An Overview of Advanced Interconnect Solutions

Prof. Krishna Saraswat
Department of Electrical Engineering
Stanford University
Stanford, CA 94305

Outline

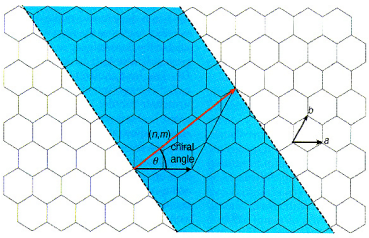
- Current Cu/low-k Paradigm: Challenges and Limitations
- Alternatives
  - Circuit Architectural/Combination Solutions
  - Technological Solutions
    - Air-gaps
    - 3-D technologies
    - Optical interconnects
    - Wireless
- Summary

Interconnect Performance Requirements

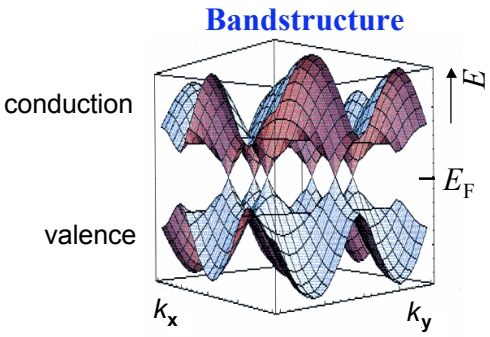
	Technology Generation 1.0 µm	0.1 µm
MOSFET Intrinsic Switching Delay	~ 10 ps	~1 ps
Interconnect Response Time (L_{int} = 1 mm)	~ 1 ps	~ 100ps
Clock Frequency	~ 30 MHz	~ 2-3.5 GHz
Supply Current (V_{dd} = 5.0, 1.0 V)	~ 2.5 A	~ 150 A
Maximum Number of Wiring Levels	3	7-8
Maximum Total Wire Length per Chip	~ 100 m	~ 5000 m
Chip Pad Count	~200	~ 2000, 4000

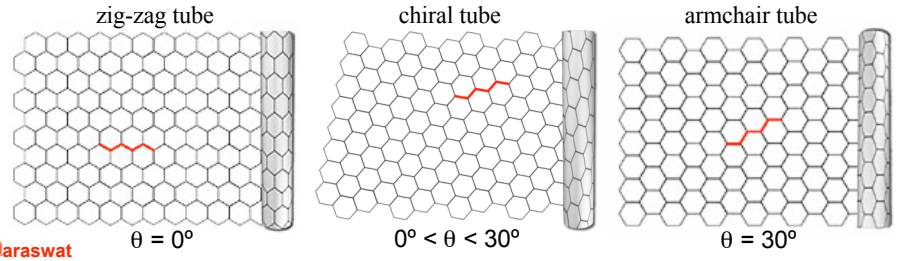
Carbon Nanotubes: Introduction

Rolled up graphene sheet

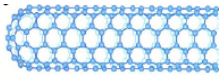


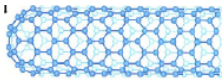
Bandstructure



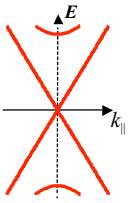


The one-dimensional subbands

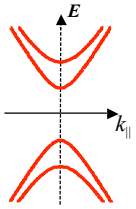




Metallic



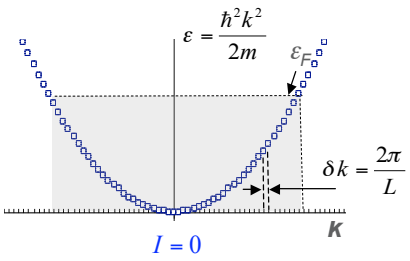
Semiconducting

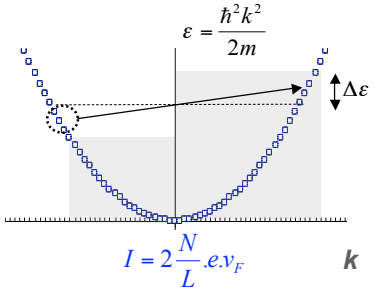


$E_g \approx 0.8/d$ (nm) *eV)*

- Chirality and diameter of the carbon nanotubes determine their properties
- Ideal carbon nanotubes don't have scattering => Ballistic transport
- Resistance of defect free nanotube-bundles, theoretically: h/(4e2n)
- Ballistic flow of electrons in metallic carbon nanotubes makes them potential candidates for nanoscale interconnections

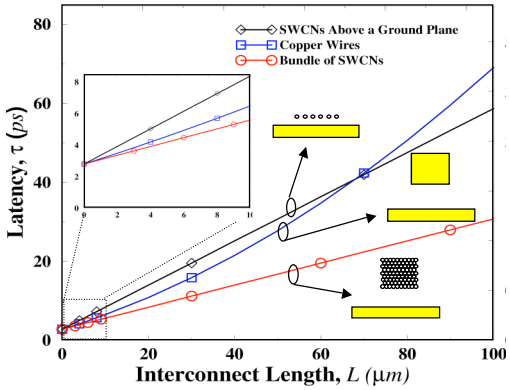
Kinetic Inductance for a 1-D System of Electrons





- In steady state equal number of electrons are moving in +ve and -ve k direction
- Nanotube is a 1D system with limited density of states
- With potential applied the carriers have to move to higher energy states resulting in increase in kinetic energy.
- Extra stored energy results in kinetic inductance
- Kinetic inductance of a carbon nanotube is ~16nH/μm, more than 4 orders of magnitude larger than its magnetic counterpart in a normal metal.

Alternatives for On-Chip Wires: Carbon Nanotubes



Naeemi, Meindl (IEDM 2004)

- Due to very large kinetic inductance, propagation speed in a carbon nanotube above a ground plane is more than 100 times slower than light in free space.
- Bundles of carbon nanotubes should be used for interconnect applications to avoid very slow signal propagation.
- Capacitance is similar to Cu-low-k. Will they have any power advantage over Cu/low-k?

Carbon Nanotubes with Finite Electron Mean Free Path

- Even initially perfect nanotubes become disordered once they are physisorbed on a surface.
- Due to interference between incident and scattered electron waves, nanotube resistance increases exponentially with length:

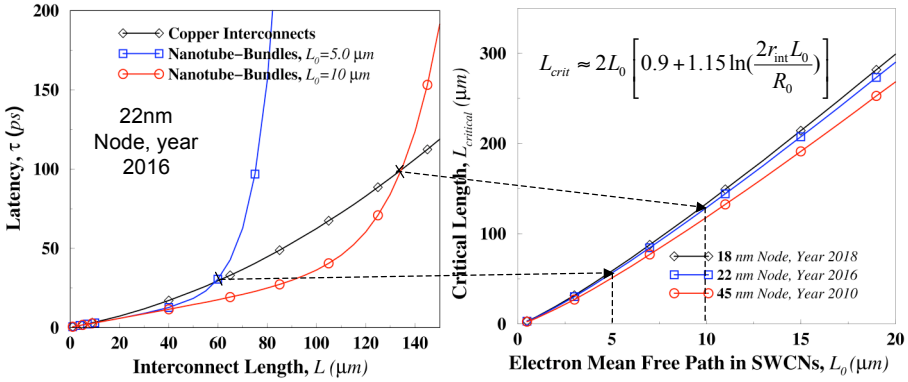
$$R = R_0 e^{\frac{L}{2L_0}}$$

R_0: Resistance of defect free nanotube-bundles, theoretically: h/(4e2n)

L_0: Electron mean free path *L: Length*

Naeemi, Meindl (IEDM 2004)

Nanotube-Bundles versus Cu Interconnects

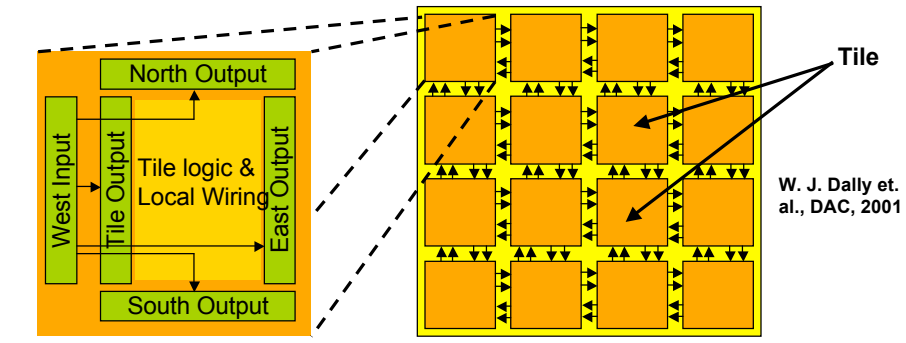


Naeemi, Meindl (IEDM 2004)

As a rule of thumb, nanotube-bundles should never be used for lengths larger than 10 times electron mean free path.

