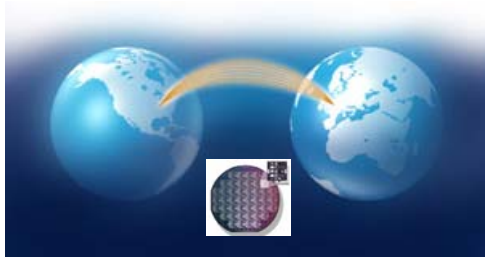
Package level Interconnect Options



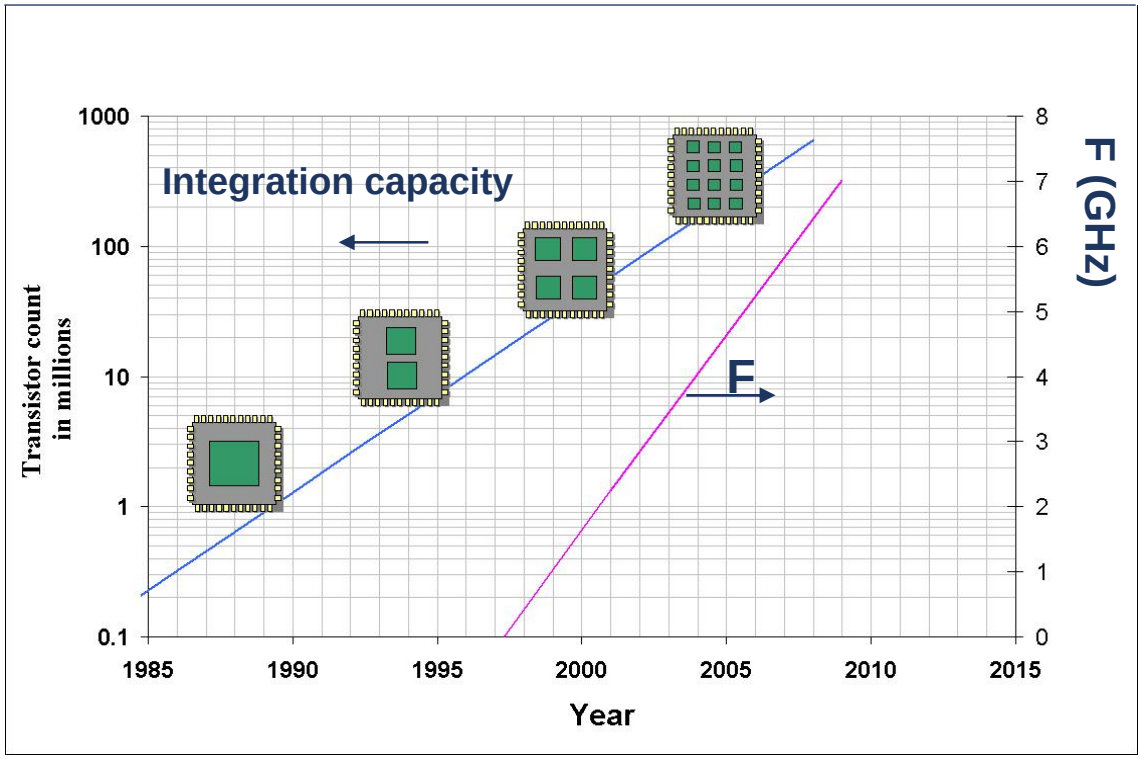
J.Balachandran,S.Brebels,G.Carchon,
W.De Raedt, B.Nauwelaers,E.Beyne

SLIP 2005

April 2 –3

Sanfrancisco,USA

Challenges in Nanometer Era



- Integration capacity ~ *billion* transistors
- Operating frequencies ~ μwave (*GHz*)
- Communication bottleneck
  - Interconnect delay far exceeds scaled transistor delay
  - Latency spanning multiple clock cycles
  - Power consumption

Computer architecture is all about Interconnect

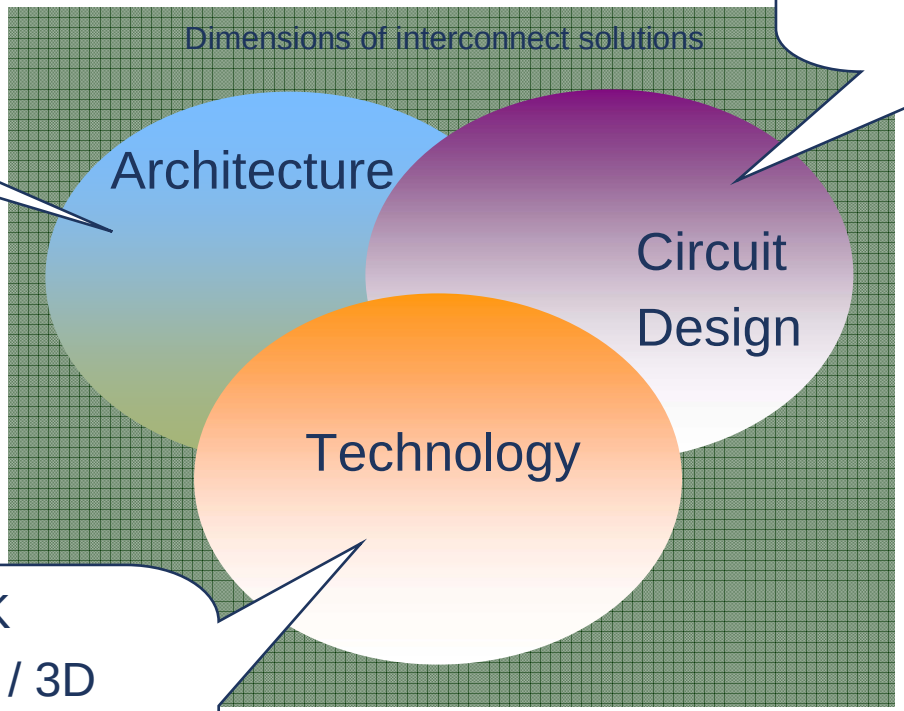
- Prof. William J. Dally, (Stanford university)

@ HPCA Panel February 4, 2002

Breaking the Interconnect bottleneck

Clustering
GALS/NoC
Wired CDMA

Repeaters
Differential signaling
Equalization

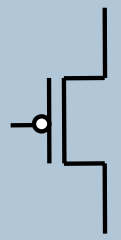


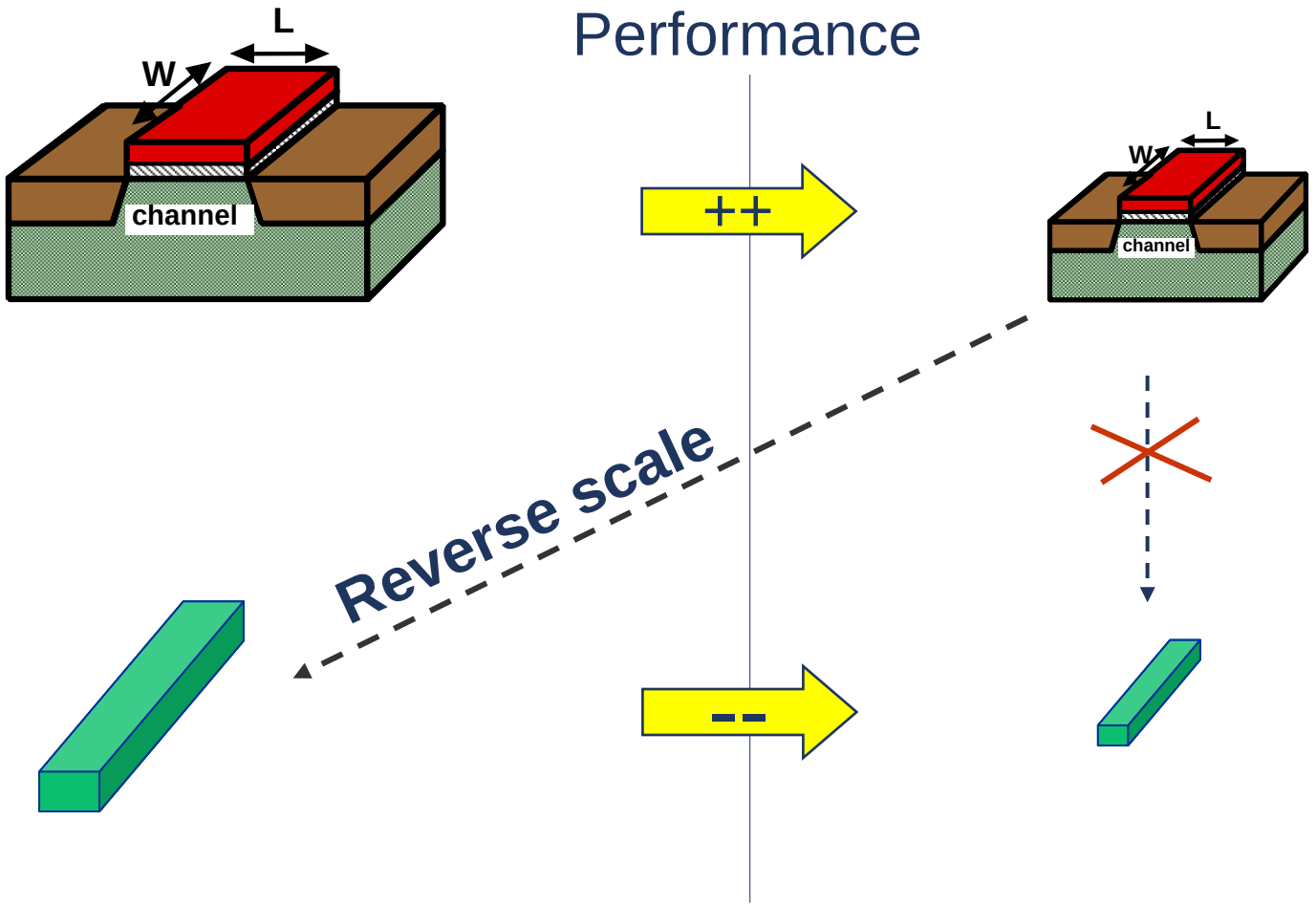
Cu/Low-K
X-routing / 3D integration
Optical
Reverse scaling

