The Light at the end of the CMOS Tunnel

(or, advice on what to do with a Silicon career)

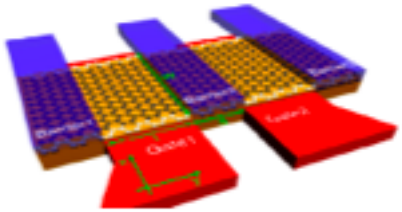
Sani R. Nassif

IBM Research – Austin

nassif@us.ibm.com

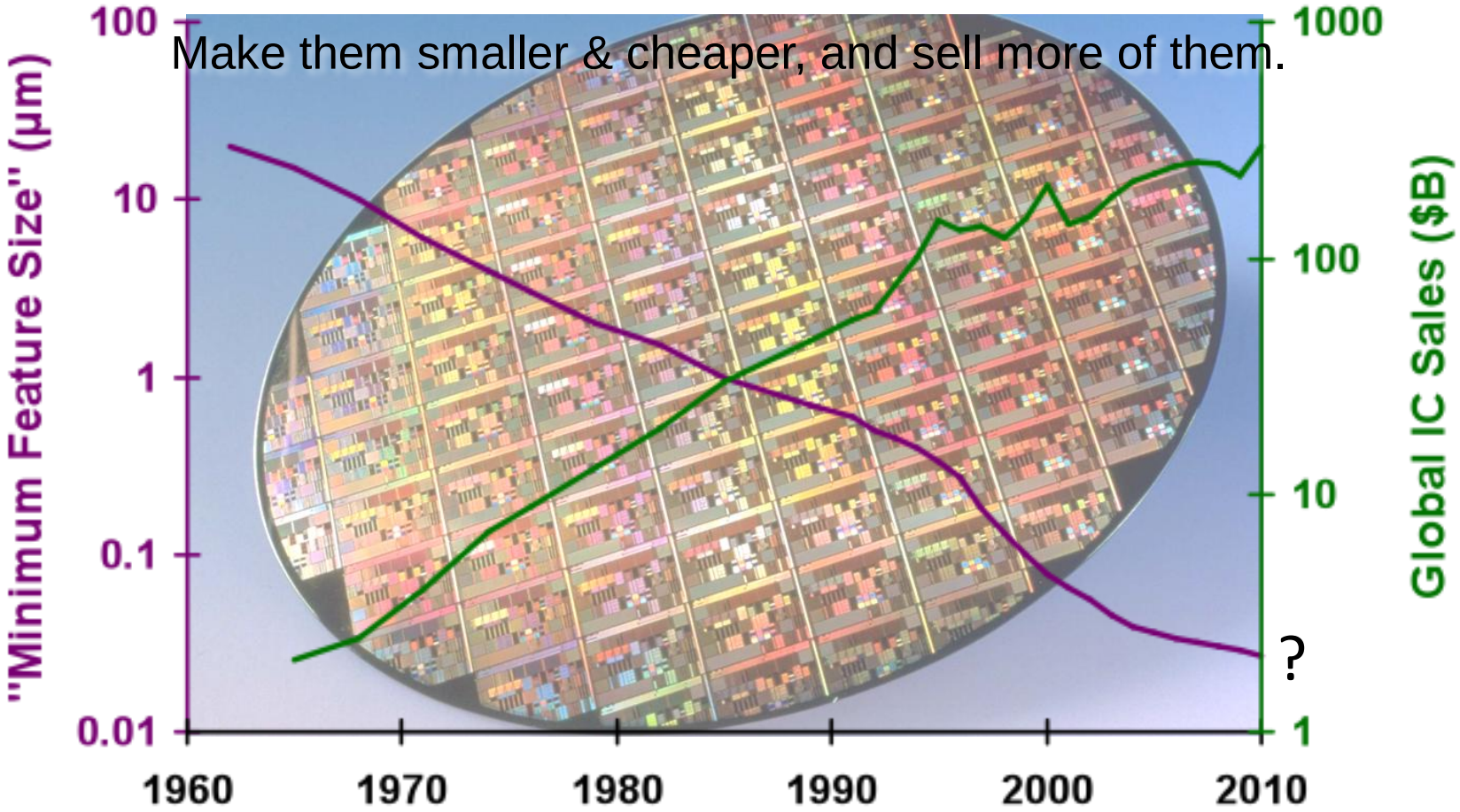
Trends and Implications

- Integrated circuits have been with us for ~40 years and now permeate every aspect of human existence.
  - Few disciplines have done so much in such a short time.
  - Few disciplines can sustain the pace of semiconductors.



- Yet now we are poised for “Post-Silicon”.
  - The industry is in the midst of considerable consolidation.
  - Many have stepped off the Moore treadmill.
  - Difficult technical challenges abound…
- What is a Silicon researcher to do?

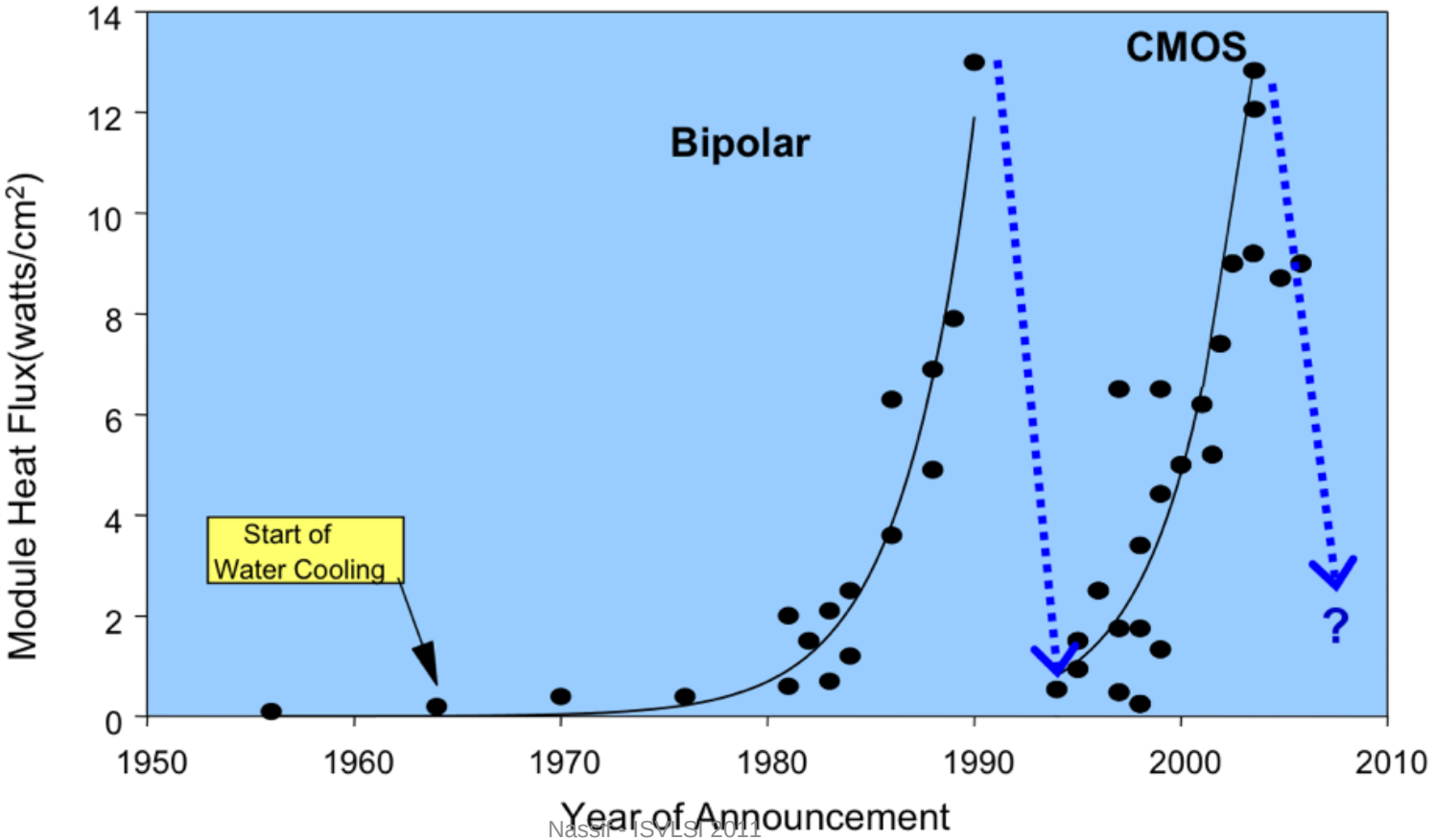
The Semiconductor Economy...



Courtesy of R. Doering, Texas Instruments;
Data from Semiconductor Industry Association, http://www.sia-online.org

Challenges to Silicon are not New!

- Many were certain we could not break the 1000nm barrier.
  - The phrase “sub-micron challenges” was quite common.
- Challenges have driven deep change: e.g. Bipolar to CMOS.



What Challenges Does the Future Hold?