Key Challenge: Low thermal budget controlled growth

On-chip Network Routers



(West input shown as an example)

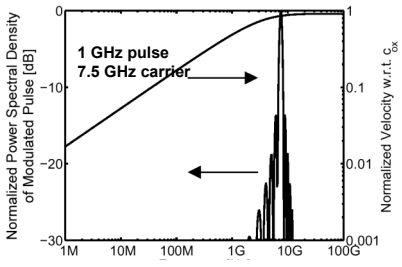
Partition of chip into tiles & network logic

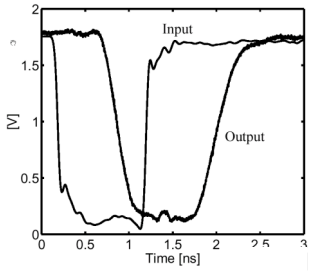
- Modular machine
  - Lots of potential compute units, w/ memory
  - On-chip network routers

Alternate Solutions in Copper Domain

Near Speed of light on-chip electrical interconnects

(R. T. Chang et. al. 2002 Symposium on VLSI circuits)

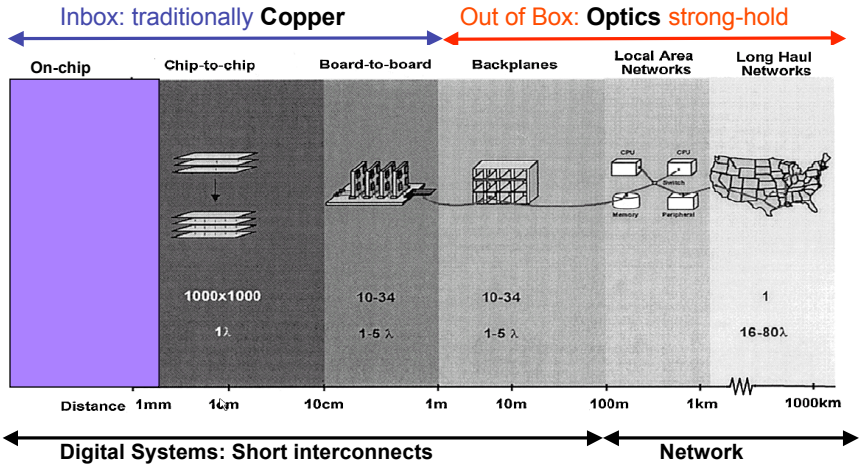




- High frequencies in a wire (LC domain) travel close to speed of light
- Modulate electrical signal to higher carrier frequencies (baseband to RF)
- But loss is high at higher frequencies (2 losses: due to conductor resistance and dielectric loss)
- Requires good low-loss transmission lines

- Near speed of light delay (10 times faster repeaters)
- Power comparable to repeater (at least at this node)
- Possible Issues
  - Requires larger area (well designed wires)
  - Some overhead due to fancy modulator/detector
  - scalability: higher freq=> higher carrier freq. => larger loss

Interconnect Hierarchy: Bird's Eye View



Y. Li et. al, Proceedings of the IEEE,vol. 88, no. 6, June 2000.

What will be the energy cost, per bit processed?

1. Logic — energy cost ~40kT per bit processed
2. Storage — energy cost ~40kT per bit processed
3. Communications — currently >100,000kT per bit processed (~1fJ)

Will this change in the future with scaling?

There are many ways to do logic and memory.
But there are not many ways to communicate:

1. Electrons on a metallic interconnect
2. Optical interconnect

Source: Eli Yablonovitch, SRC Workshop, Asheville, 2005

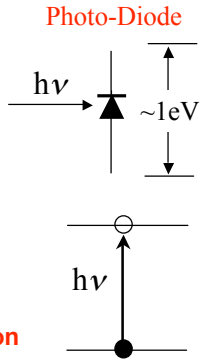
Impedance Matching Crisis

What is the impedance of a wire?

C_{wire} ~ femto-Farads $\quad Z_{wire} = 1/j\omega C_{wire}$

The natural capacitance of a nano-scale device is much smaller:

C_{nano} <10^{-18} Farads $\quad Z_{nano} = 1/j\omega C_{nano}$



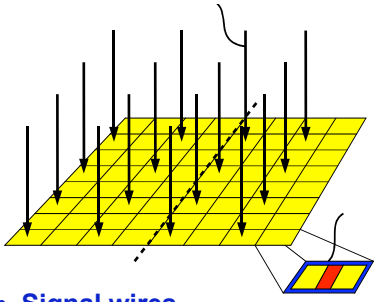
- The natural voltage range for wired communication is rather low: The wire wants many electrons at few mVolt each
- The natural voltage range for the nano-device is the voltage required for switching the electron out of the potential well: ~1Volt
- The nano-device wants one electron at 1 Volt
- An impedance matching device is needed at the Nano-scale.

Optoelectronic devices may provide the best solution

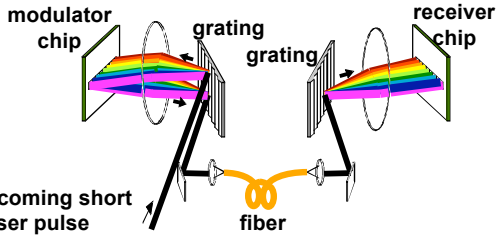
Sources: (1) David Miller, Optics Letters, 1989, (2) Eli Yablonovitch SRC Workshop, Asheville, 2005

Can Optical Interconnects help?

On-Chip Optical Interconnects

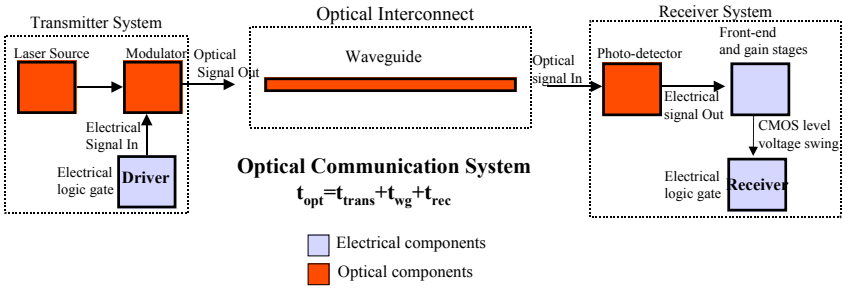


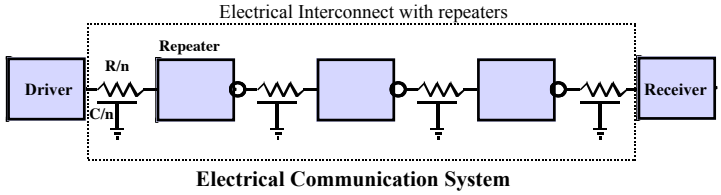
Chip-to-chip Optical Interconnects



- Signal wires
  - Reduce delay
  - Increase bandwidth
- Clock distribution
  - Reduce jitter and skew
- I/O with very high bandwidth
- Reduce power