Outline

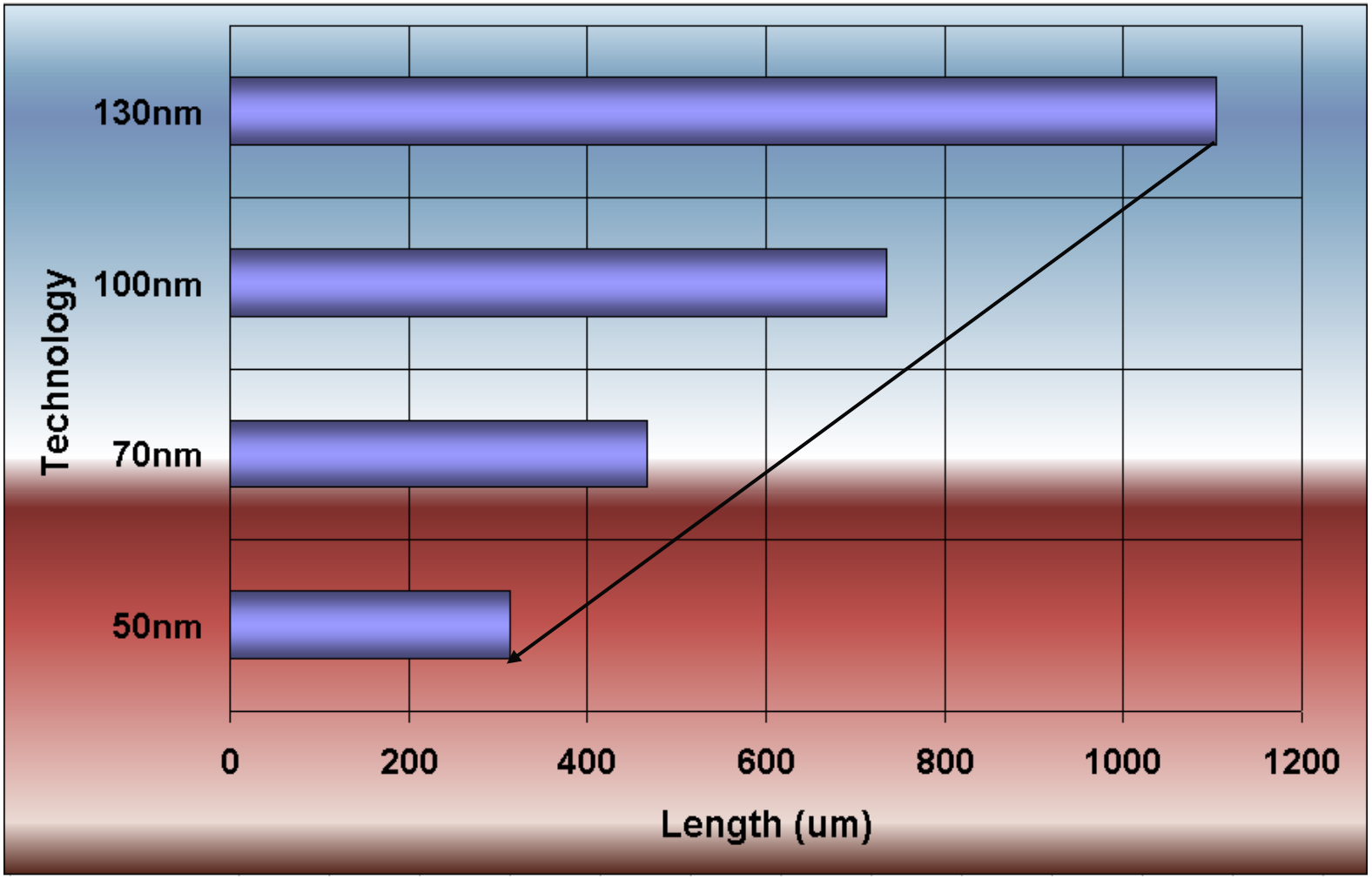
- (Interconnect) **Challenges in Nanometer Era**
- **Review of contemporary solutions**
- **Interconnect Scaling & Motivation for Packaging Approach**
- **WLP Technology**
- **Performance Comparison**
- **Applications**
- **Conclusion**