- One can easily get confused when discussing Silicon, Post-Silicon, and the emerging world of Nano-Stuff.
  - *Or at the very least, the long-term research funding agencies certainly appear to be confused.*

Two possible views, using the same two words:

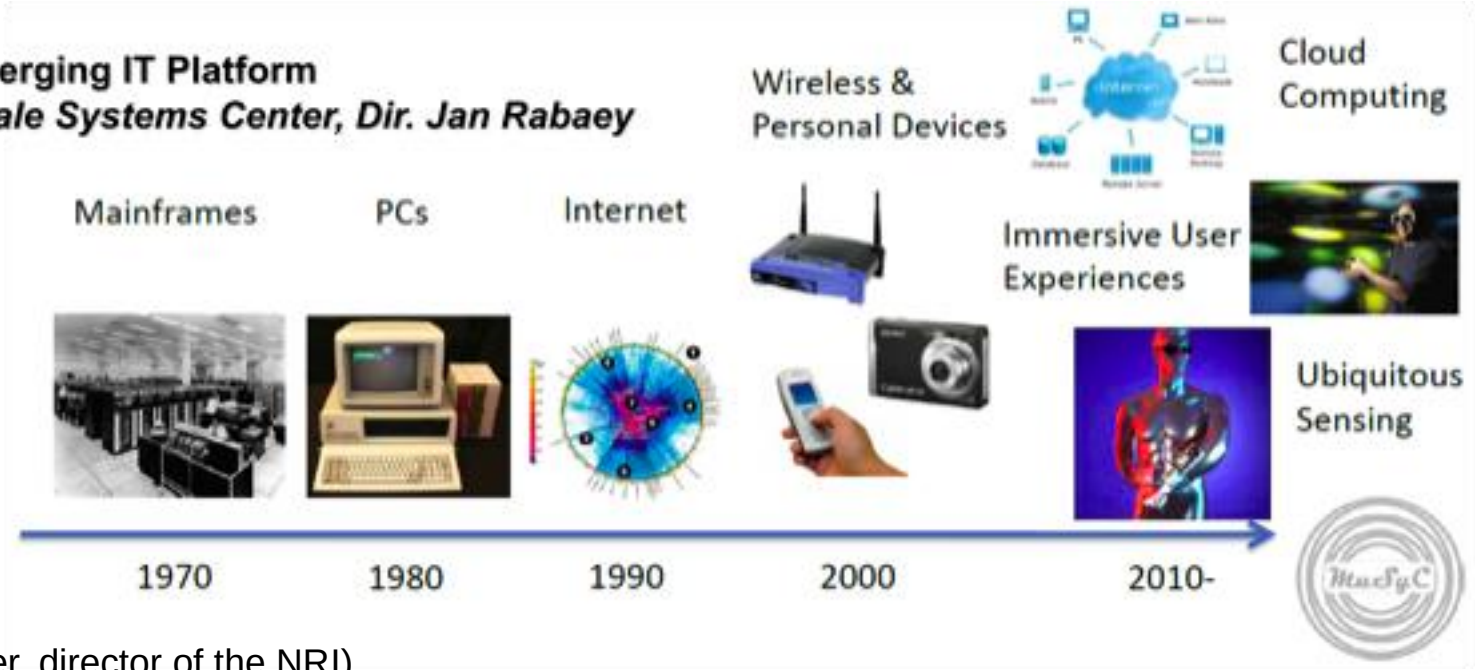
- What post? (there is no technology post CMOS).
- Post What? (the new technology will be so good we will not even think about Silicon CMOS any longer).

- Neither extreme view is adequate or correct.

Vision for the Future (from Jeff Welser, director of the NRI)

"HOW" to Build to "WHAT" to Build!

- Barrier to future scaling changing from "how do we make them smaller?" to "how do we reduce power to make them usable?"
  - Change happened between the 100nm and 10nm barriers!
- Many new materials with new physics, properties & functionality.
  - Shift from single device focus to circuits/architecture integration



(from Jeff Welser, director of the NRI)

The Key Question: When?

- For that, we must ask the Gods.

We are almost there... 2015?

CMOS will be around until 2030 or longer!

Ganesha

God of Success

Saraswathi

Goddess of Knowledge

What do we do while Waiting?

- There are a number of problems to overcome:

Lithography.

  - This is a complex issue, with the uncertainty around UV making things harder. Computational Lithography, Double/Triple patterning, and layout regularity are all in the mix.

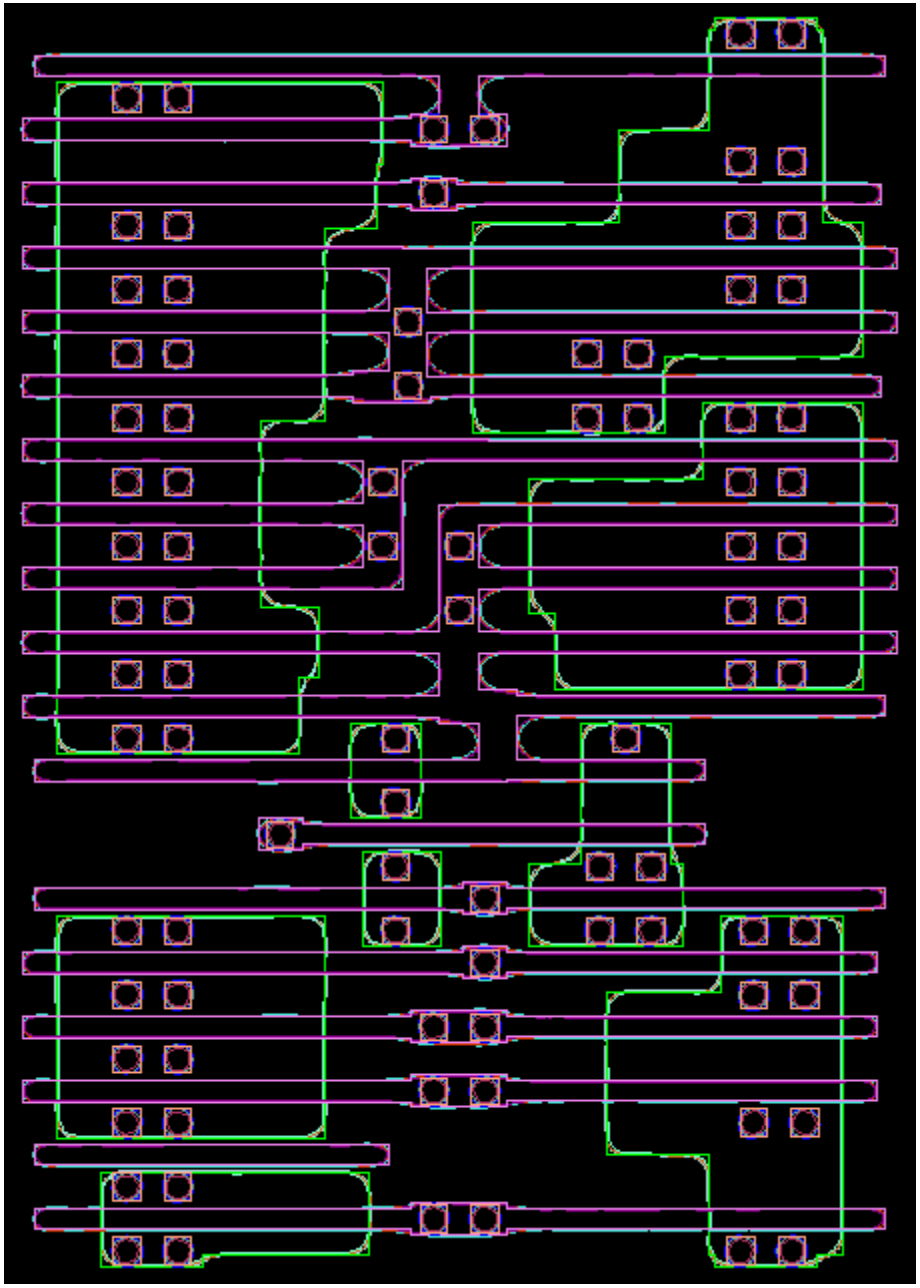
Power.

  - This will remain a hard problem. While some benefit can come from device and technology improvements, much remains at the software and system levels.

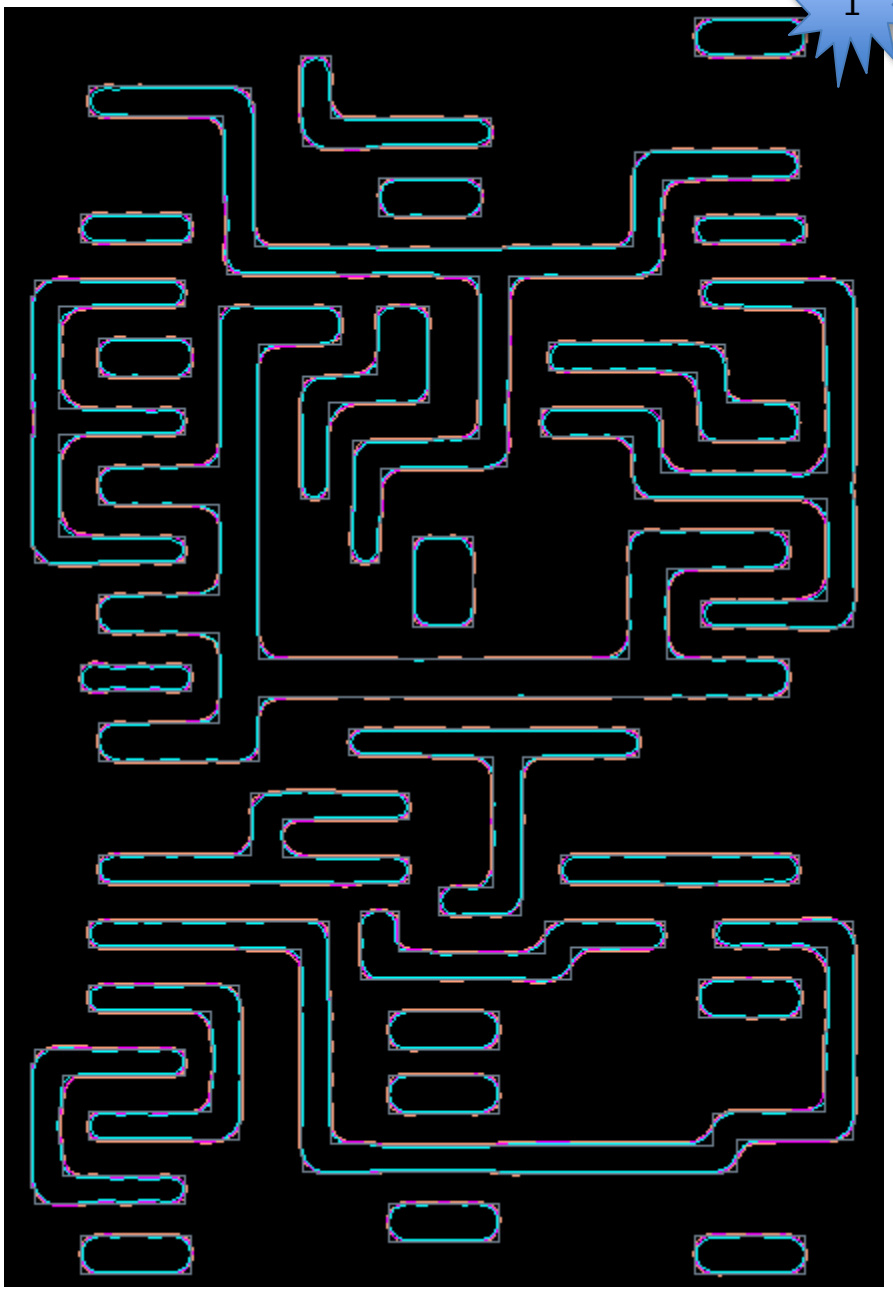
Resilience.