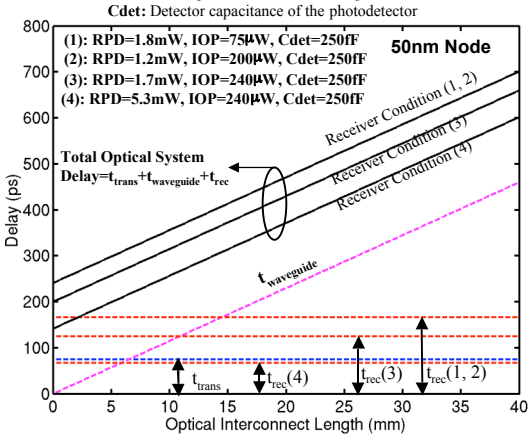
Optical Vs. Electrical Wires





Optical System Delay

RPD: Receiver Power Dissipation; **IOP:**Incident Optical Power at the receiver
Cdet: Detector capacitance of the photodetector

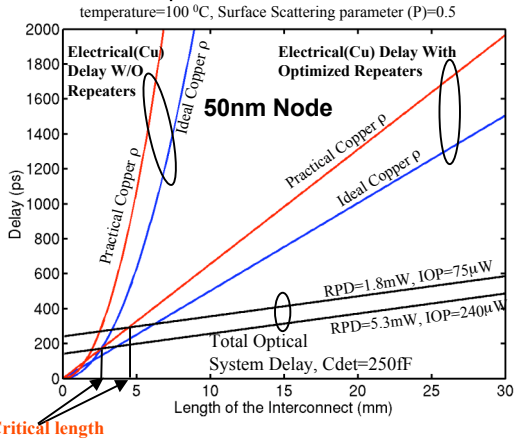


- **Transmitter delay** using buffer chain
- **Waveguide delay**: assumed simplistic: 11.5 ps/mm
- **Receiver Delay**: different conditions
- Waveguide delay dominates after 15mm

Global Signaling Delay: Optical Vs. Electrical Wires

- Optical Interconnects are faster than repeated wires beyond a length well within chip size
- However for Signaling both delay and power are important
- 1.8 mW is approximately power dissipated by a repeated chip edge long wire

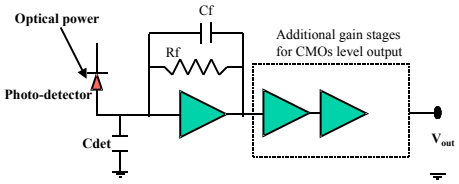
IOP: Incident Optical Power at the receiver
Practical Cu ρ: ALD Barrier, Barrier Thickness=10nm, temperature=100 ^{0}C, Surface Scattering parameter (P)=0.5



Critical length above which optical System is faster than even the electrical (Cu) repeated wires

Optical Receiver Power Dissipation

A simple receiver analysis for studying scaling



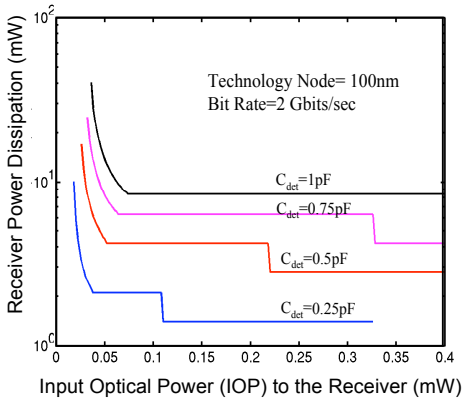
Constraints

- Bandwidth
- SNR
- Supply swing at output

→ Front-end width & feedback resistance

→ # of gain stages

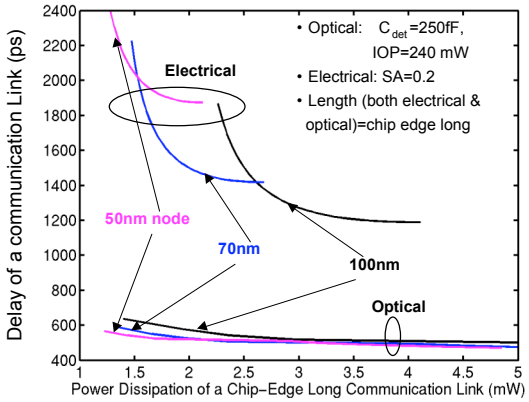
- Short channel effects for transistors incorporated
- Transistor specs from ITRS
- Receiver power dissipation (Static)



Lower detector capacitance and higher IOP for low receiver power dissipation

Kapur and Saraswat, IEEE IITC, June 2002

On Chip Optical Vs. Electrical Wires: Delay & Power Scaling



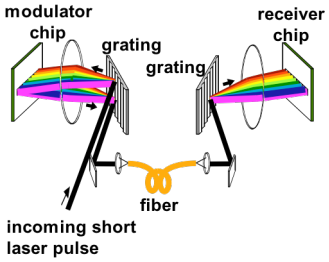
Kapur and Saraswat, IEEE IITC, June 2002

- **Scaling: delay advantage increases for optics, power advantage diminishes**
- **Good for long global wires whose number is not large**
- **Even power advantage exists if switching activity (SA) is higher**
- **Power can be reduced with better detectors**

Off-Chip Interconnect Requirements

Requirements

- High Bandwidth or bit rate
- Low latency: some applications
- Acceptable data fidelity
- Low power

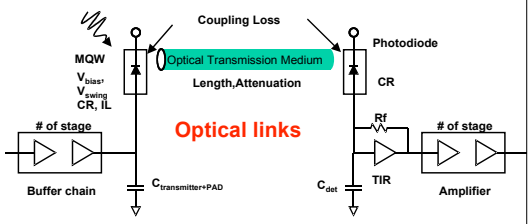


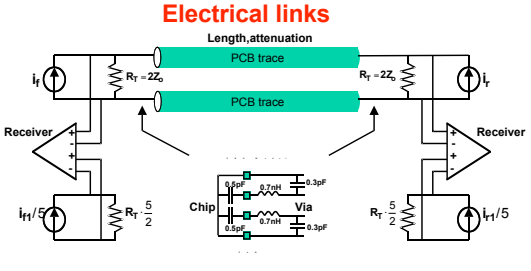
Different beast than on-chip interconnects

- Larger widths (~100-200μm) and longer distances (1cm-1m)
- Better defined return paths (controlled L, R and C)
- Wires are mostly "low loss" LC
- No silicon => no repeaters
- Equalization

Off-Chip Electrical Vs. Optical Links

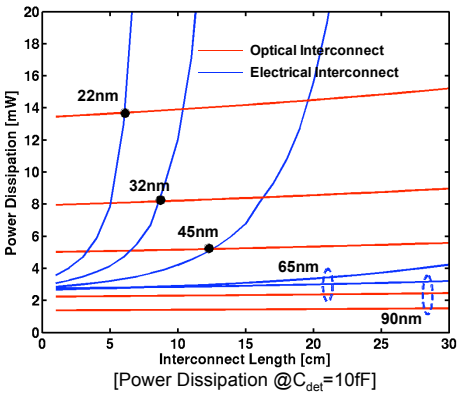
- Share many components with electrical links
  - Transmitter to drive the laser or modulator
  - A CDR to recover timing
- Also have some new components
  - Laser/modulator, photo-diode
  - Need driver for the modulator/laser, TIA for photo-diode
  - Optical channel
    - Connectors, optical wire (board, fiber, free space?)