Reverse Scaling to Resolve Global Interconnect Challenge





Scaling decreases the repeater-less Communication radius

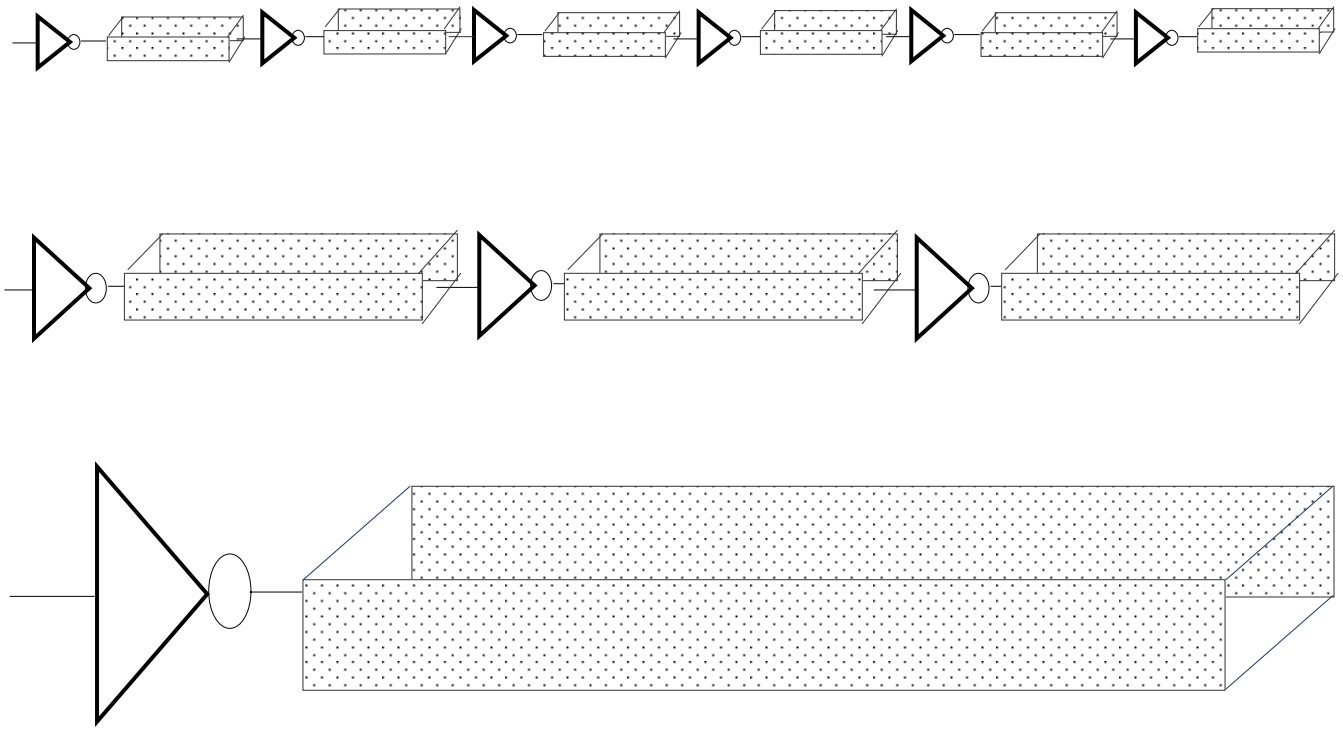


Scaling Trend: Interconnect Sizing vs. Repeater insertion

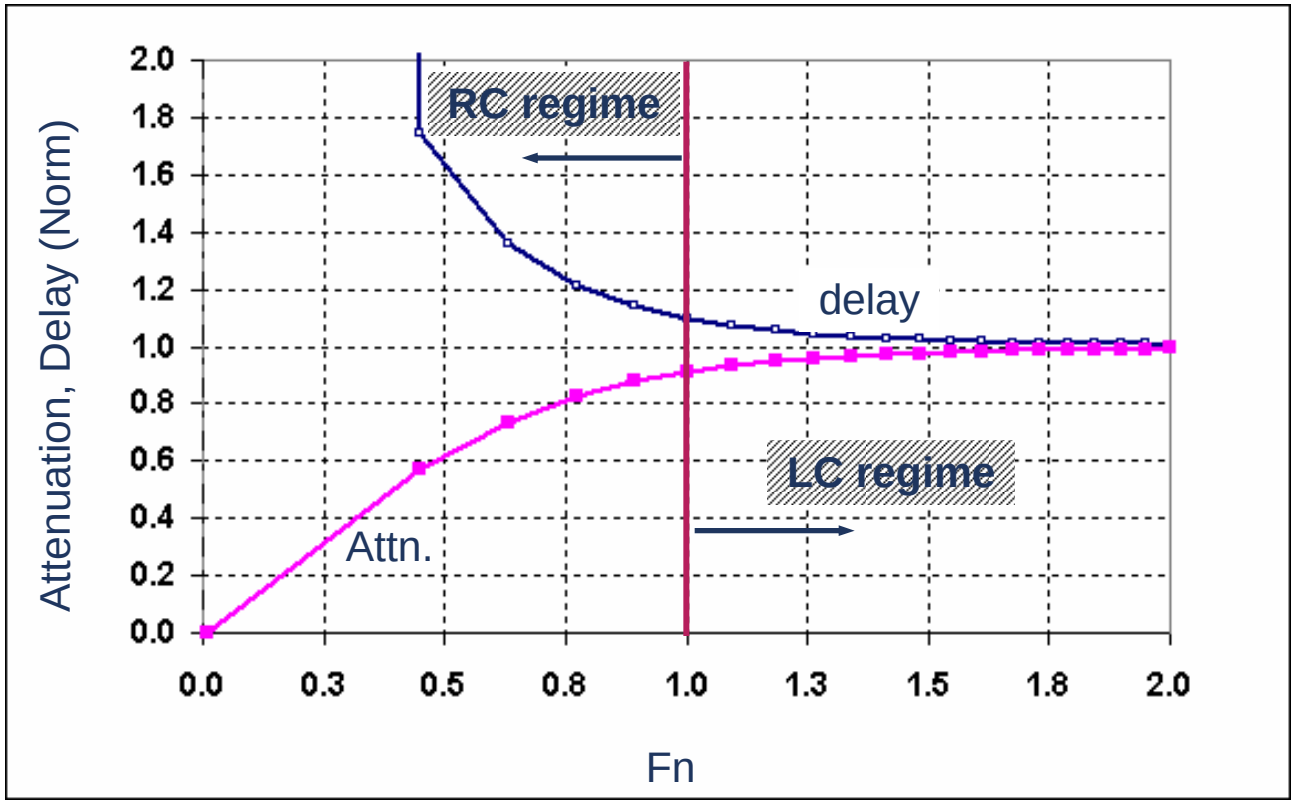
Low

Performance

High



Metal Interconnect Signal Propagation Characteristics

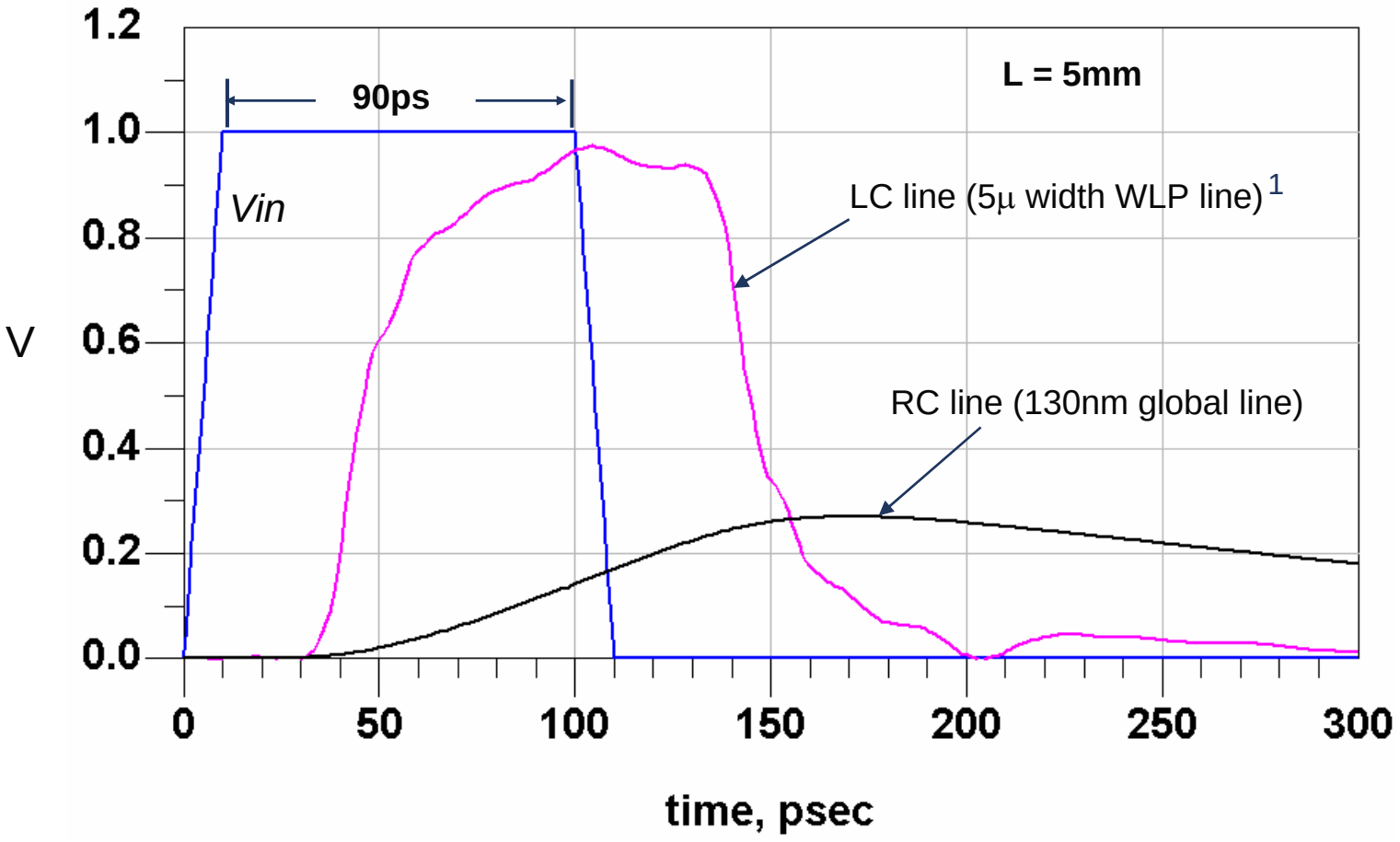


- Delay normalized to near speed of light delay
- Fn – frequency normalized as $\sqrt{6.28\, f\, (L/R)}$
- Speed of light propagation within 'tolerable' attenuation levels

RC vs. LC Characteristics

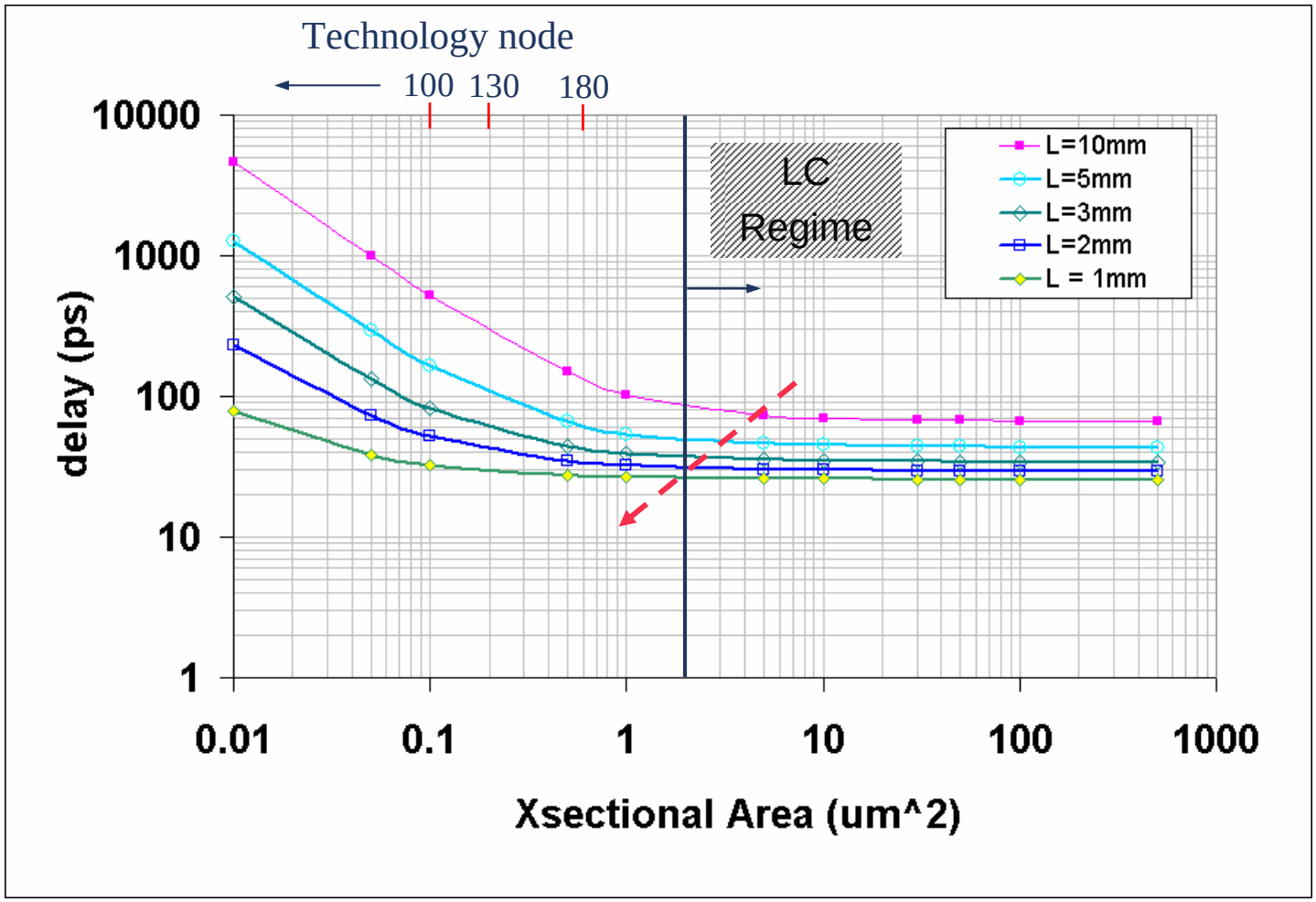
Attribute	RC Lines	LC Lines
Signal propagation	Diffusion – slow	Near speed of light transmission *(Limited only by dielectric constant of the ILD)*
Dispersion	High	Low
Attenuation	High	Low
Delay with length	Quadratic	Linear
Repeaters	Required (depending on distance)	Not required
Return path	Not significant	Significant
Pulse response	Poor	Good
Type	On-chip interconnects	Off-chip interconnects