  - This gets less press time, but is just as hard a problem. We cannot count on 100% functionality any longer!

A Latch in 45nm Technology

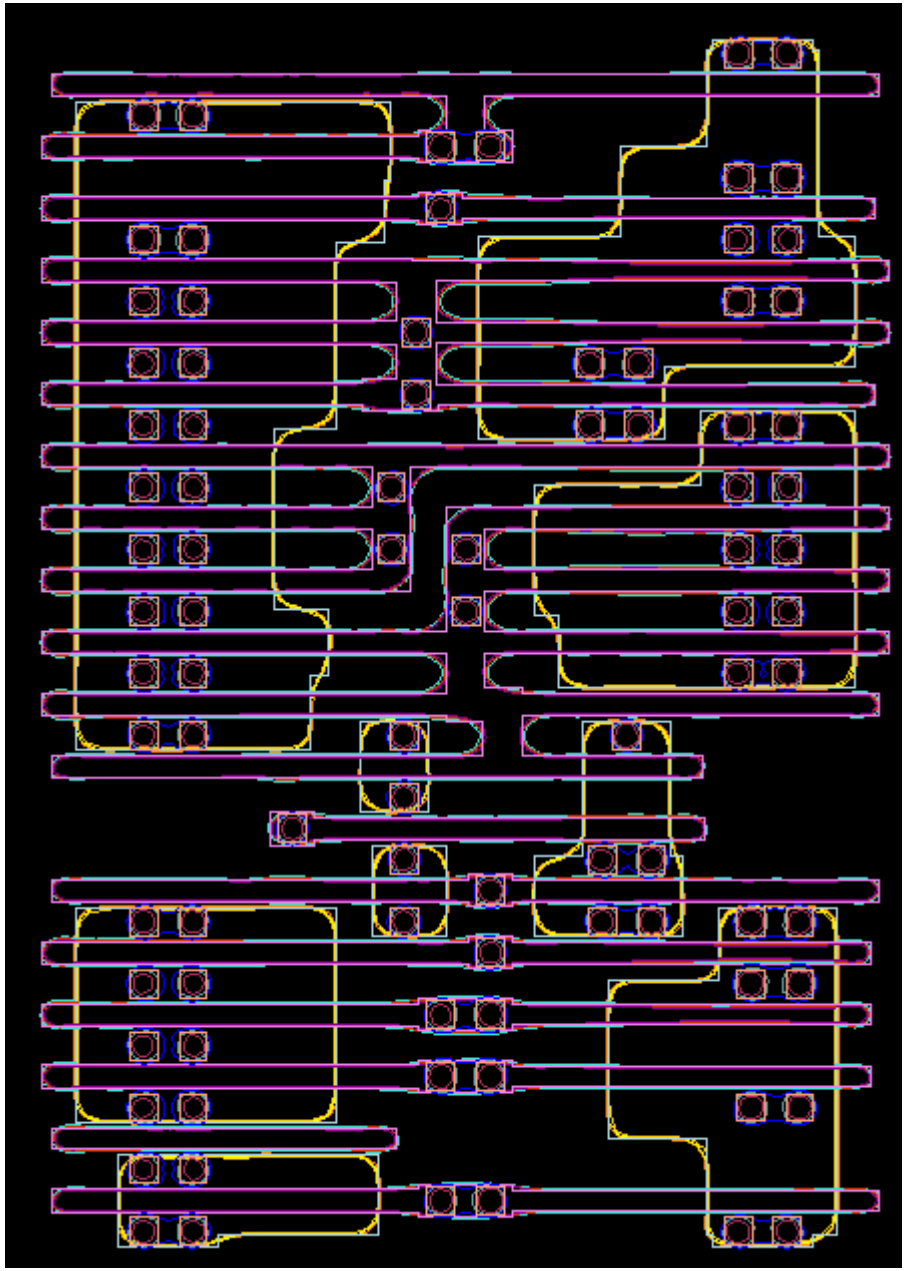


Polysilicon, Diffusion, and Via contours

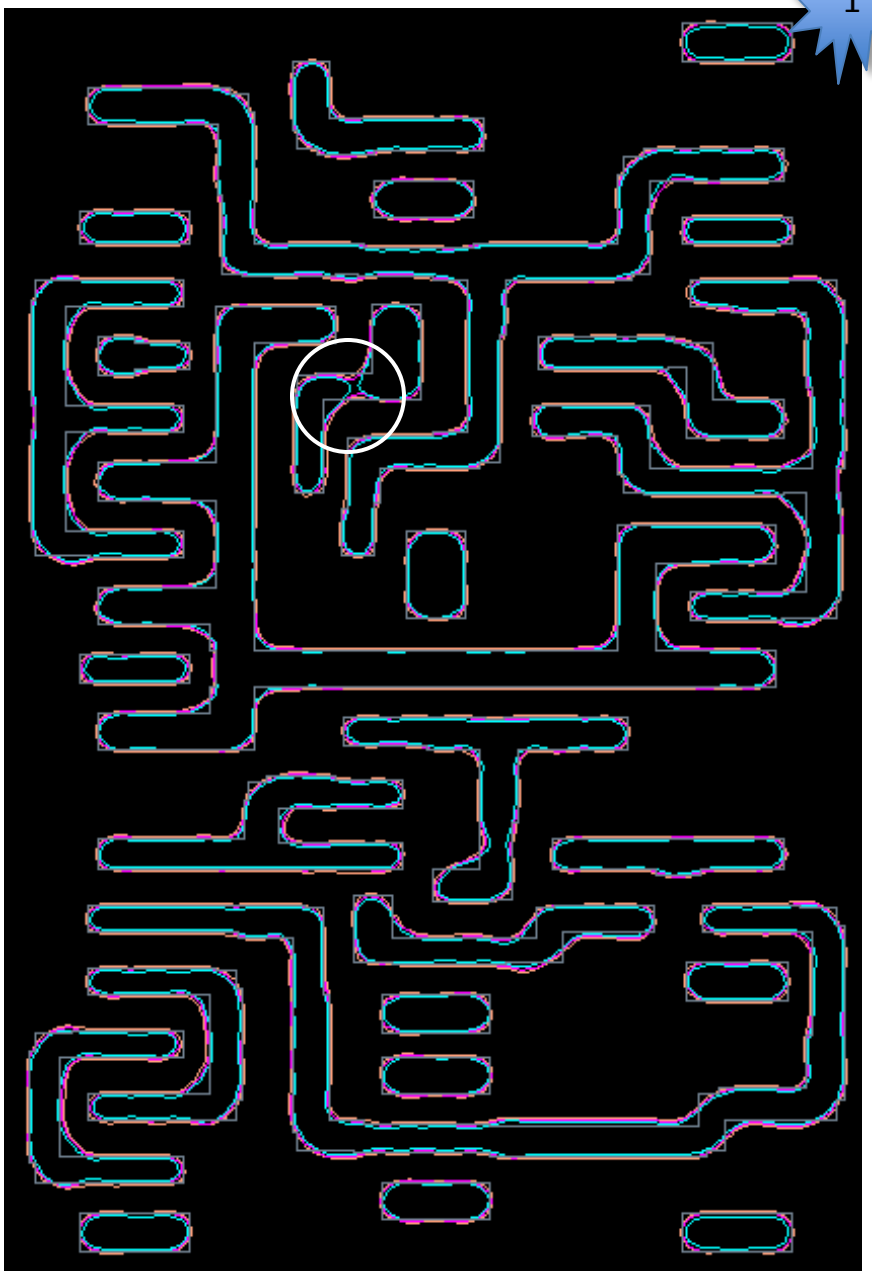


M1 contours

A Latch in 45nm Technology shrunk to 32nm

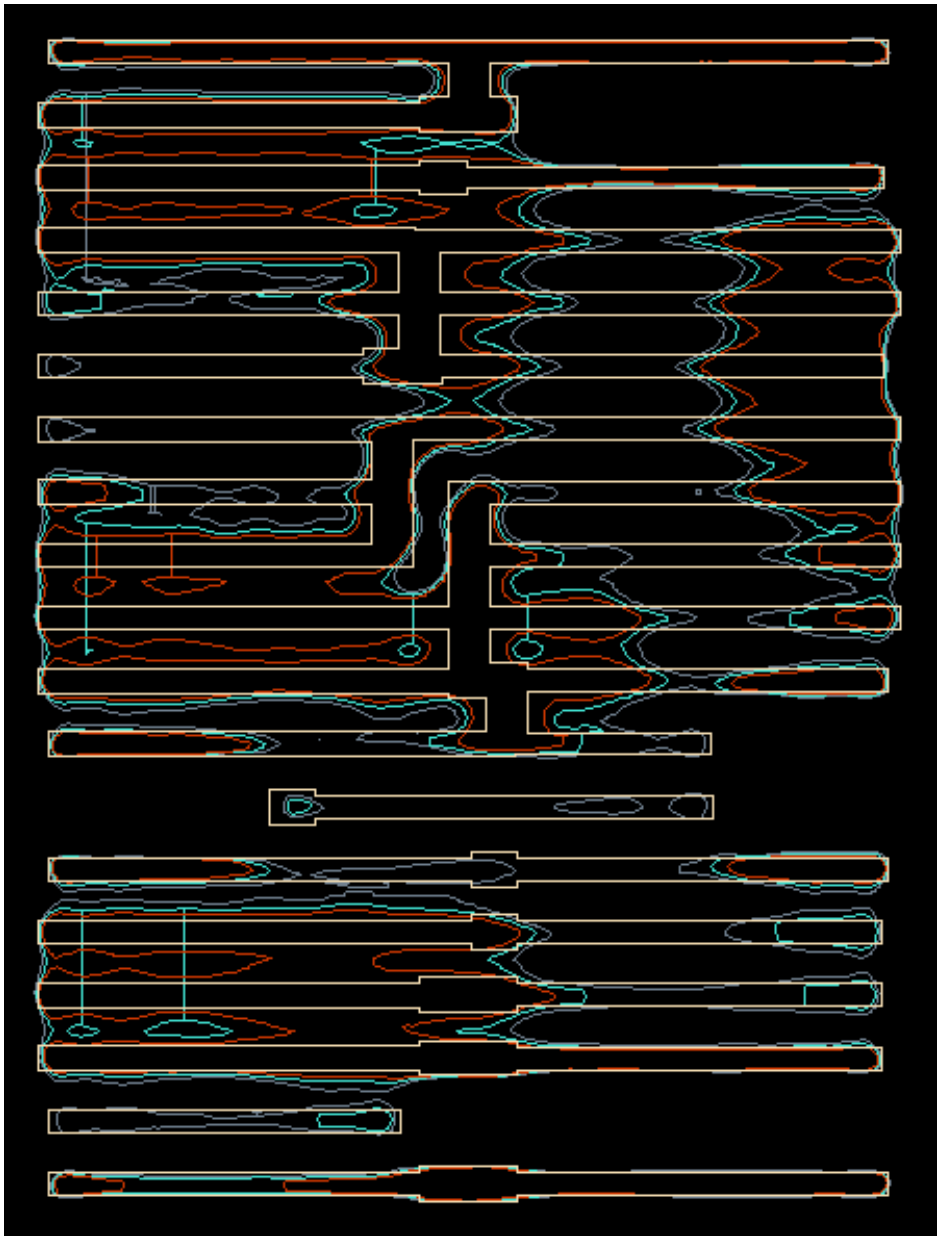


Polysilicon, Diffusion, and Via contours

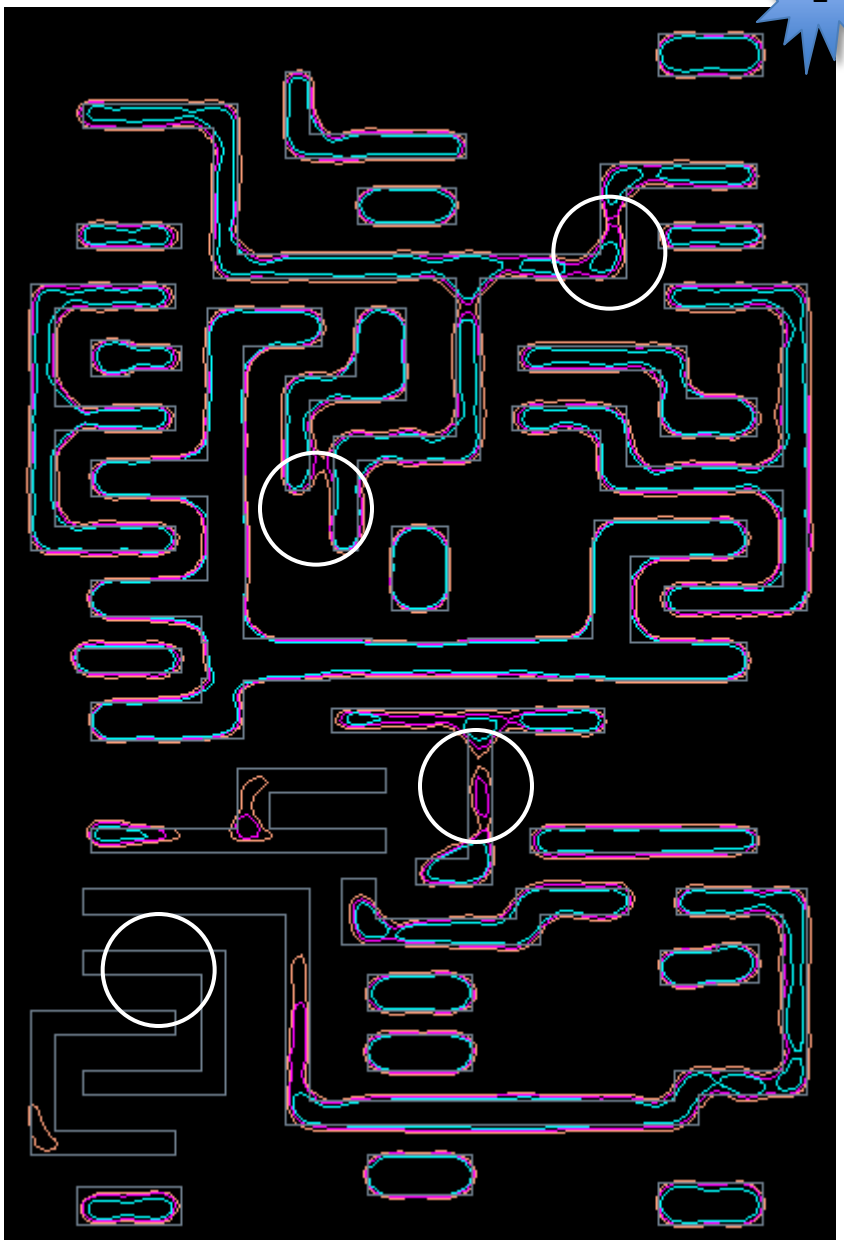


M1 contours

A Latch in 45nm Technology shrunk to 28nm



Polysilicon, Diffusion, and Via contours



M1 contours

How Real is This?

iGoogle | 2011 overview | ISVLSI 2011 | Inbox (46) – sani.nass

semimd.com/blog/tag/14nm/

14nm node design rules already set for HVM in 2013. An Intel 10nm node in HVM would follow in 2015.

At Nikon's LithoVision workshop, Sam Sivakumar, Intel fellow (*figure*), explained that the DR set for 14nm used 193nm-immersion double-patterning (193i-DP), and that for the 10nm node—featuring 20nm actual line width, and 40nm pitch—design rules will be frozen early in 2013. "So