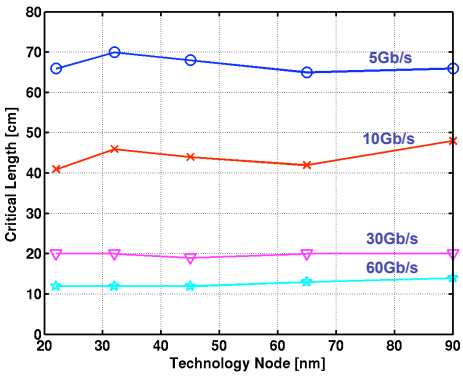


Cho, Kapur and Saraswat, IITC 2004, JLT 2004

Comparison Between Electrical and Optical Interconnects for Off-Chip Communications



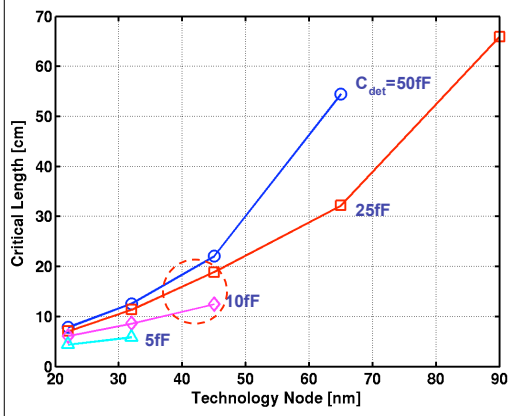
[Power Dissipation @C_{det}=10fF]



- **Beyond certain length optical I/O is more power efficient**
- **Critical length decreases at higher bit rate**
- **Beyond 32nm Technology node critical length < 10cm**

Cho, Kapur and Saraswat, IEEE IITC, June 2005

Power Comparison: Technology Scaling

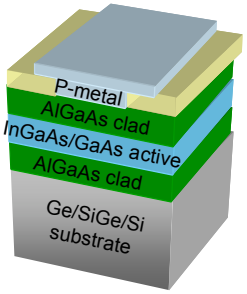


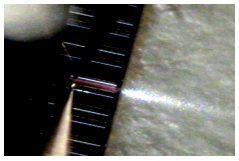
- **Optical interconnects will become more and more favorable in the future by showing a rapid reduction in critical length with technology scaling**
- **Importance of lower detector capacitance in facilitating the insertion of optical interconnects**

Cho, Kapur and Saraswat, IEEE IITC, June 2005

Technology for Optical Interconnects on Silicon : Monolithic Integration of Lasers

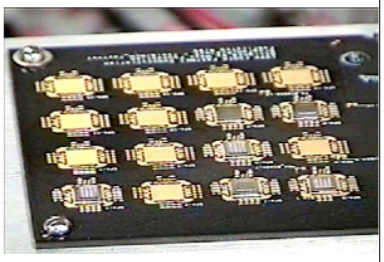
GaAs/AlGaAs QW laser on Ge/GeSi/Si





Fitzgerald, MIT

Arrays of surface emitting lasers

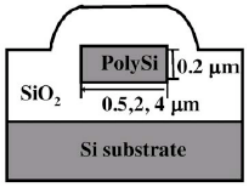


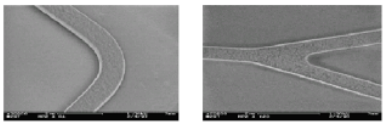
(Haney et al.)

- **InGaAs/AlGaAs QW laser on Ge/SiGe/Si operates cw at λ = 858nm**
- **Critical issues are power consumption, stability and life time**

Technology for Optical Interconnects on Silicon: Optical Transmission Media

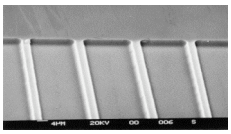
Waveguides



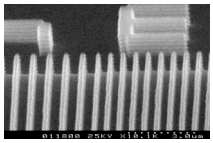


(Kimmerlingr, MIT)

Arrayed Waveguide Grating

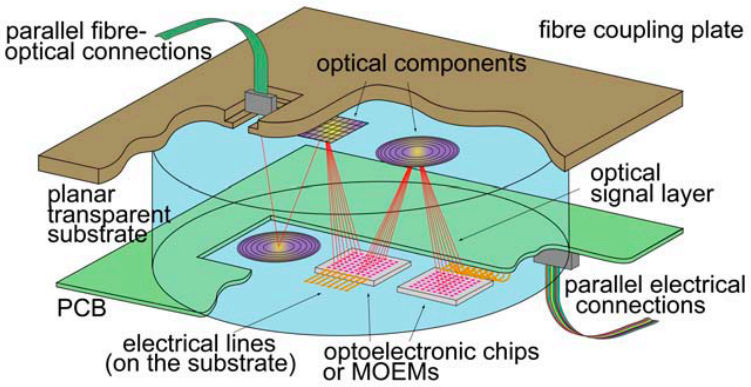


Directional Coupler



(Jalali, UCLA)

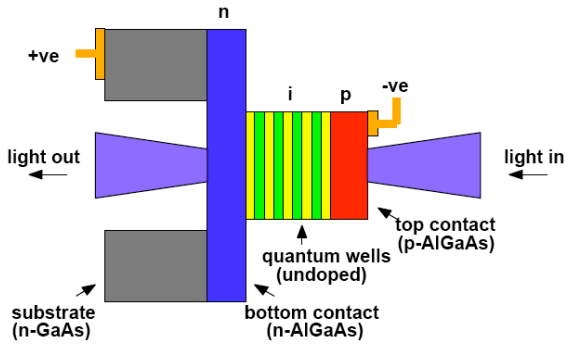
Technology for Optical Interconnects on Si: Free Space Transmission



Critical issue is packaging technology

(Jurgen Jahns, Fernuniversität Hagen, Germany)

Technology for Optical Interconnects on Si: Quantum Well Modulator

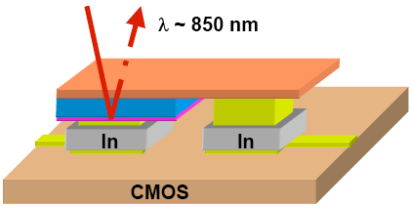


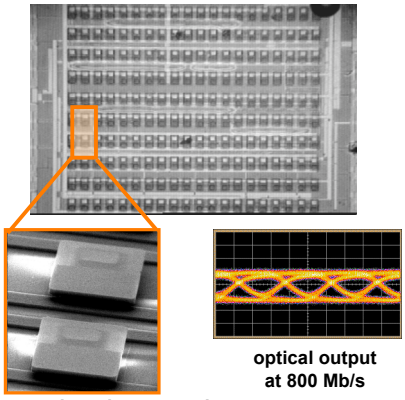
(Miller, Harris, Stanford)

Application of E-field changes optical transmission through quantum well modulators

Flip-Chip Bonding of GaAs Modulators and Photodiodes to Silicon CMOS

(Miller Group, Stanford)





optical output at 800 Mb/s

- Flip-chip bonding enables post-processing integration:
  - Mature CMOS process + optimized optoelectronics
  - Dense 2D arrays bonded in parallel with high yield
  - Low noise and high speeds due to reduced device capacitance

Germanium – A Prospective Material

Small optical bandgap

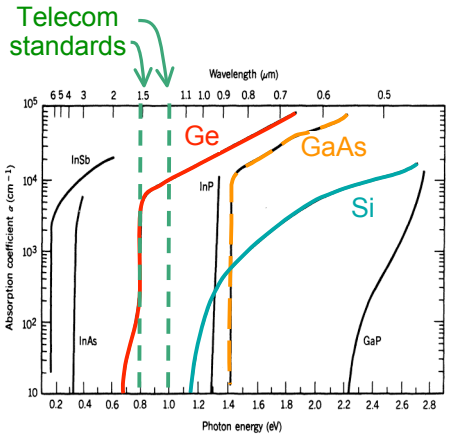
- ✓ λ transparent to Si, no extra circuit noise
- ✓ Direct gap below 1.53μm ⇒ Broadens λ spectrum for opto-electronic integration to enhance CMOS functionality
- × Higher thermal leakage

High carrier mobilities

- ✓ Short detector transit time

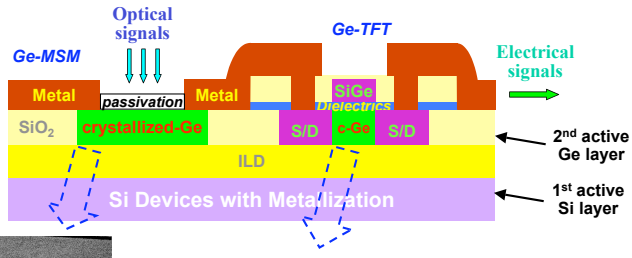
Compatible with Si IC

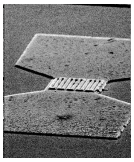
- ✓ Low processing temp



(Stillman *et al.*, *IEEE TED*, 31, p.1643, 1984)

3D Integration of Ge Optoelectronic Devices on Si





Ge Photodetector

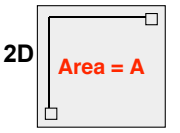
Ge Transistor

• Piecewise technologies ready for the monolithic receiver integration in the near future

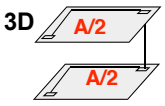
- **Employ recrystallization or layer transfer technique for Ge on Si**
- **Integration of optical receiver in the upper active (Ge) layer**
  ⇒ On-chip optical clock distribution in 3D-ICs

Saraswat, Stanford Univ.