LC lines offer superior pulse response



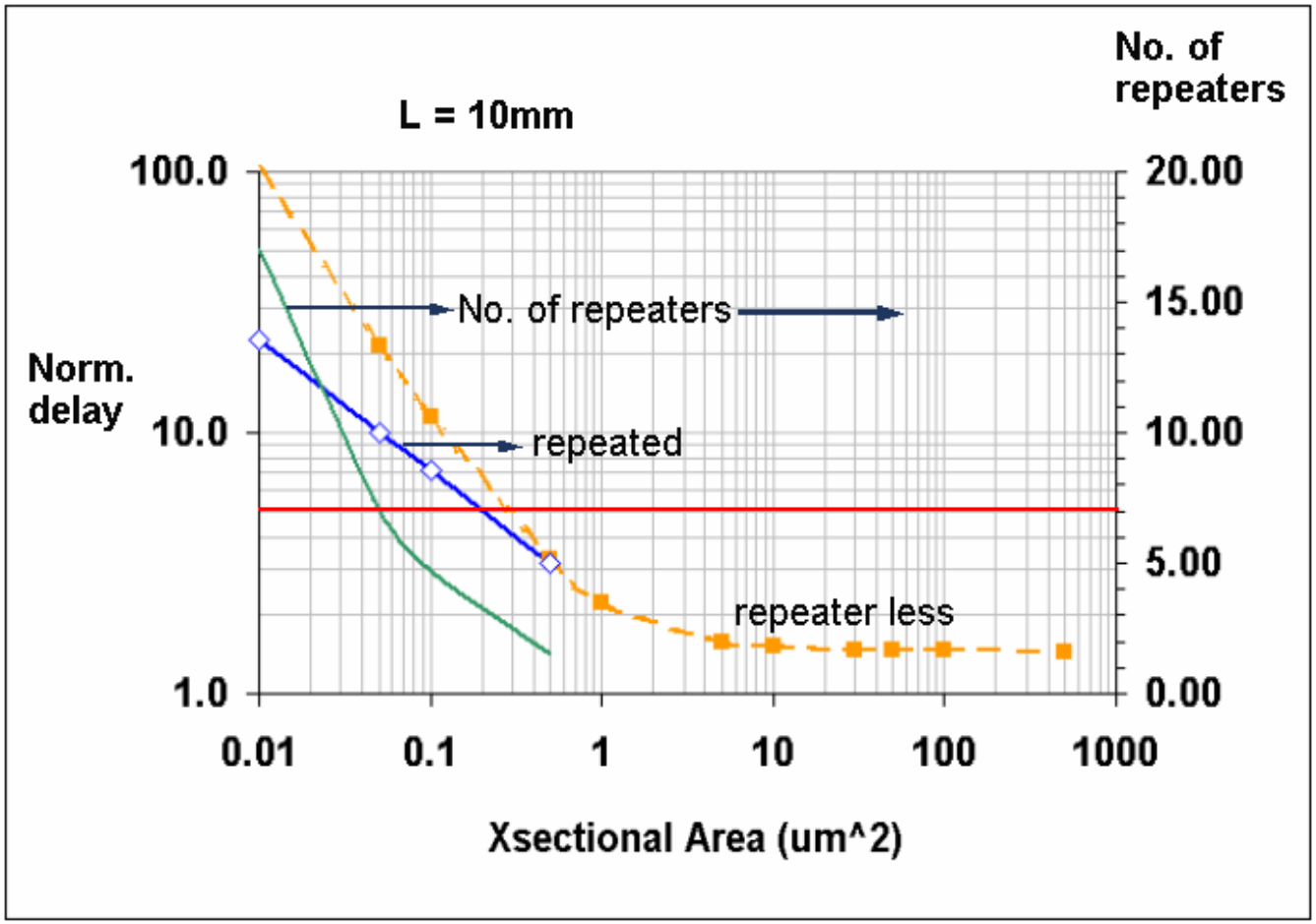
1 – Spice simulation from measured S-parameters

LC regime occurs for XSectional Area > 1um²



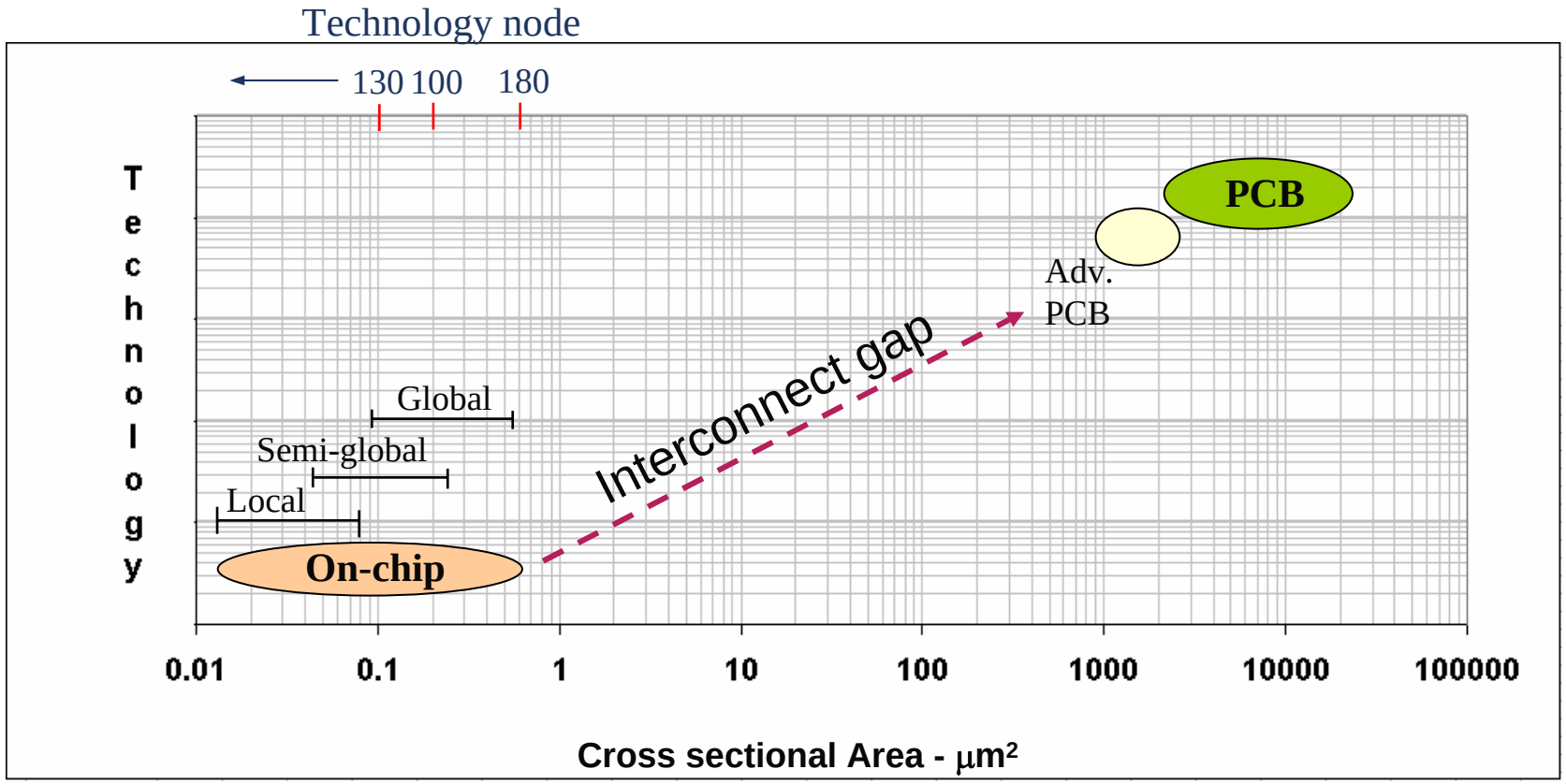
Results from spice simulation
130nm BPTM model for driver

Wiring sizing outperform Repeater Insertion



- Diminishing returns with repeaters for increasing wire x-section
- Below red line wire sizing outperforms repeaters

