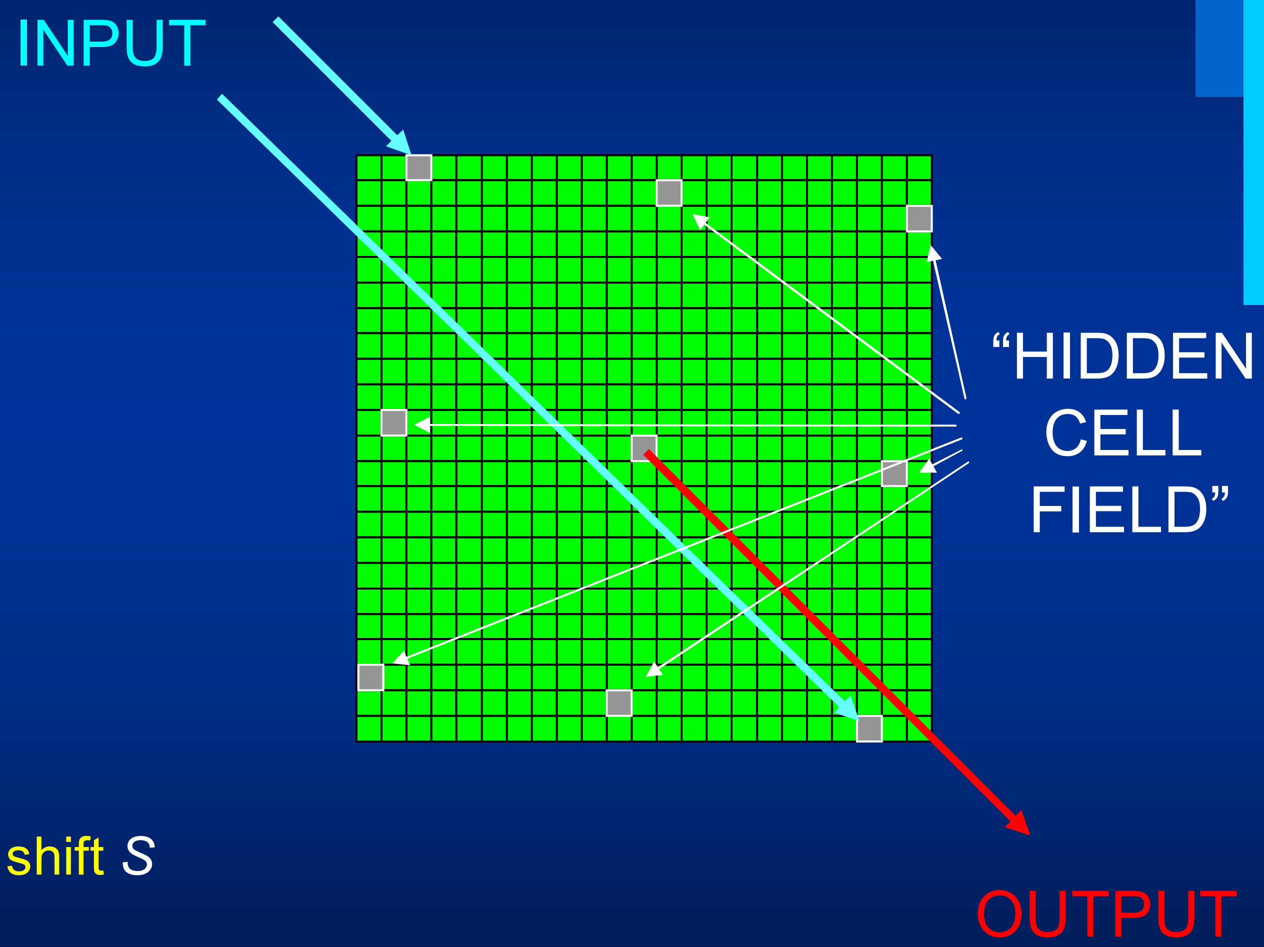
  $x_j(t) = x_j^i(t) + x_j^e(t)$

- Outputs:

  $x_k\ (t)$

- Training:

  change (quasi-) global shift S



CROSSNETS: ULTIMATE PERFORMANCE ESTIMATE

Synaptic plaquette footprint: $A_s = 256F^2$

for F = 2 nm: $A_s \sim 32\times32$ nm^2

Synapse density (for L = 5): $\sim 3\times10^{12}$ cm^{-2}

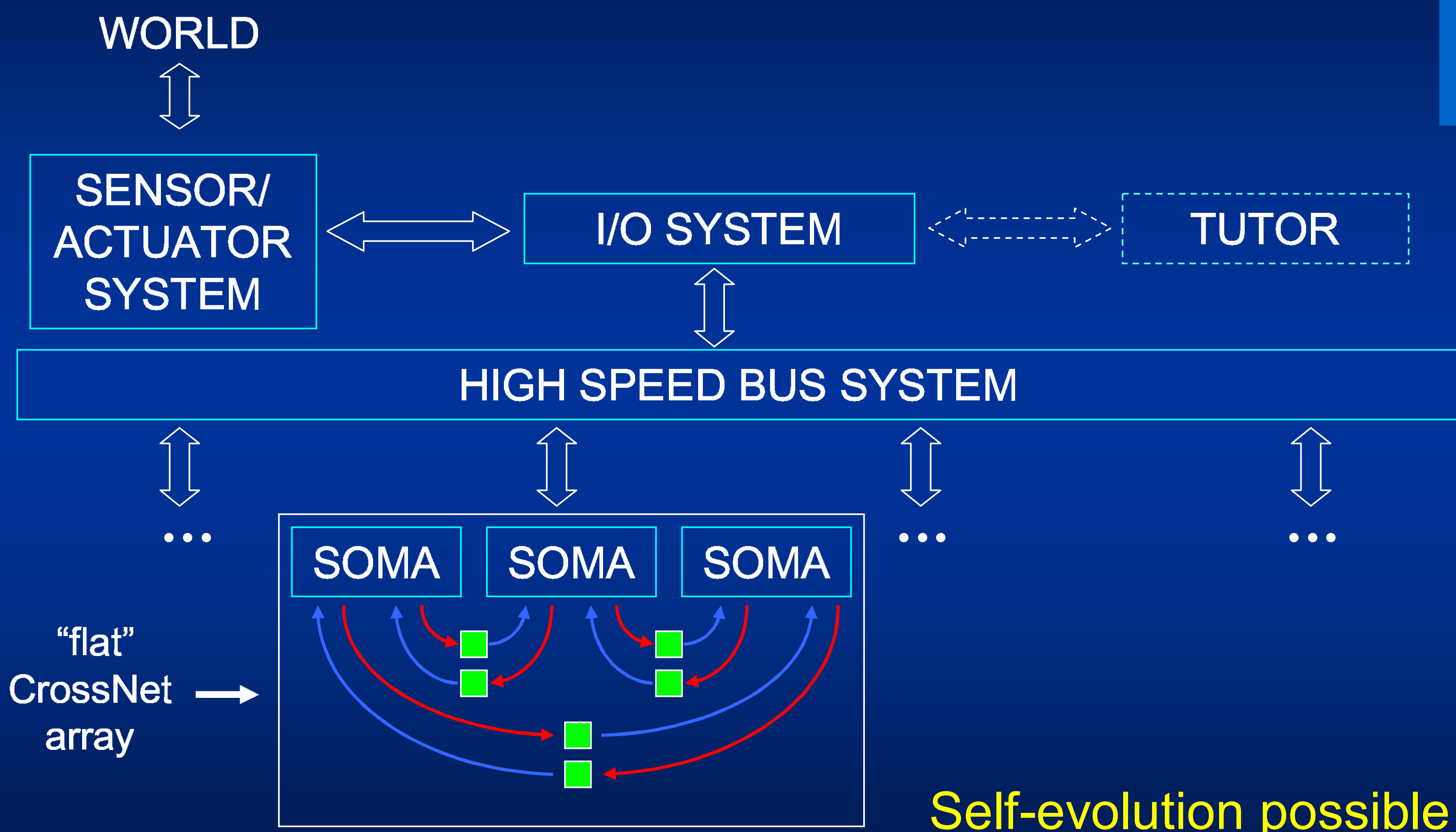
Cell density: for $4M = 10^4$: $\sim 1\times10^7$ cm^{-2} (close to bio)

Speed (intercell latency)**:** ~ 20 ns @ 100 W/cm^2 ($R \sim 10^{10}\ \Omega$)
or: ~ 2,000 ns @ 1 W/cm^2 ($R \sim 10^{12}\ \Omega$)
(cf. ~10 ms = 10,000,000 ns in bio)

Performance (for 100 W/cm^2)**:**
$\sim 3\times10^{12}$ cm^{-2} / 20 ns $\sim 10^{20}$ ops/cm^2-s
(cf. $\sim 10^{16}$ bits/cm^2-s for Prescott)

(Note: even here, $E \sim 10^{-18}$ J/op >> $k_BT \ln 2$)

CROSSNET SYSTEM HIERARCHY



Self-evolution possible ?

CONCLUSIONS

Fundamental power limitations:

- none
- even the perceived "limits" are irrelevant

Much more important: (quasi-) fundamental size scale:

- $F \sim |\lambda_B| \approx$ ◎$/(mE)^{1/2} \sim 1$ nm
- hardly feasible without molecular devices

CMOL:

- the future of microelectronics (?)

CrossNets:

- natural for CMOL
- ultimately high density @ high speed
- may reproduce any neural networks
  (at much higher speed and input vector size)
- (promise of) self-development:
  THE final frontier

THANK YOU!

Suggestions/comments to:

klikharev@notes.cc.sunysb.edu