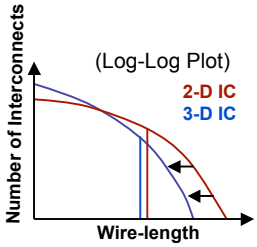
3-D Integration: Motivation

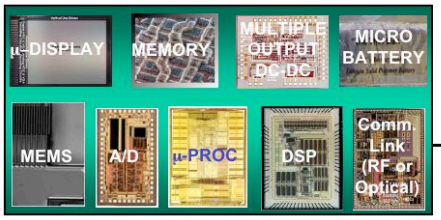


Very Long Wire

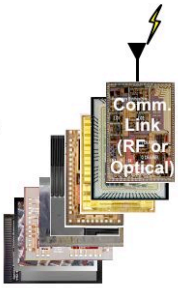


Shorter Wire





2-D System



3-D System

- **Integration of heterogeneous technologies possible, e.g., memory & logic, optical I/O**
- **Reduce Chip footprint**
- **Replace long horizontal wires by short vertical wires**
- **Interconnect length ⇓ and therefore R, L, C ⇓**
  - **– Delay reduction**
  - **–Power reduction**

Importance of Form Factor

GPS Hand-Held Receiver

STAR TREK COMMUNICATOR

Cell phone

PDA

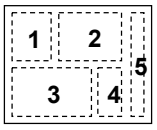
MP3 player

Digital Camera

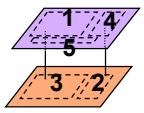
Wire-length Distribution of 3-D IC

Microprocessor Example

Number of Logic Gates	93.6 million
Number of Memory Devices	86.4 million
Minimum Feature Size	50 nm
Number of wiring levels,	9
Metal Resistivity, Copper	1.673e-6 Ω-cm
Dielectric K, Polymer	ε_r = 2.5

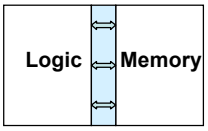


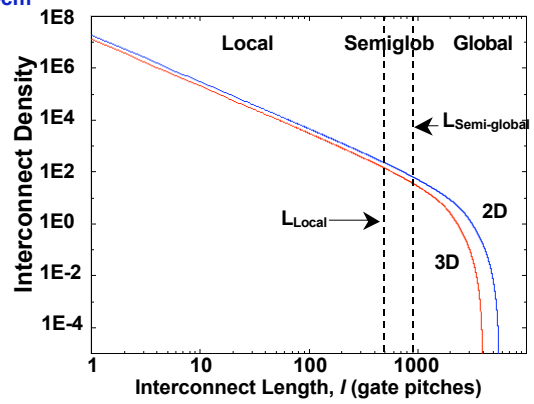
Single Layer



2 Layers

A significant fraction of horizontal wires replaced by vertical wires



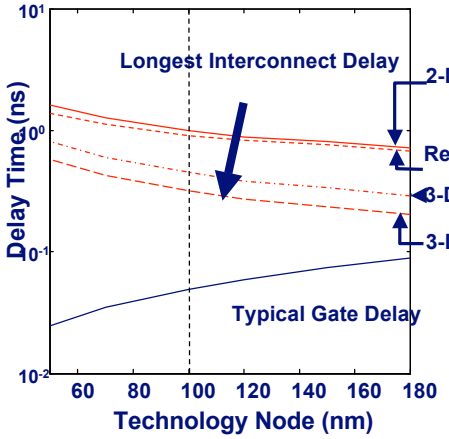


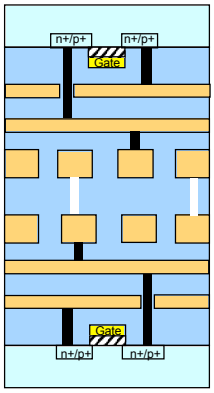
Souri, Banerjee, Mehrotra and Saraswat, ACM Design Automation Conf., June 2000.

Delay of scaled 3D ICs with multiple Si layers

- Summary of delay results for ITRS technology nodes
  - Moving repeaters to upper active layers reduces delay by 9%
  - 3D (2 Si layers) shows delay reduction to 62%
  - Increasing metal layers reduces delay further by 25%

3D can alleviate interconnects limits





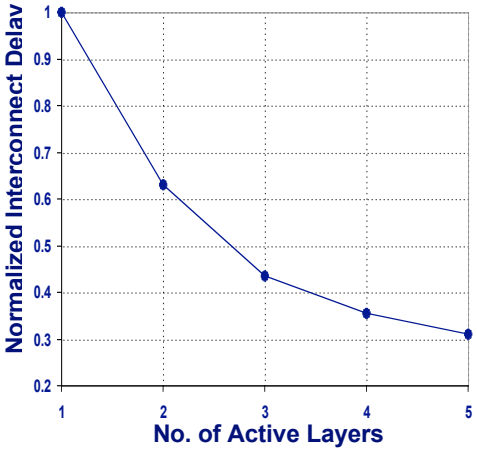
Banerjee, Souri and Saraswat, Proc. IEEE, May 2001

Effect of additional active layers

ITRS 50 nm Technology Node

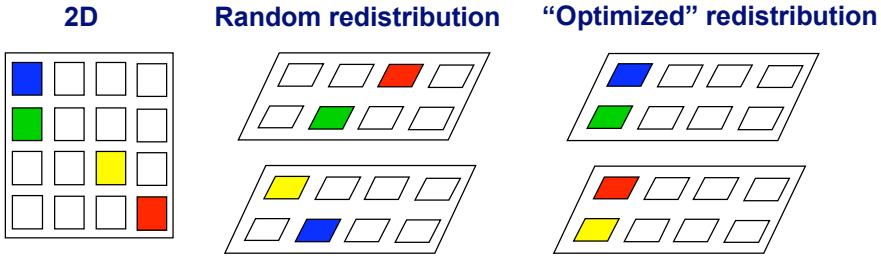
N_{Logic}	$769X10^6$
N_{Memory}	$6284X10^6$
λ	50 nm
Metal levels	9
ρ_{Cu}	$1.67X10^{-6}$ Ω-cm
ε_r	1.5

- Boundary condition: all footprints increased to 2D
- Limited returns with increased active layers due to wire routing to higher tiers
- 2 active layers show largest delay improvement



S. Souri, PhD Thesis, Stanford Univ., 2003

Random vs. Optimized Redistribution of Logic



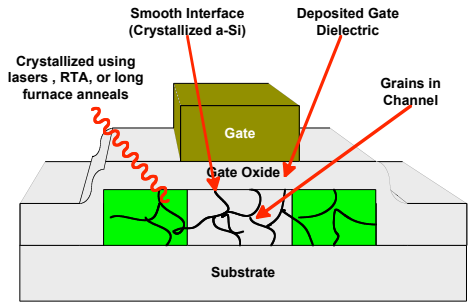
- Random redistribution: indiscriminate replacement of short/long wires with VILICs for generalized analysis
- Optimal redistribution: only critical paths replaced. Dependent on particular IC design
- Analysis so far conservative due to randomizing redistribution. With optimized redistribution gains would be higher