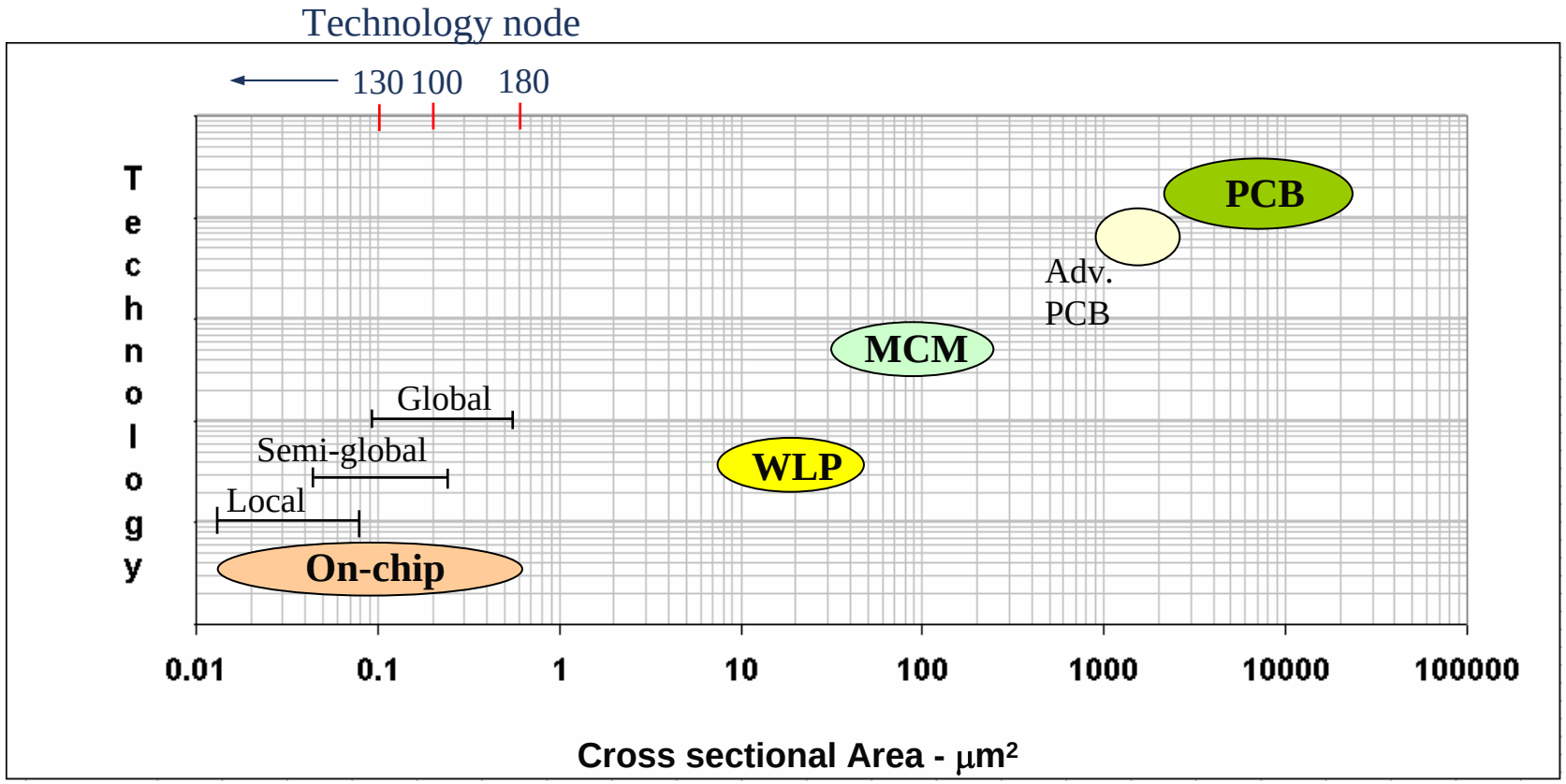
Wide Gap in the interconnect Hierarchy



Fat wires on-chip?

- Technologically challenging
- Signal integrity
- Cost

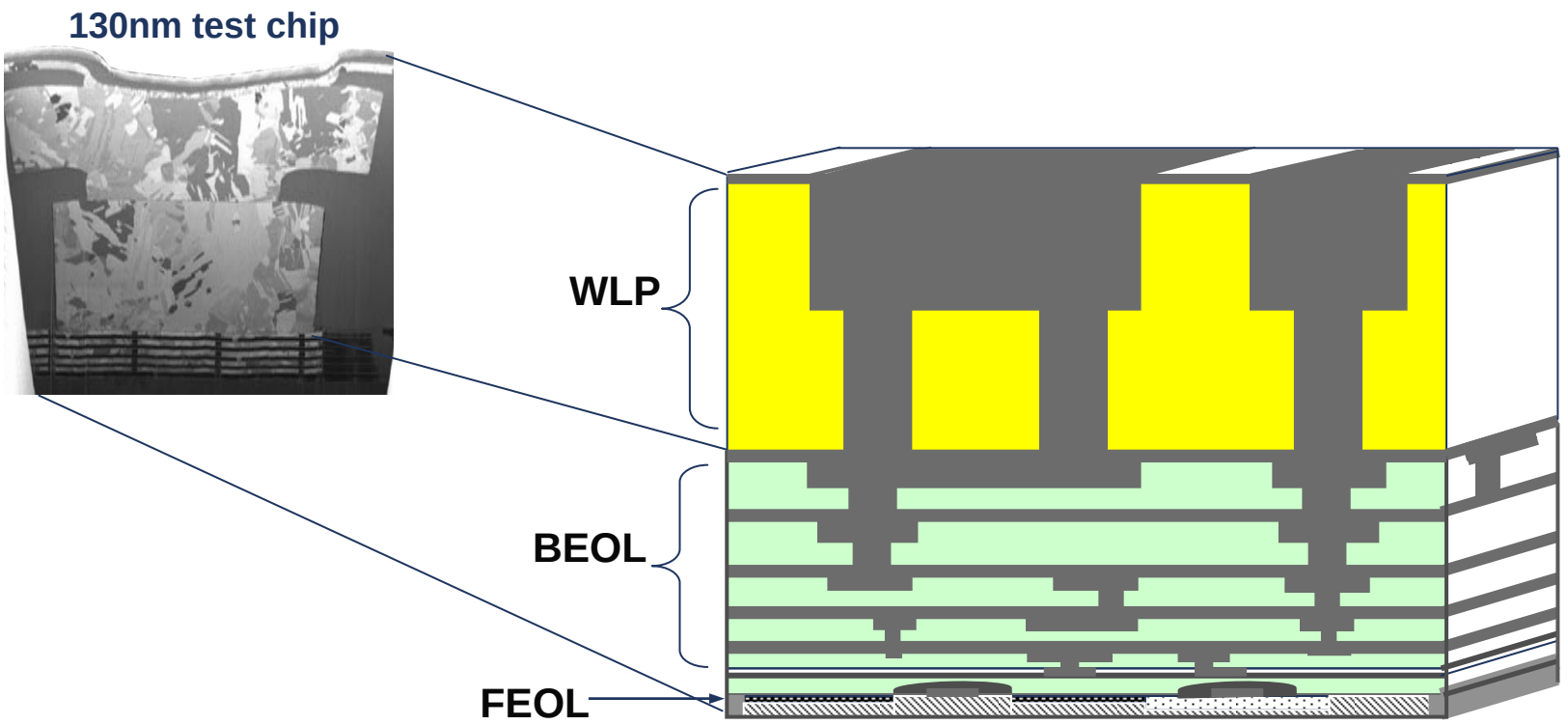
Bridging the Interconnect gap with Wafer Level Packaging



Fat wires on-chip?

- Technologically challenging
- Signal integrity
- Cost

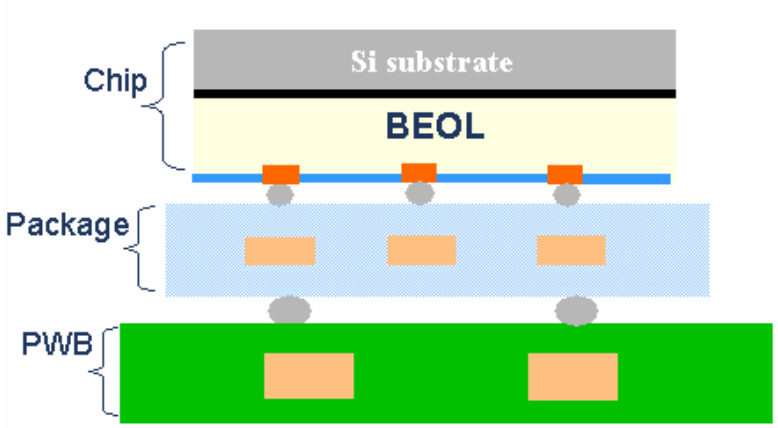
WLP Concept



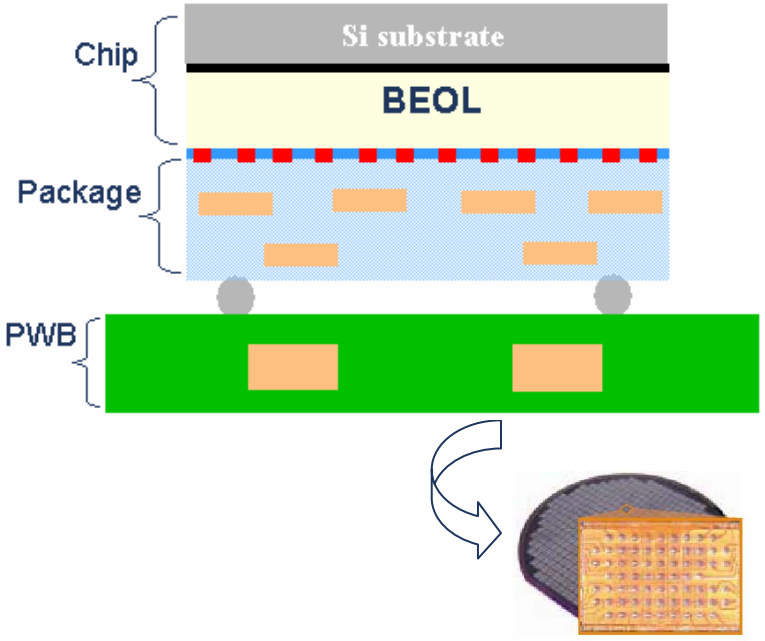
- WLP wiring layers are post processed on top of active wafer
- Modular approach to wire stacking
- Thick low-k ILD with larger metal cross section