production EUV tools will be delivered too late to meet the need to develop DR for the 10nm node, though Intel remains committed to going into production with EUV" said Sivakumar. Regarding the possibility of re-insertion, it is possible but only after surmounting a barrier. "We'd need to reset the DR for EUV," elaborated Sivakumar, "because it doesn't make sense to use DR developed for 193i with EUV. The key point is that DR flexibility needs to be built in, so that we can smoothly insert EUV and derive maximum benefit."

Sam Sivakumar (source: Intel)

[***UPDATE 3/1***: In an evening panel session, Sivakumar asserted Intel's official position as, "Our primary plan is to use EUV for 10nm, but we need ArF double-patterning as a backup." Intel will have a pre-production 3100 tool this year, and likely will want a production 3300 whenever available. "When going to immersion, it

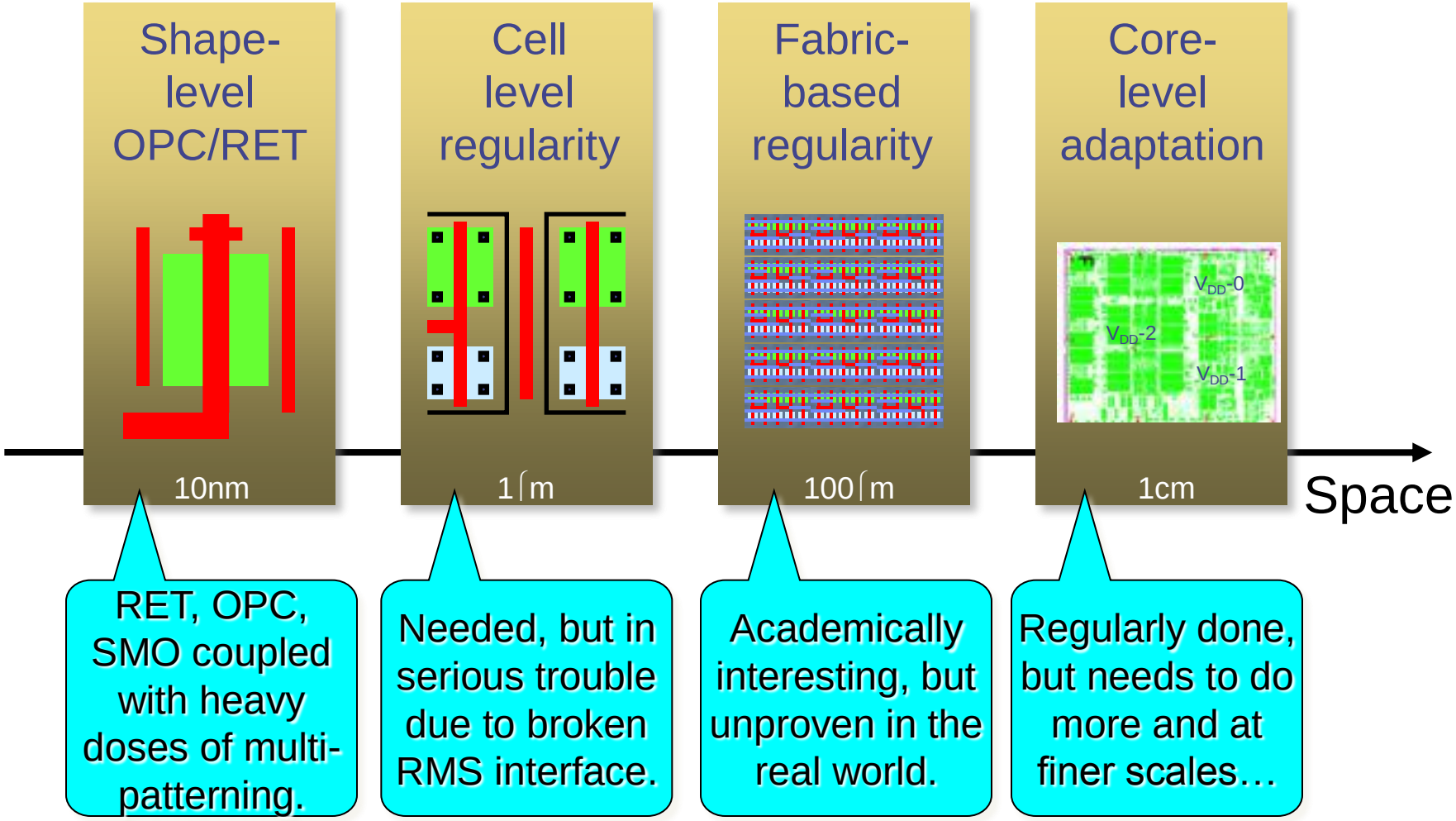
materials used in solar cells and modules is expected to grow to $16.9 billion in 2015, up from $6.5 billion in 2010, according to a