Technology to Fabricate 3D ICs

- Bonding
  - Metal-metal
  - Dielectric glue
- Epitaxial growth
- Crystallization
  - Laser melting and crystallization
  - Seeded crystallization
  - Liquid phase crystallization
- Self assembled molecular devices
  - Si and Ge nanowires
  - Carbon nanotubes
  - Organic semiconductors

Statistical Variations in Poly-TFT Properties

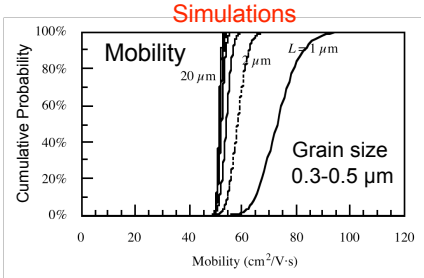
Conventional Poly-TFT



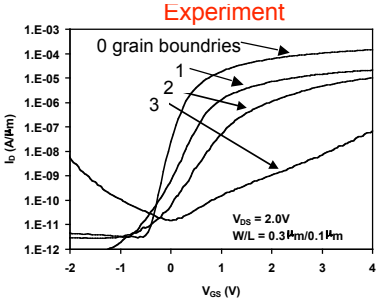
- As channel length ⇒ grain size, statistical variation increases
- Poly-Si TFTs not suitable for 3-D ICs.
- Grain boundaries need to be eliminated

Wang and Saraswat, IEEE Trans. Electron Dev., May 2000.

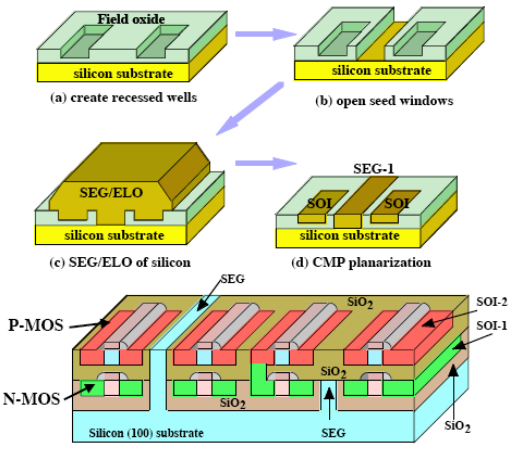
Simulations



Experiment



3D Approaches: Epitaxial Lateral Overgrowth

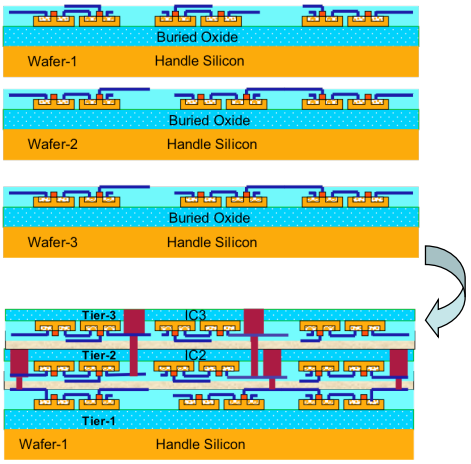


Key Challenges:

- High temperature compatible, low resistance interconnect technology
- Low temperature epitaxy (≤ 400 °C)

Source:
G. Neudeck, Purdue Univ.

3D Approaches: Cu Contact Wafer Bonding

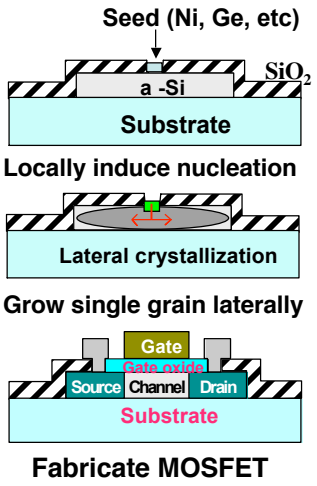


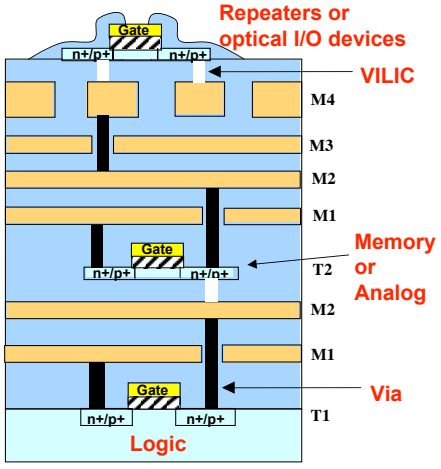
Wafer Bonding (MIT, IBM, and many more)

Key Challenges: Precise alignment of wafers/dies

3D Approaches: Crystallization of α-Si

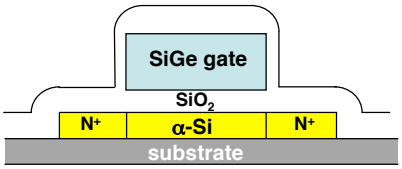
Fabricate devices at any metal level if thermal budget is low



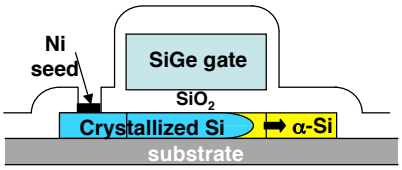


Key Challenge: Low temperature defect free crystallization

Ni Seeded Lateral Crystallization

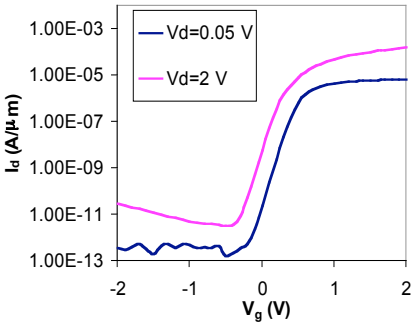


Fabricate amorphous device



Use seeding for simultaneous crystallization and dopant activation

NMOS L = 0.1μm, W = 0.3 μm

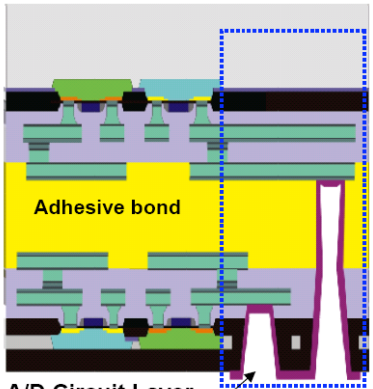


- Low thermal budget (T_{max} < 500°C)
- Devices on top of a metal line
  - MOSFETs
  - Optical detectors

Joshi & Saraswat, IEEE Trans. Electron Dev. 2003

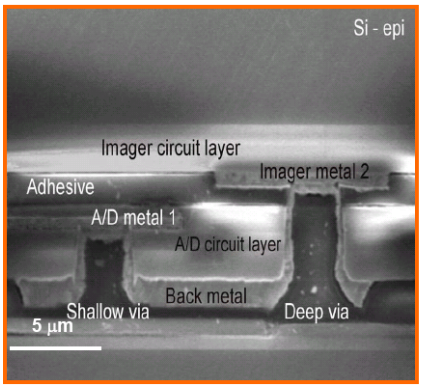
3D Image Sensor

Imager Circuit Layer



A/D Circuit Layer

2-step connection between layers

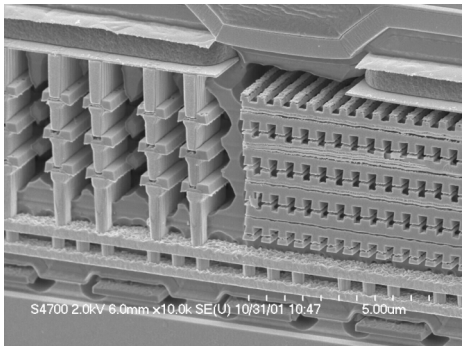


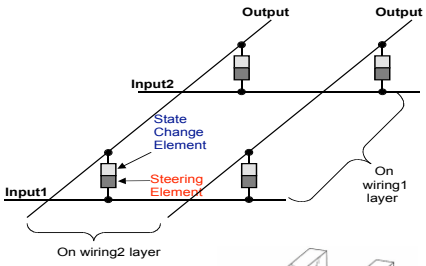
3-D Interconnect Cross Sectional SEM

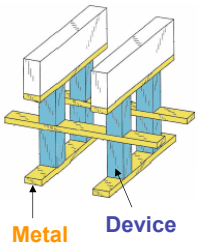
First demo of 64x64 APS stacked on fully parallel ADC circuit
4000 vertical interconnects