WLP enables higher wiring density than conventional packages

Conventional Package wiring



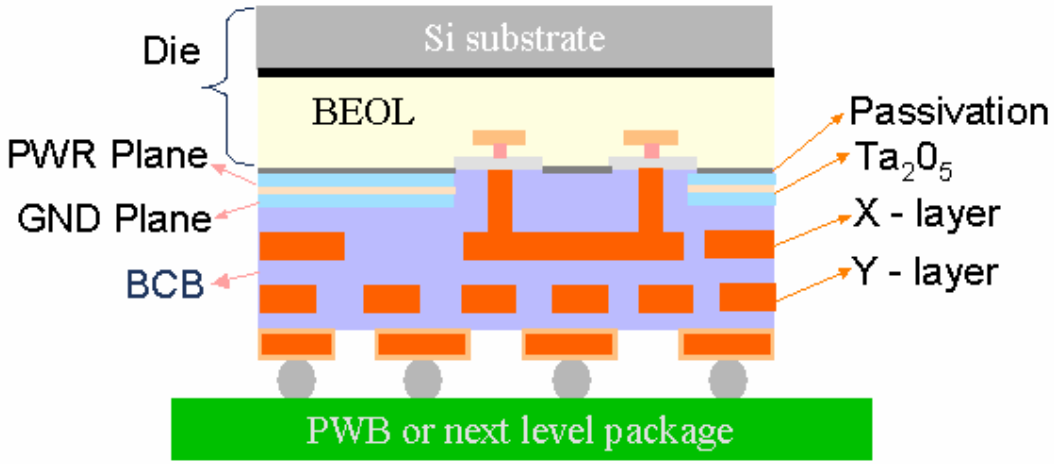
WLP wiring



WLP

- uses *vias* for inter layer connection
- Package wiring layers form a part of chip

WLP Technology in our Test chips



- 2 user routable layers
- Cu metals and low-k BCB dielectric (ε_r = 2.65)
- Two process configurations
  - Config1 – 6 masks, IMPS T-lines
  - Config2 – 8 masks, Ta_2O_5 decoupling capacitors, Microstrip T-lines

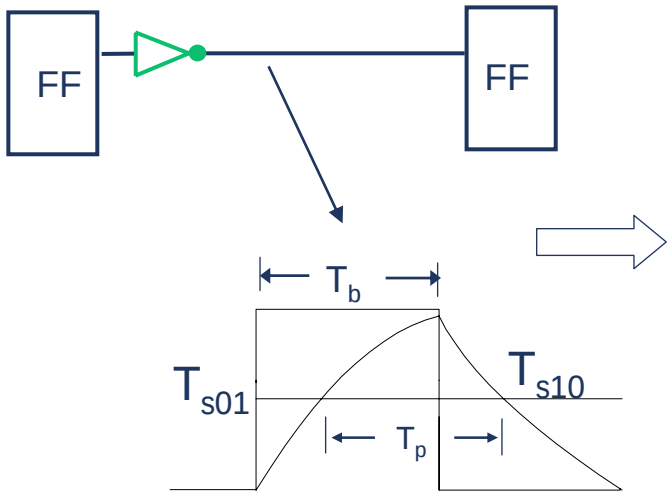
Outline

- (Interconnect) **Challenges in Nanometer Era**
- **Review of contemporary solutions**
- **Interconnect Scaling & Motivation for Packaging Approach**
- **WLP Technology**
- **Performance Comparison**
- **Applications**
- **Conclusion**

Metrics for Package-BEOL Comparison

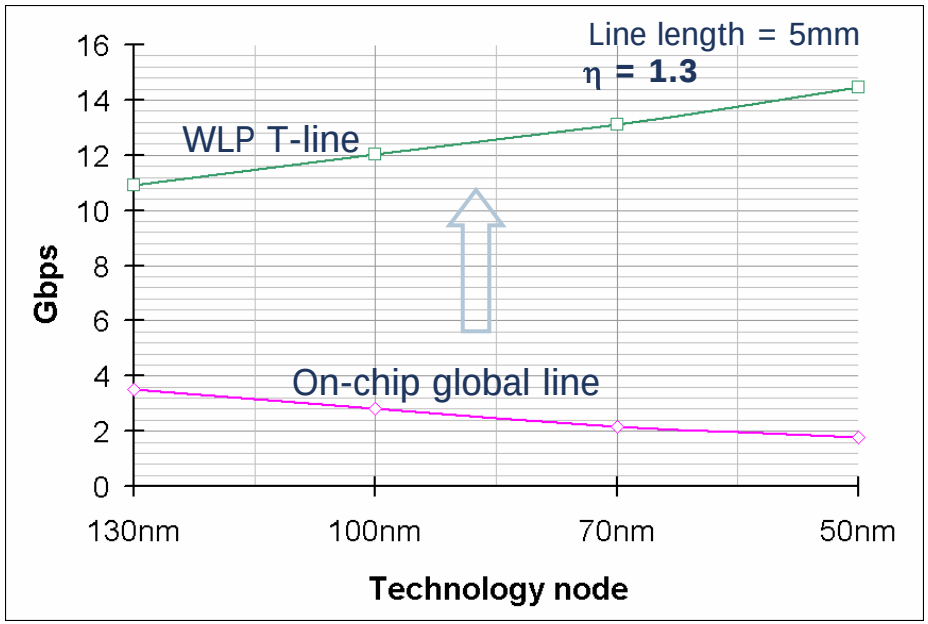
- Bandwidth (BW)
  - No. of bits transmitted per second per wire
- BW density
  - BW across unit routing area
  - Measure of wiring density
- Latency
  - Delay with respect to clock period
- Power

3-7x BW improvement possible with Packaging Approach



$$BW_{RC} = \frac{1}{\eta\left(T_{s01} + \frac{Tp}{2}\right)} \qquad BW_{LC} = \frac{1}{\eta\left(T_{s01} + T_p\right)}$$

$\eta \geq 1$ - signal integrity factor



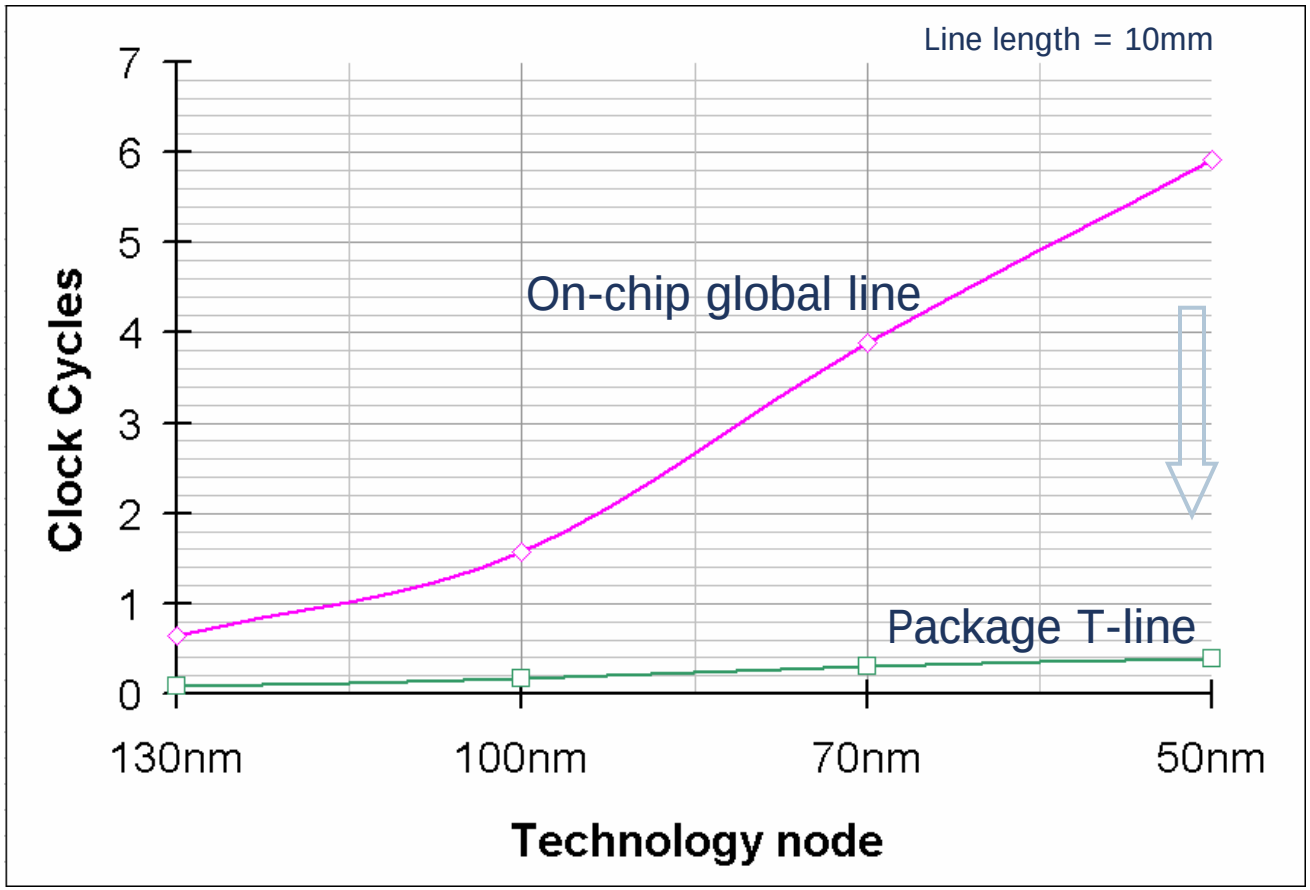
Package Bandwidth is limited by *register setup and hold times*