Intel will make 14nm features using 193nm light.

Like using a broomstick to sign your name…

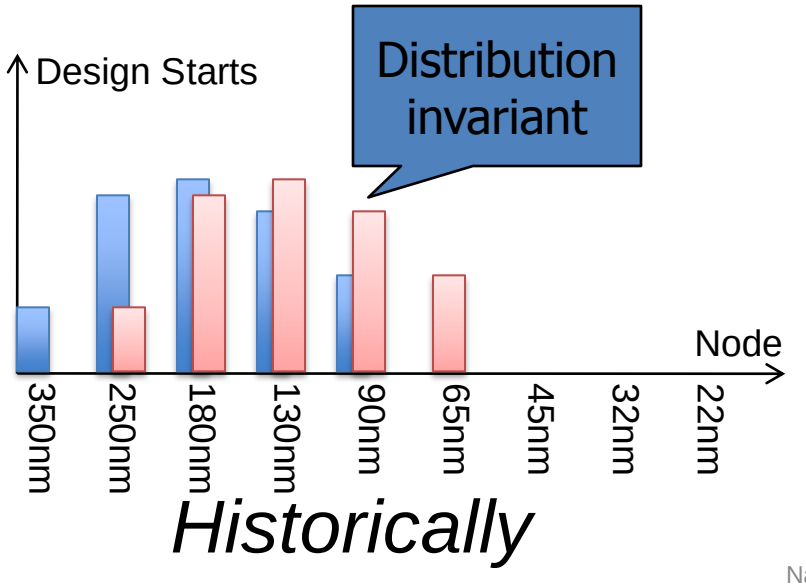
Responses to Lithography Difficulties

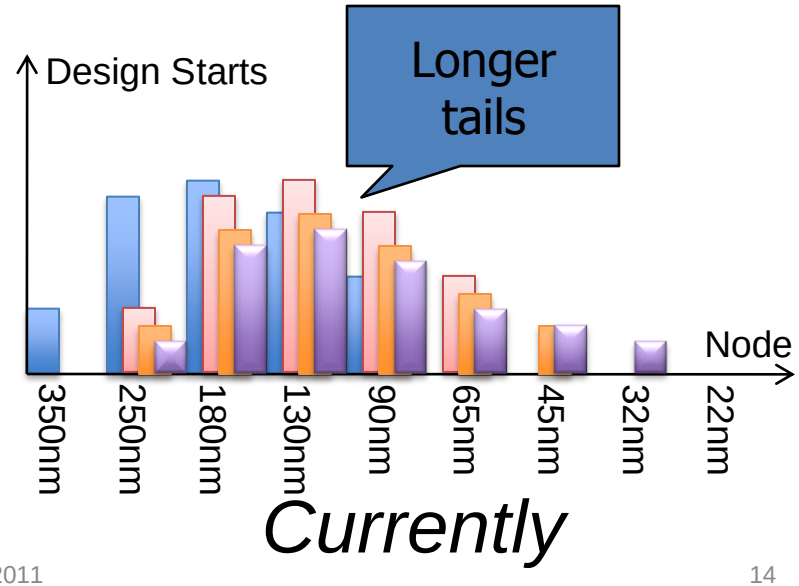
- Responses at different levels of granularity!



RET, OPC, SMO, DP (Litho Alphabet Soup)

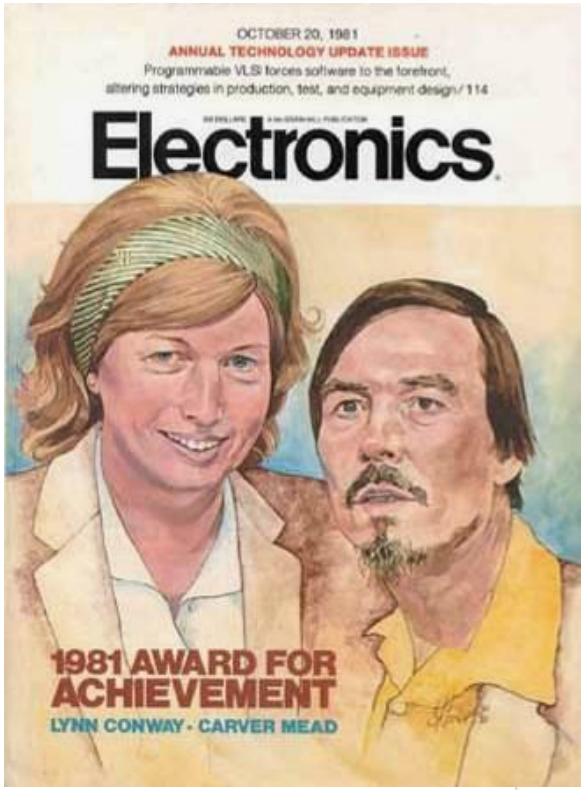
- The industry is going to extraordinary lengths to keep the scaling roadmap moving (due to EUV being late).
  - Additional cost and complexity make scaling less attractive for many market segments.
- Challenge: doing more with mature technologies?