Source: Crais Casey, MIT Lincoln Laboratory

Matrix Memory Array (8 layers@0.25μm)





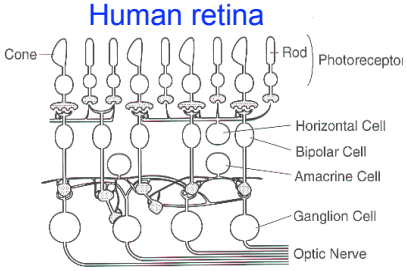


- Matrix memory is a 3D array of diode-antifuse cells.
- Diodes are inserted into the circuit at desired locations by blowing an antifuse.
- Matrix memory is write-once and archival.

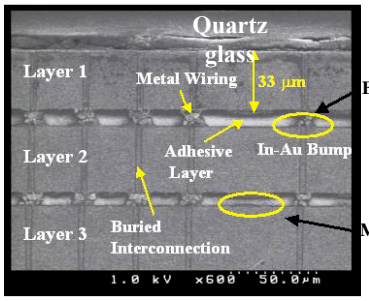
Source: T. Lee, Matrix Corp.

3D Artificial Retina Chip

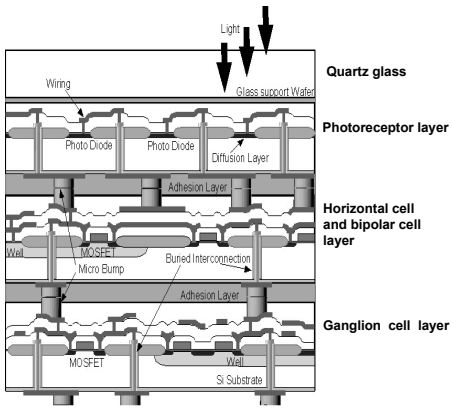
Human retina



3D retina chip

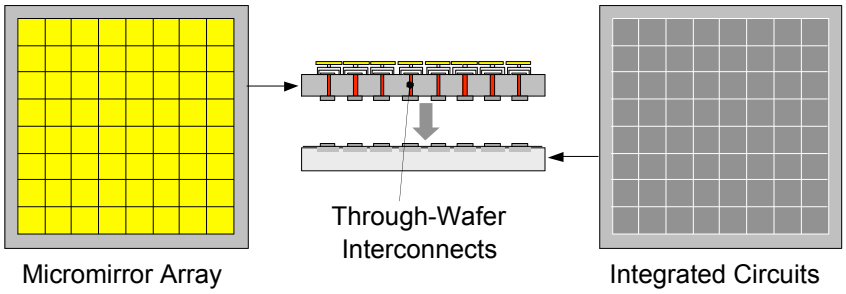


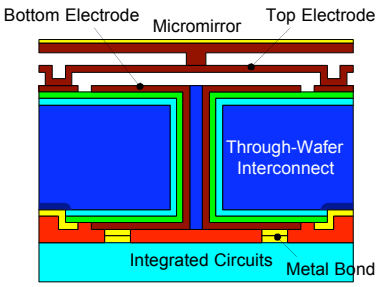
3D retina chip (schematic)



Courtesy: M. Koyanagi, (Tohoku University)

3D Integration of Photonic Devices with Electronics

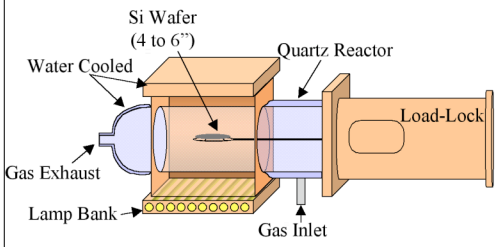




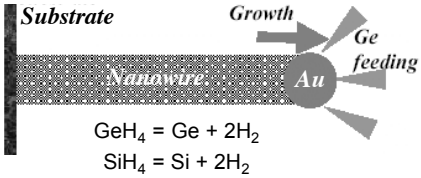
Khuri-Yakub, Stanford University

Ge nanowires synthesized by low temperature CVD

Cold-Wall NW Growth Reactor

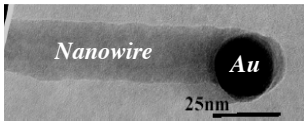


- **Vapor-Liquid Solid (VLS) mechanism uses super saturated liquid eutectic to locally deposit Si.**
- **Growth mechanism: wires grow as the catalytic end moves**
- **Size of the wire is decides by the size of the nano particle**
- **10-20 nm single crystal Si or Ge nanowires without a crystalline substrate**
- **CVD temperature: 275 - 400°C**



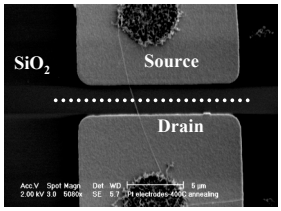
$GeH_4 = Ge + 2H_2$

$SiH_4 = Si + 2H_2$

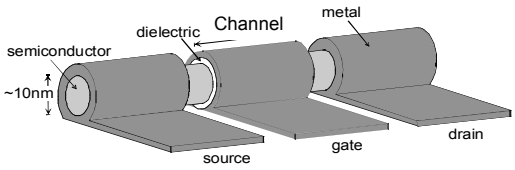


Ge Nanowire FET with High K gate dielectric

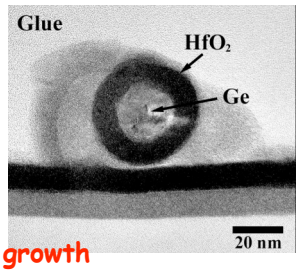
Top view SEM



Ge NW MOSFET



ALD HfO_2 Coating of Ge NW

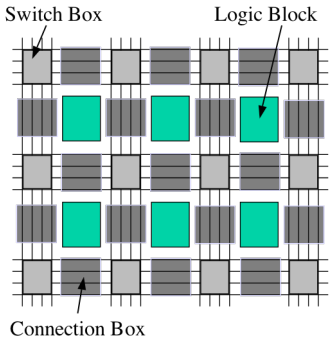


- **Initial fabricated transistor shows great promise μ_p ~ 500**
- **No crystalline seed needed => 3D integrable technology**

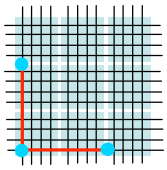
Key Challenge: Controlled growth

Ref: Wang, Wang, Javey, Tu, Dai, Kim, McIntyre, Krishnamohan, and Saraswat, Appl Phys. Lett, 22 Sept. 2003