J.Balachandran et.al SLIP 2005

Package Level Interconnects Enable Single Cycle Communication

$$LT = D_{\text{int}} * F_{clk}$$

LT = latency
D_{int} = interconnect delay
F_{clk} = Clock frequency specified by ITRS

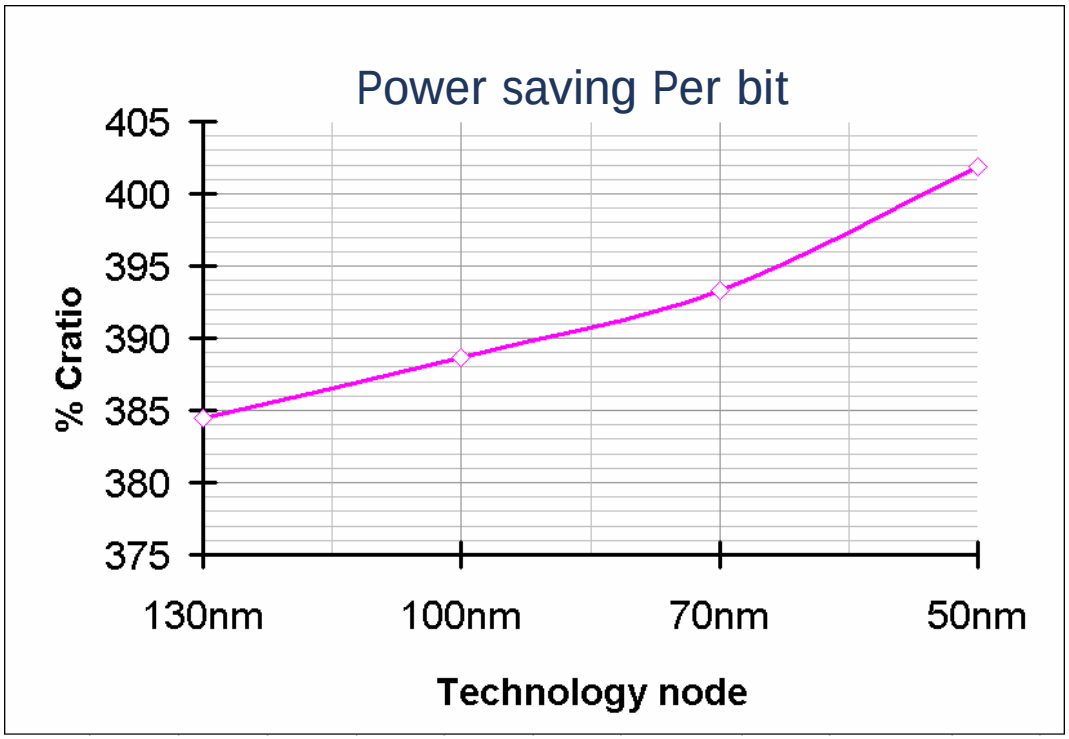


Significant Power savings with Package level Interconnects

For comparing power, Capacitance ratio is derived as

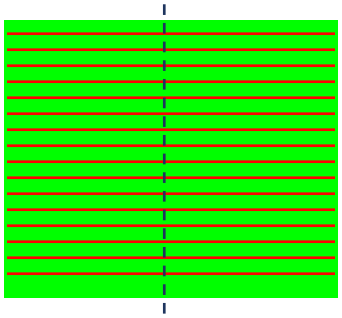
$$\frac{C_{chip-global}}{C_{pkg-T-line}} = \frac{1.75 * Zo * C_{i-onchip}}{\frac{\sqrt{\varepsilon_r}}{C_o} + R_t C_t}$$

Zo = characteristic impedance of transmission line
Ci-on chip =On-chip global interconnect capacitance
Rt,Ct – minimum sized driver output resistance and input capacitance

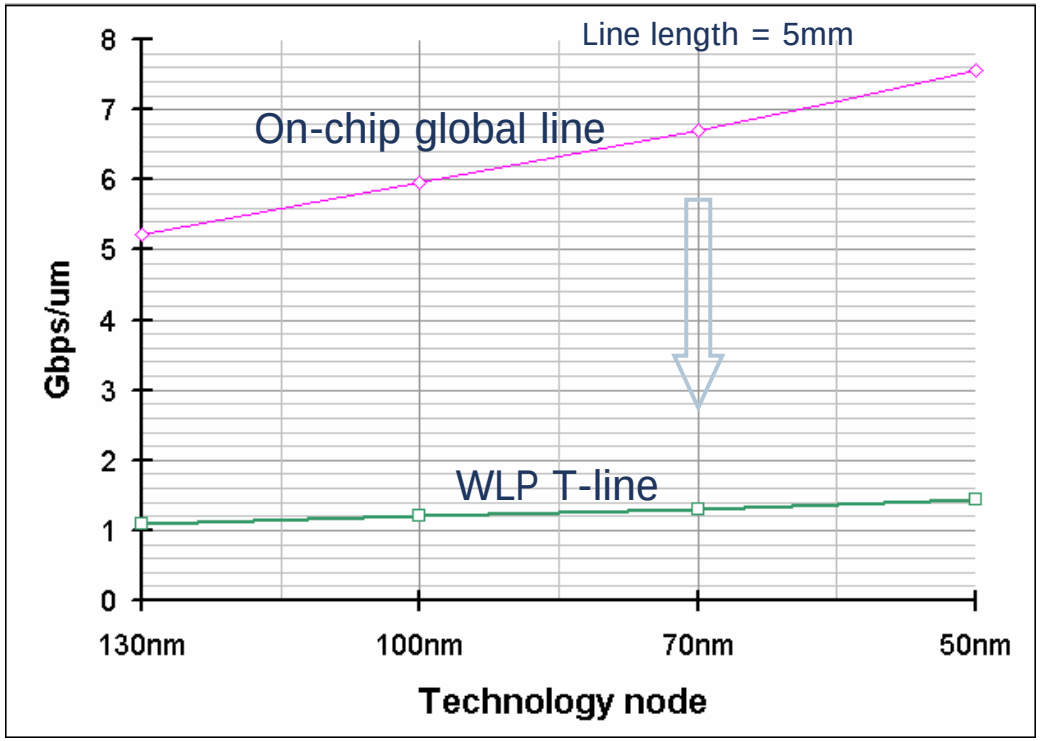


Zo = 75Ω

Bandwidth density is lowered in Package level Interconnects



$$BD = \frac{BW}{W+S} = \frac{BW}{P}$$



J.Balachandran et.al SLIP 2005

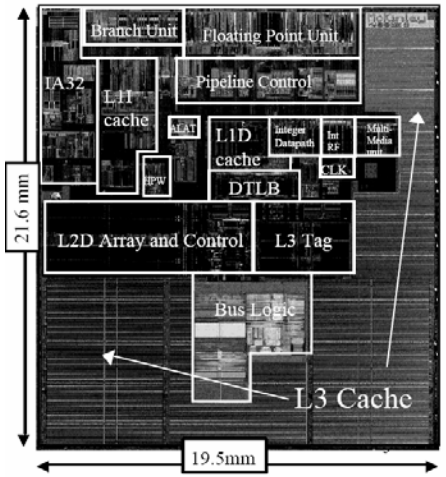
Outline

- (Interconnect) **Challenges in Nanometer Era**
- **Review of contemporary solutions**
- **Interconnect Scaling & Motivation for Packaging Approach**
- **WLP Technology**
- **Performance Comparison**
- **Applications**
- **Conclusion**