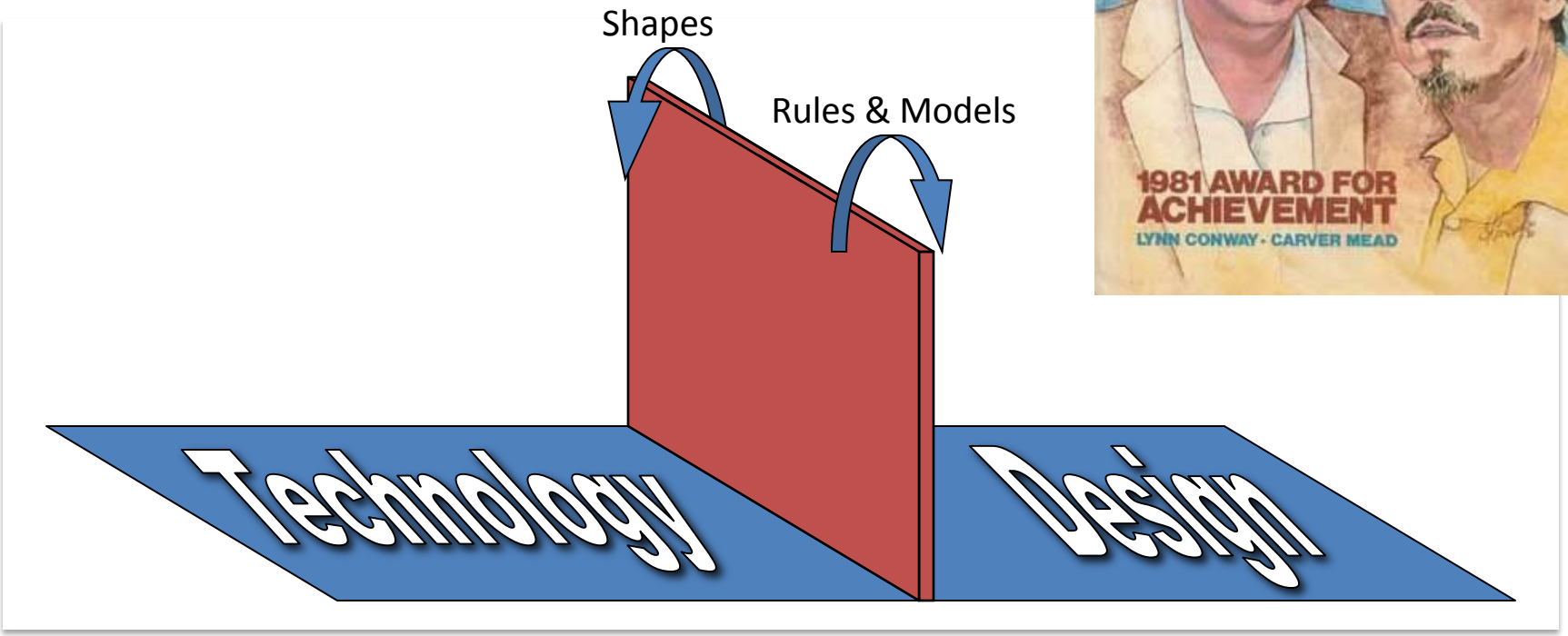




The Rules/Models/Shapes Contract

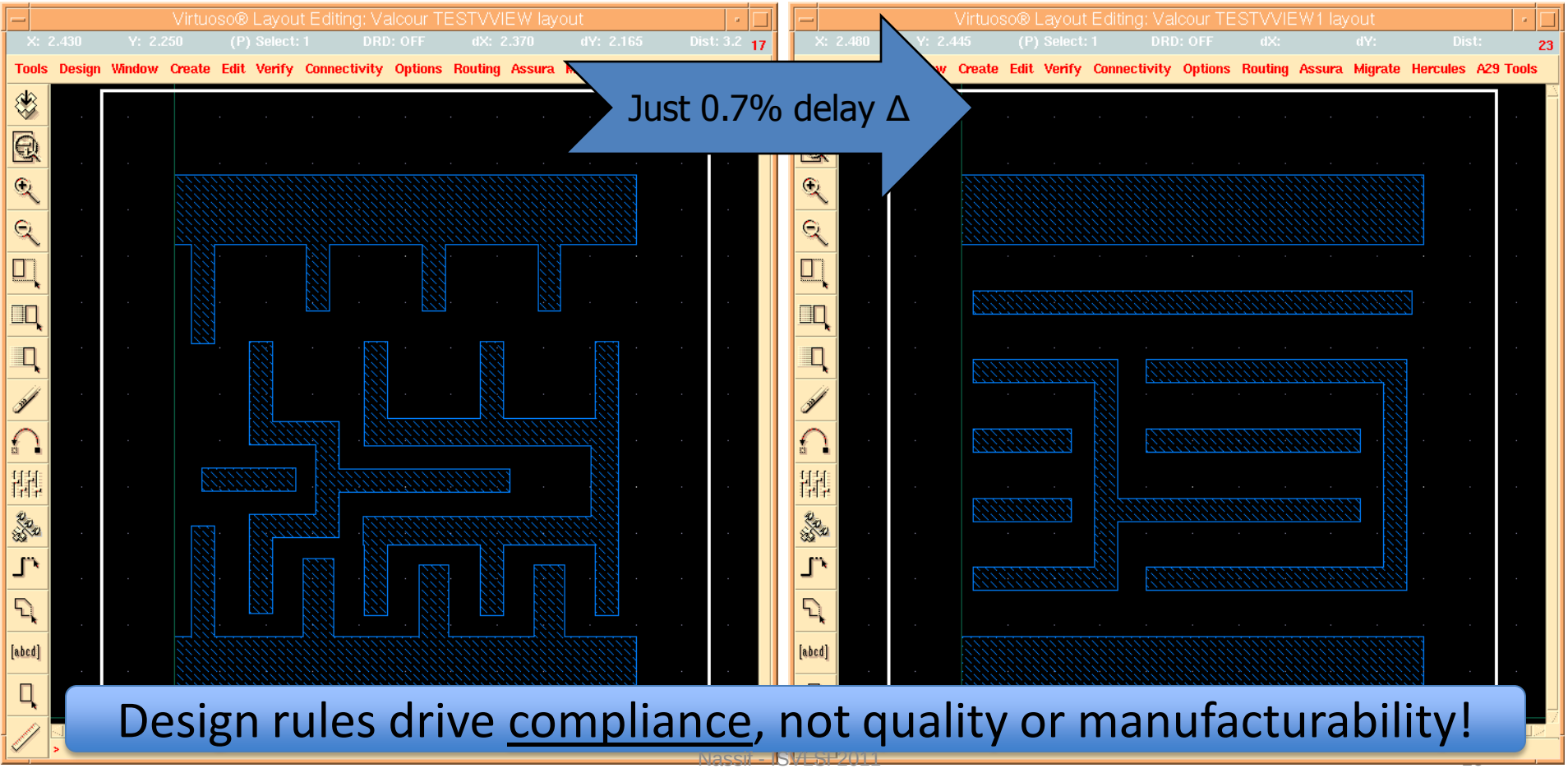
- The classical view, going back to Mead-Conway's work, is fast breaking down at advanced nodes.





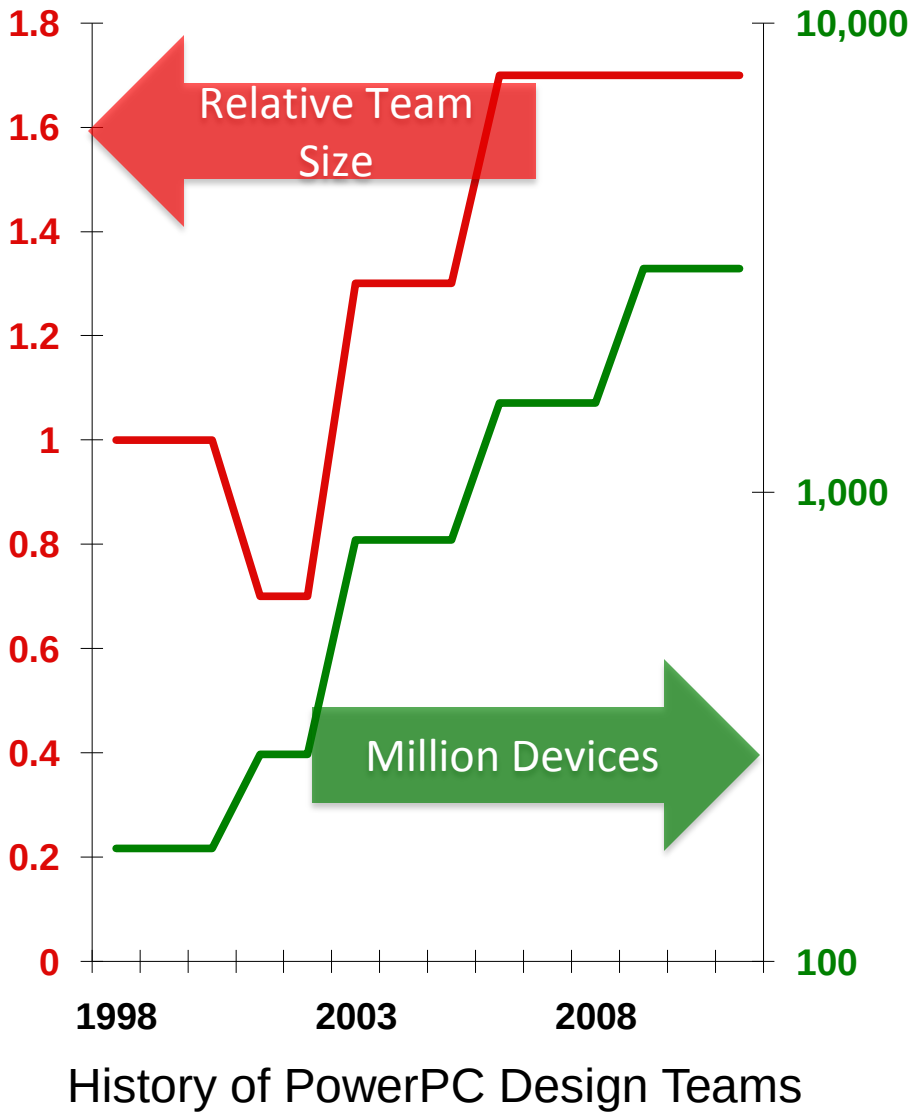
Shapes vs. Design Intent

- 32nm library, local interconnect. Radical changes to improve lithography have zero impact on delay!



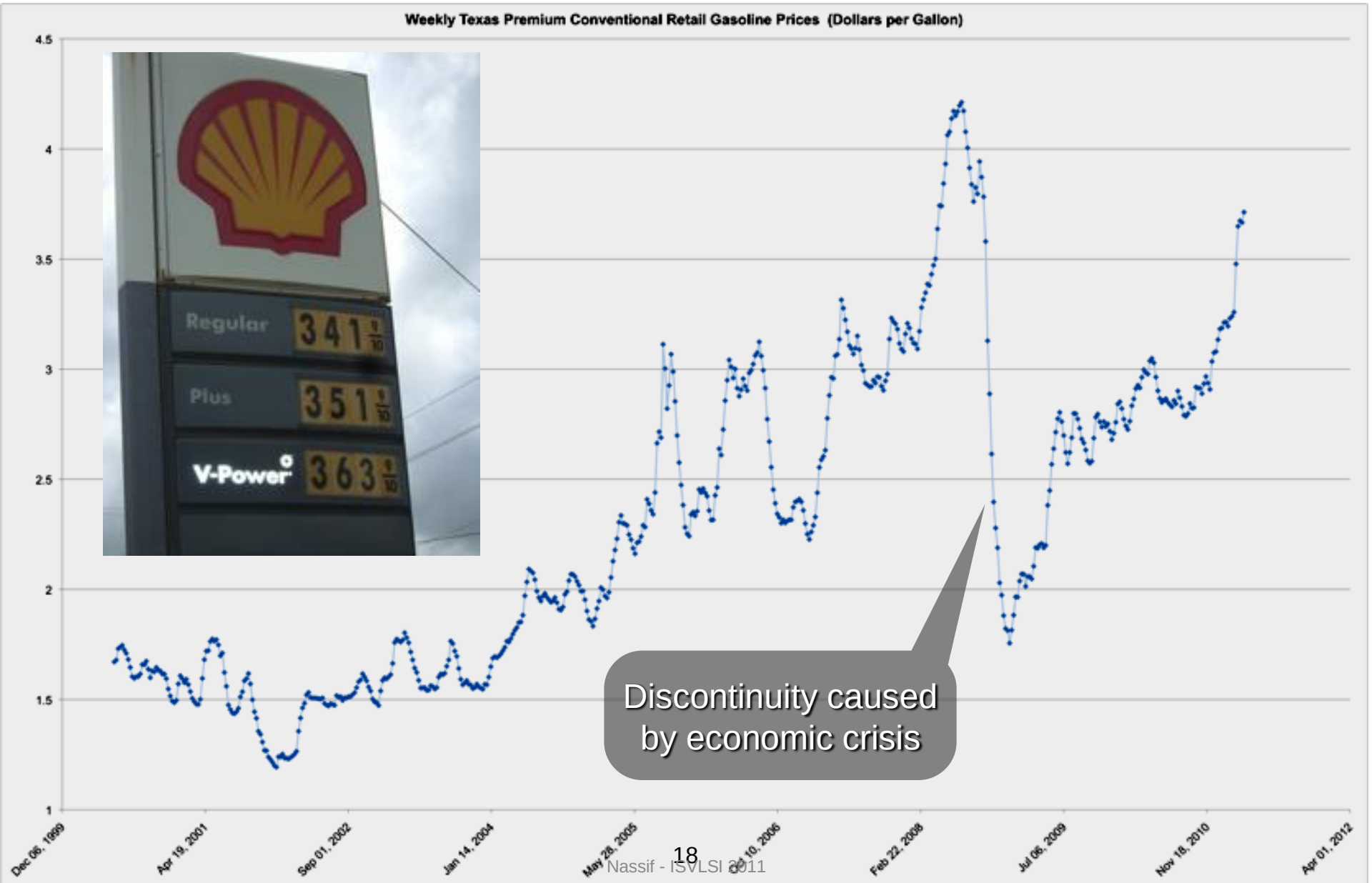
What Should We Do?