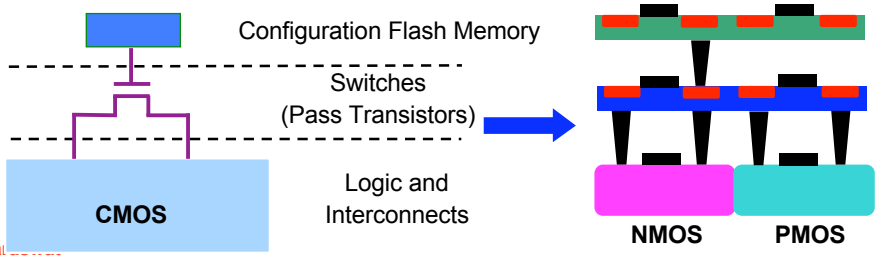
3D Monolithically Stacked FPGA



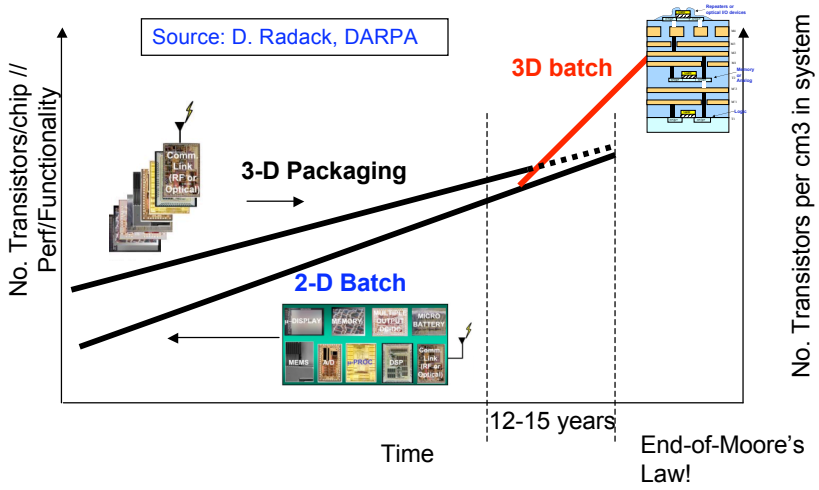
Move programming on top of logic



- Down to 0.14 area
- Shorter interconnect
  - Higher performance
  - Lower power



Motivation: Integration Density



The Best Integrators of Electronic Devices Will Own the Heart of Every System – We have <15 Years to Figure it out

Wireless Clock Distribution Network

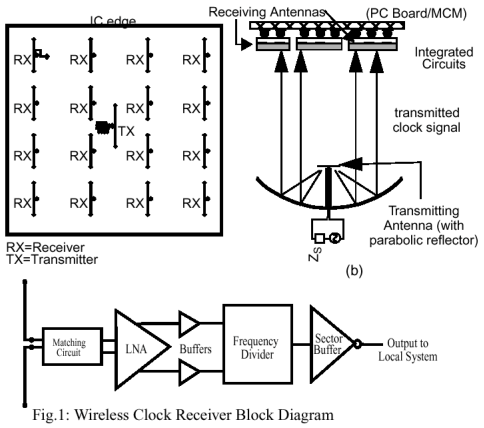


Fig.1: Wireless Clock Receiver Block Diagram

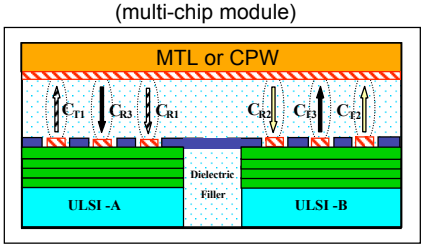
- Clock signal transmitted as a monotone cosine wave
- No signal dispersion (**1st advantage with wireless over conventional clocking**)
- Received signal amplified, divided and buffered

RF-Interconnect System Concept

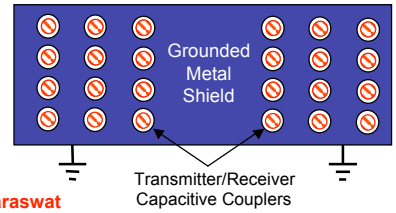
Courtesy: F. Chang, UCLA

Chip to Chip Interconnect using
Wireless RF through capacitive coupling

Miniature LAN in MCM
(multi-chip module)



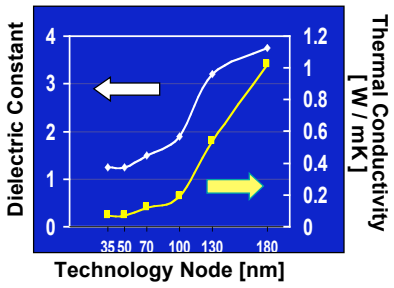
Top View

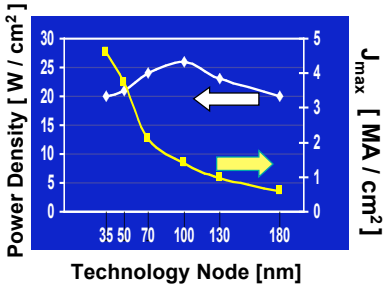


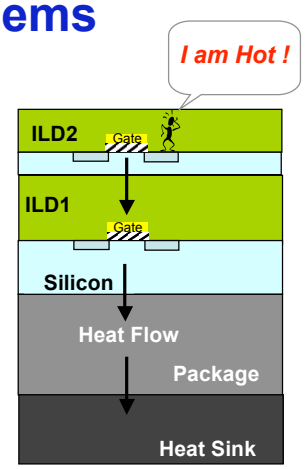
- Low loss, dispersion-free, ultra-high data rate (**100Gbps/channel** & **20Tbps/chip**)
- Multi-I/Os per channel, simultaneous communications via shared MTL or CPW using FDMA/CDMA multiple access algorithms
- Reconfigurable network for on-line system-level rewiring (**Architecture reconfigurable on-the-fly**)
- Coherent chip-module combined interconnect scheme, compatible with mainstream ULSI, MCM or surface-mount PCB

Key Challenges
•Signal to noise ratio?
•Power consumed?
•Interference?

Thermal problems





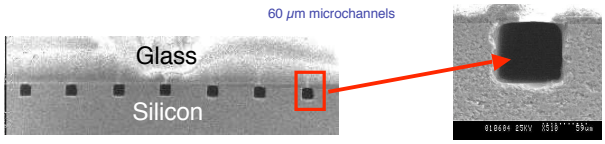


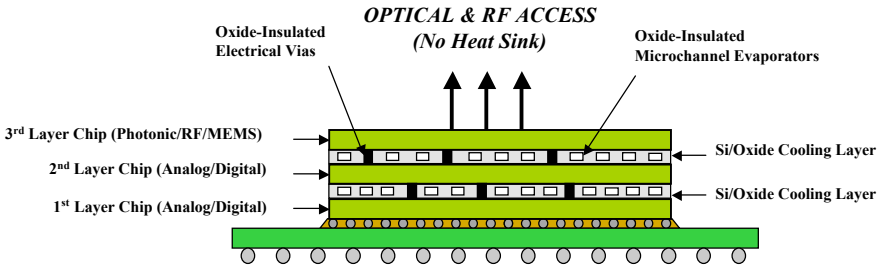
- **Thermal conductivity of low-k insulators is poor**
- **Thermal impedance increases**

Chiang and Saraswat, VLSI Symp., June 2001

Integrated Microchannel Cooling for 3D

Prof. Goodson. (Stanford Univ.)





Microchannel: Remove heat from targeted areas by means of convective boiling

Deliverable: 200 W from 2 x 2 cm Closed-Loop System with Electrokinetic Pump

Summary

- **On-Chip Interconnects**
  - **Conventional Interconnects: Challenges and Limitations**
    - Importance of looking at all figures of merit simultaneously to decide dimensions
    - Dimensional scaling challenges/limitations
      - Cu effective ρ rises dramatically at all tiers: technology effects
    - Realistic interconnect delay rises significantly w.r.t clock period even with repeaters
    - Interconnect power also rises in future
  - **Alternatives in Future**
    - Copper domain solutions, such as air-gaps and circuit/architectural combination will take us a long ways
    - Carbon nanotubes promising for local interconnects
    - 3D technology promising
    - Optical Interconnects promising but needs extensive R&D
    - Wireless: an intriguing technology, however, with several key challenges
- **Off-Chip Interconnects**
  - Optics may be viable around >35 cm either from bit rate or power perspective
- **Thermal dissipation in 3-D ICs may require innovative packaging solutions.**