Interconnect Options: Memory buses

- Off-chip BW is limited – strong trend for more on-chip memories

Memories occupies significant die area

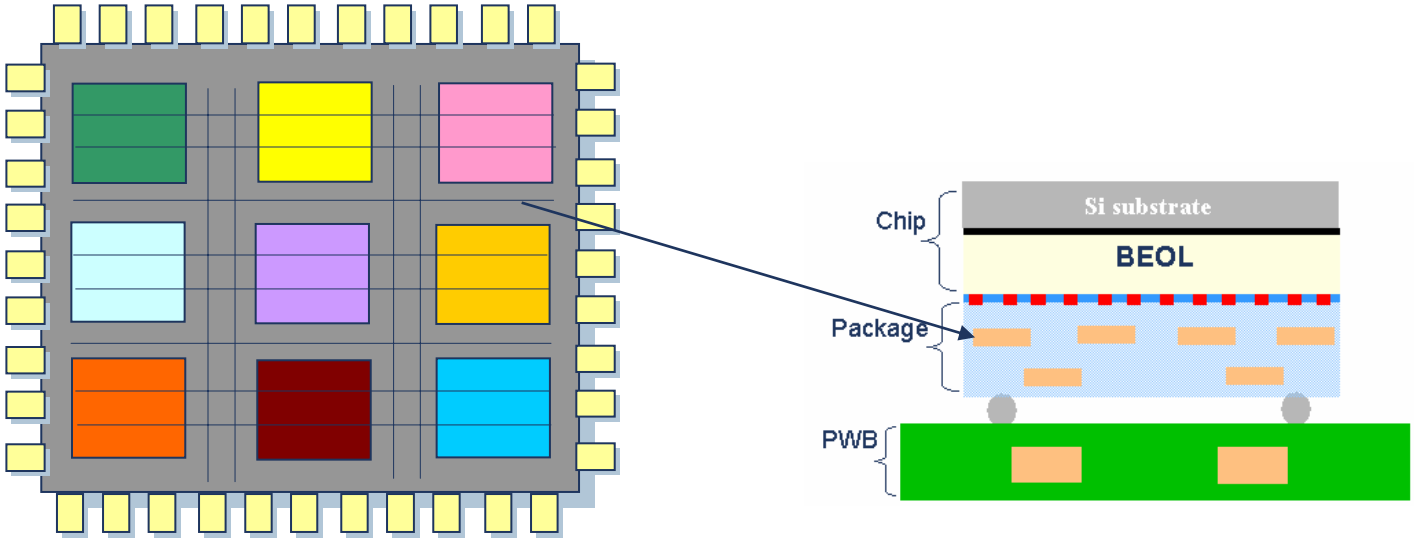


Micrograph of a leading Processor

Source: ISSC 2002

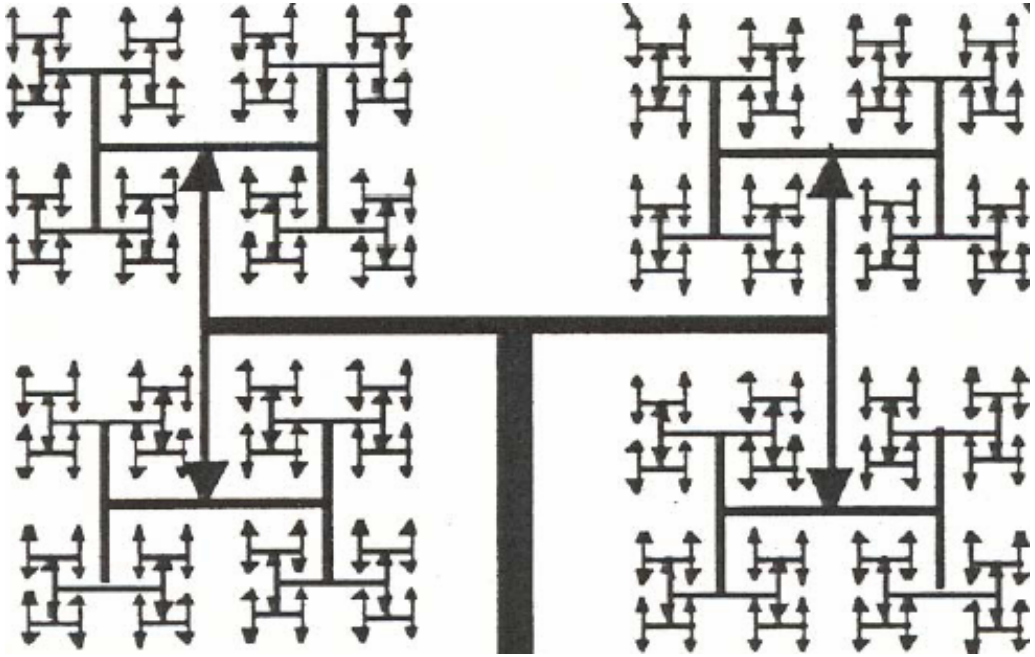
- WLP interconnects allow low latency memory accesses spread across the die
- Repeater less communication enables routing over hard macro blocks
- Can be routed over sensitive analogue blocks
- Flexible floor planning
- Low power

Interconnect Options: Inter Tile Communication



- Exploit LC Transmission line properties of WLP interconnects
- NOC or Bus ?

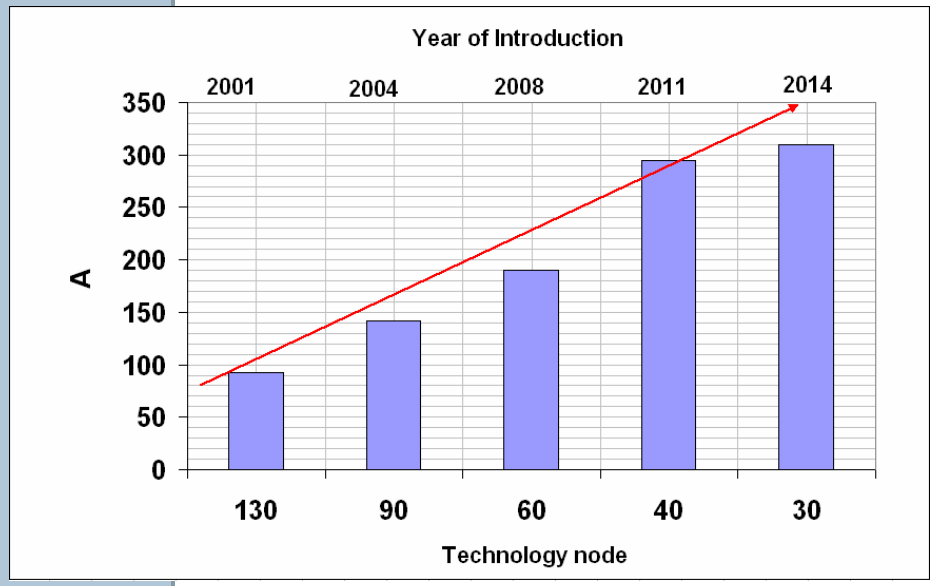
Interconnect Options: Clock distribution

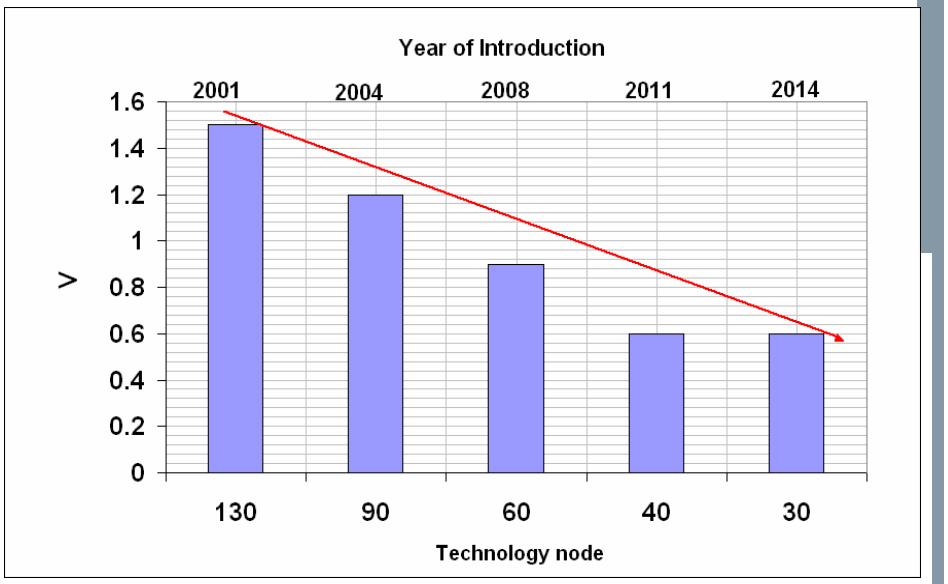


- First few levels of clock trees can be in WLP
- Inherent low rise time as compared to on-chip lines
- Low latency -> relaxed skew tolerance
- Low power

Interconnect Options: Power distribution

DC Current and Voltage requirements for ITRS Technology nodes





At 40nm node, to have 10% IR drop tolerance,

Total wire Resistance should be less than *0.2milliohms!*

⇨ Package planes and lines can be used for power distribution

Conclusion

- Global interconnect delay and power delivery issues dominate nano-CMOS designs
- Reverse scaling enabled by WLP Packaging approach is promising :
  - *Extends on-chip Wiring hierarchy*
  - *High BW @ low power and latencies*
  - *Simultaneously address signal and power distribution*
  - *However limited wiring density*
- Options
  - Memory buses, Intertile Communication, Clock and Power distribution

Thank you

Acknowledgements:

Scott List, Maarten Kuijk,Franky Cathoor, Philip Christie,Mandeep Bamal