- The RMS (**R**ules **M**odels **S**hapes) contract needs to be updated.
  - Cannot legislate better layout with more and more design rules.
  - And need to maintain designer productivity.
- Need better implementation styles that solve this problem without impacting density!
  - And with the same level of automation existing now.
- Why can't this be solved in the context of a cell library?

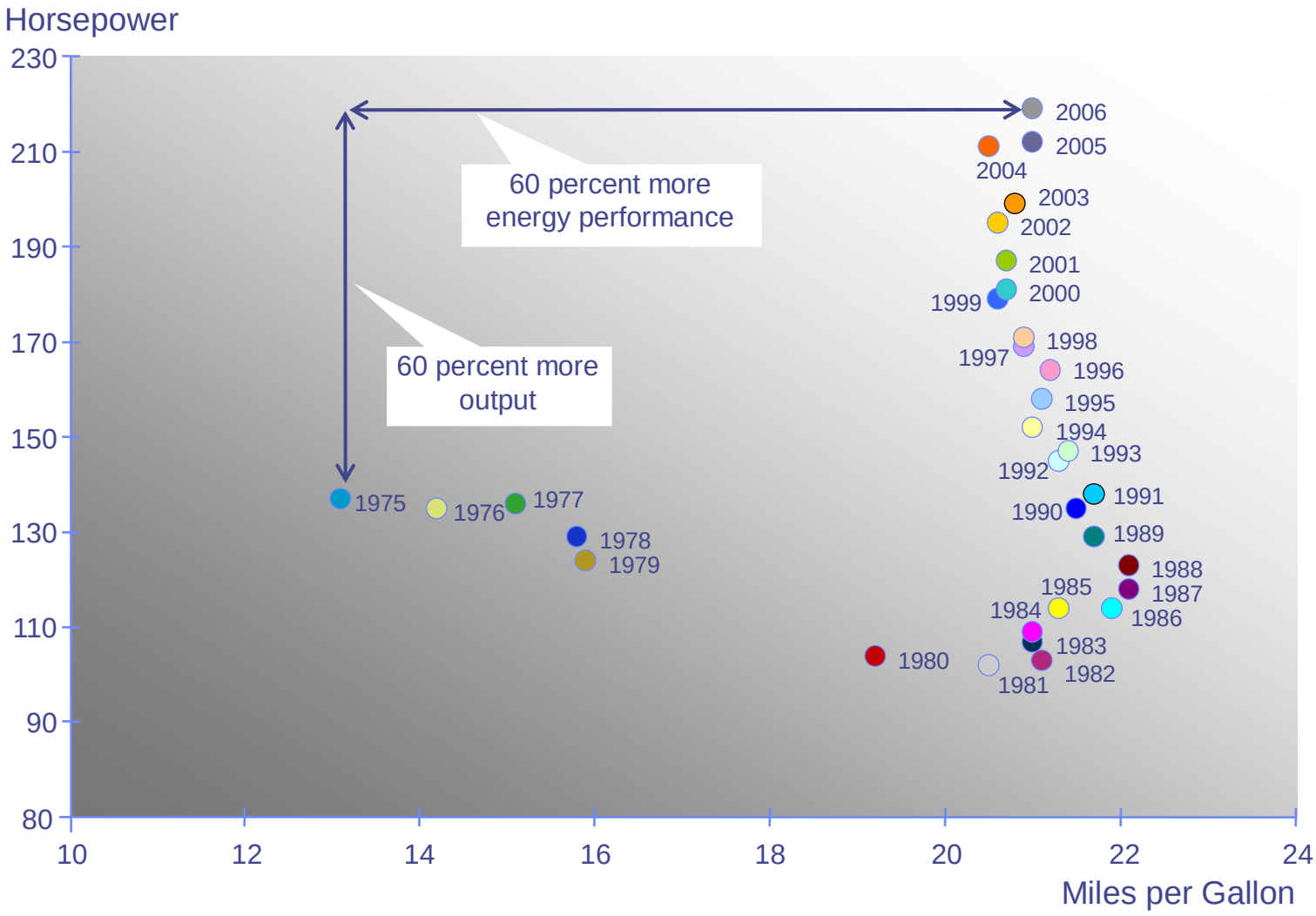


History of PowerPC Design Teams

Gas Prices…

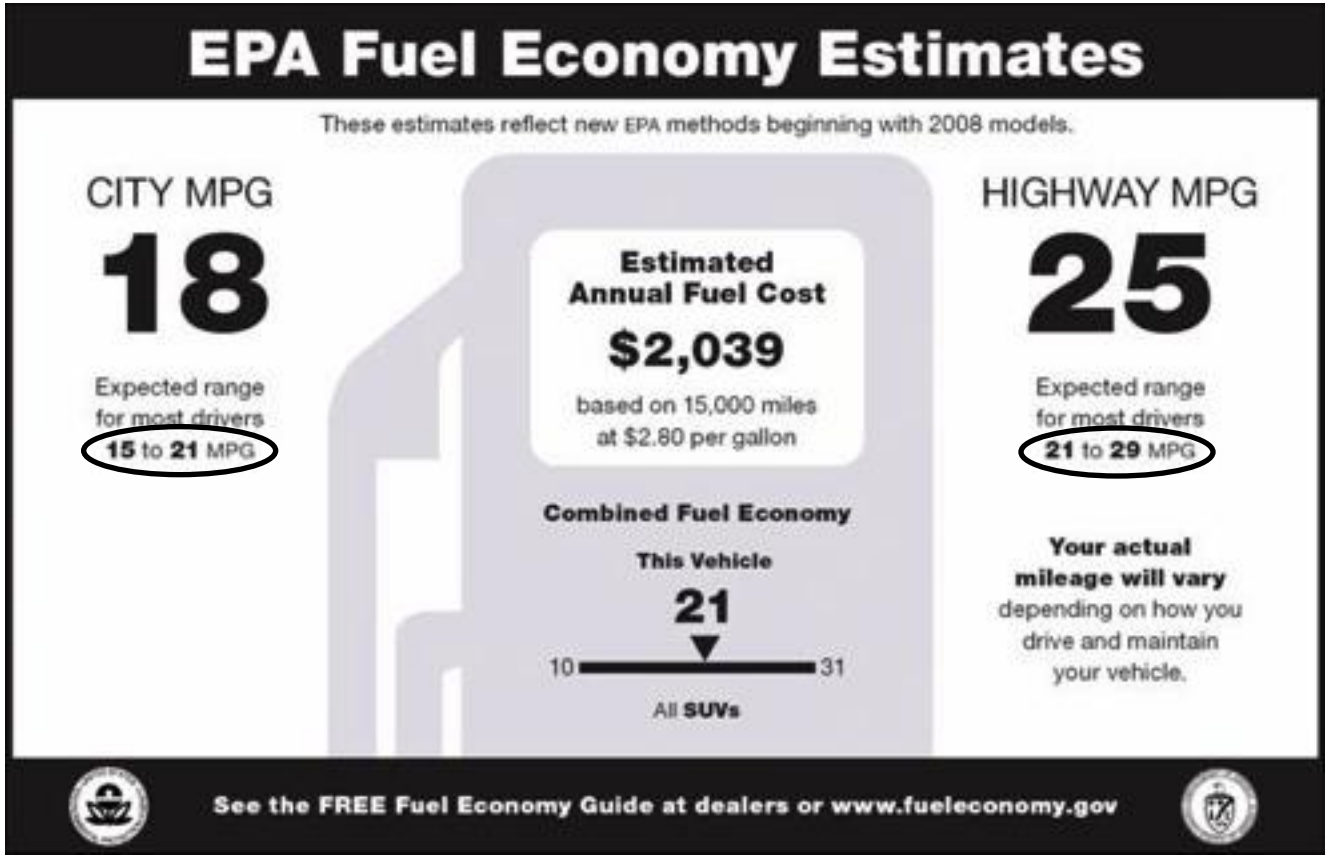

Weekly Texas Premium Conventional Retail Gasoline Prices (Dollars per Gallon)
4.5
4
3.5
3
2.5
2
1.5
1
Regular
341
Plus
351
V-Power
363
Discontinuity caused by economic crisis
Dec 06, 1999
Apr 19, 2001
Sep 01, 2002
Jan 14, 2004
May 28, 2005
Oct 10, 2006
Feb 22, 2008
Jul 06, 2009
Nov 18, 2010
Apr 01, 2012

R&D Impact on Car Efficiency…



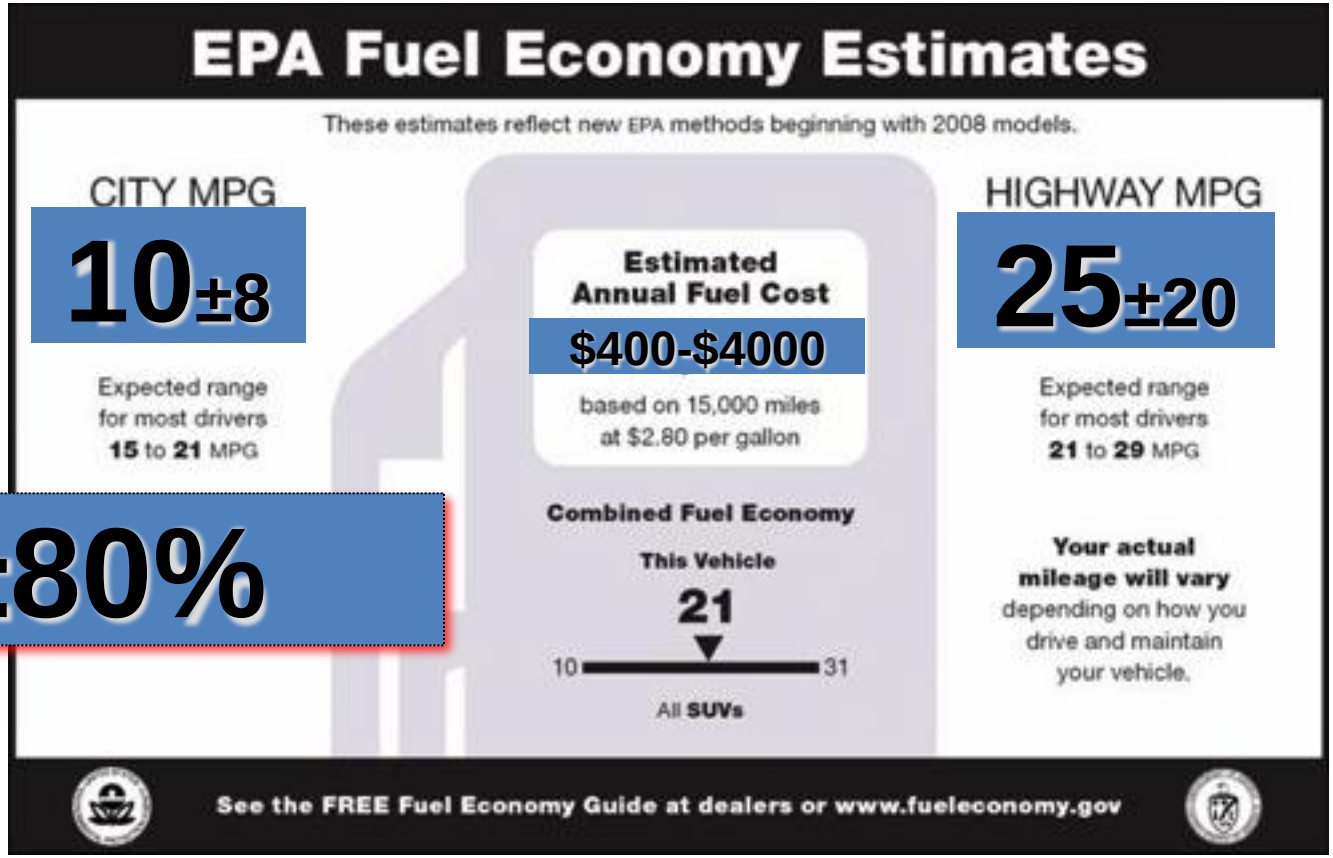
Source: Environmental Protection Agency, Light Duty Automotive Technology and Fuel Economy Trends: 1975-2006, July 2006.

Certainty… A valuable commodity



±16%

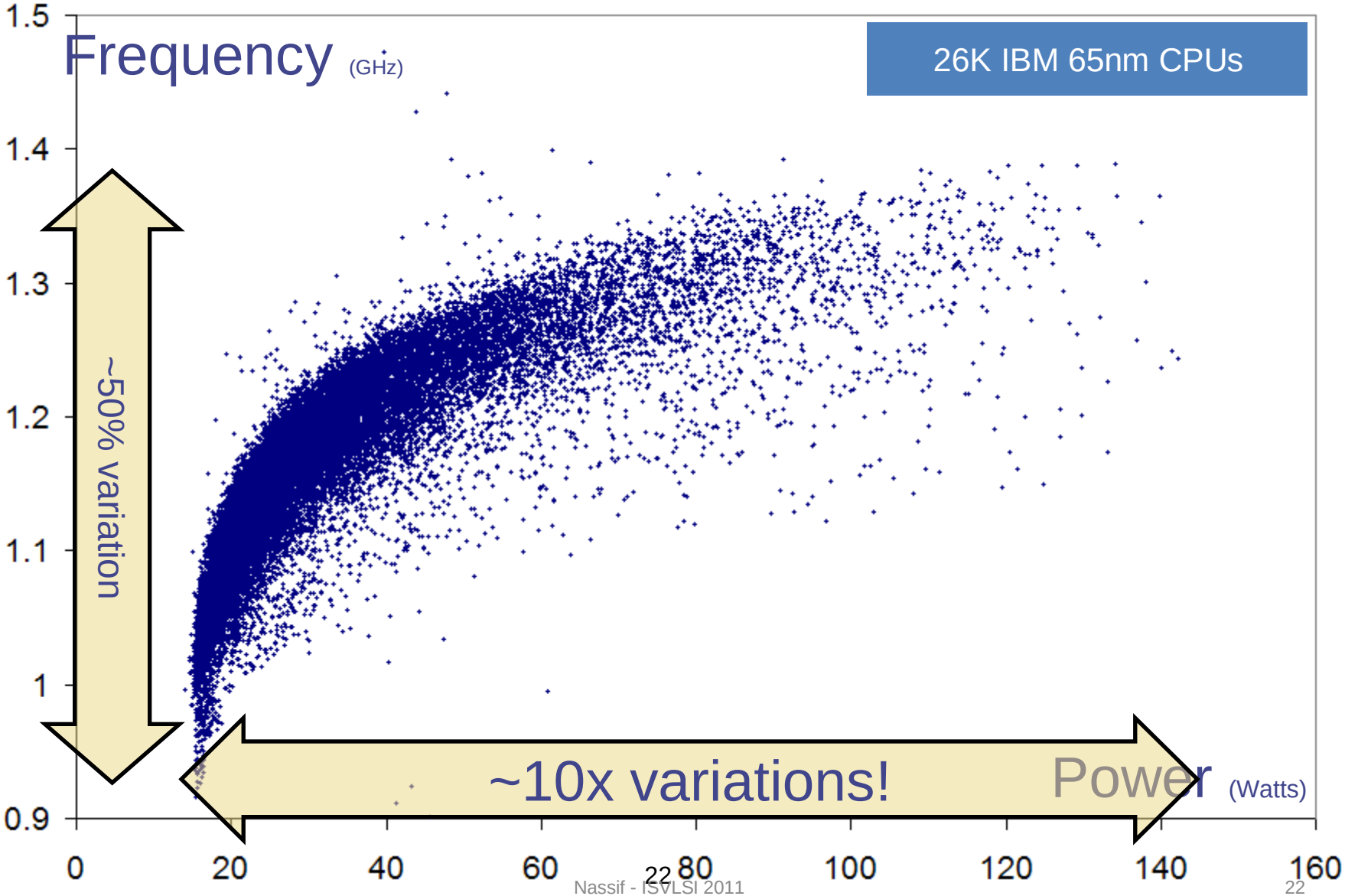
High Variability = Low Quality?



±80%

- Would you buy this car?

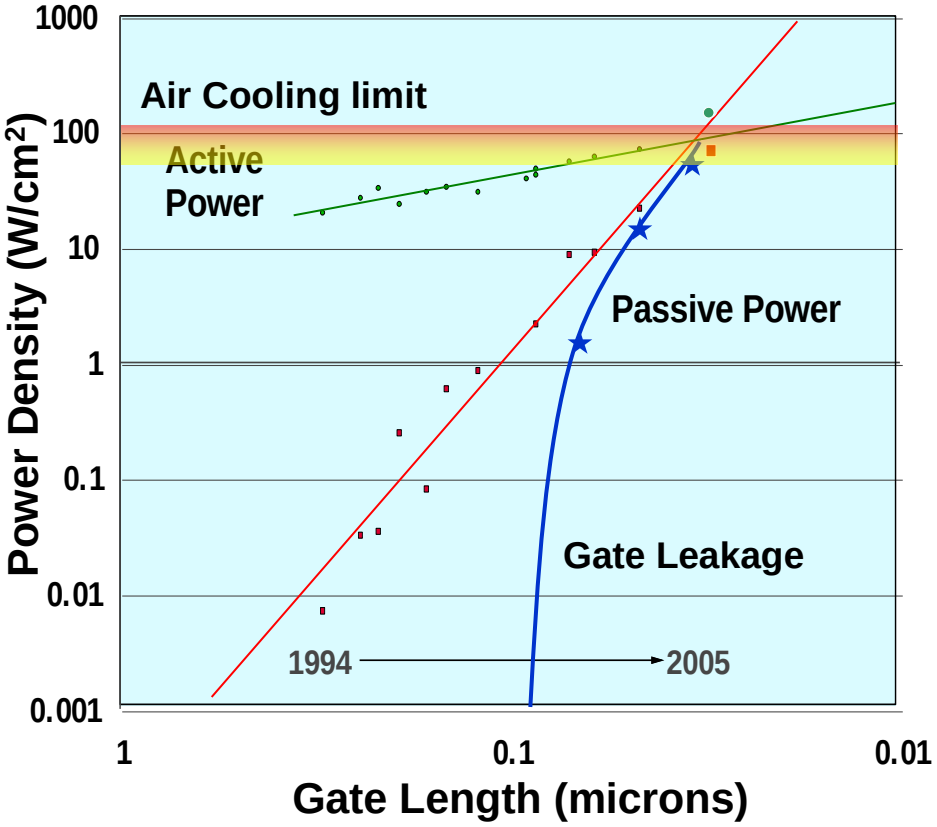
Would you buy this chip then?



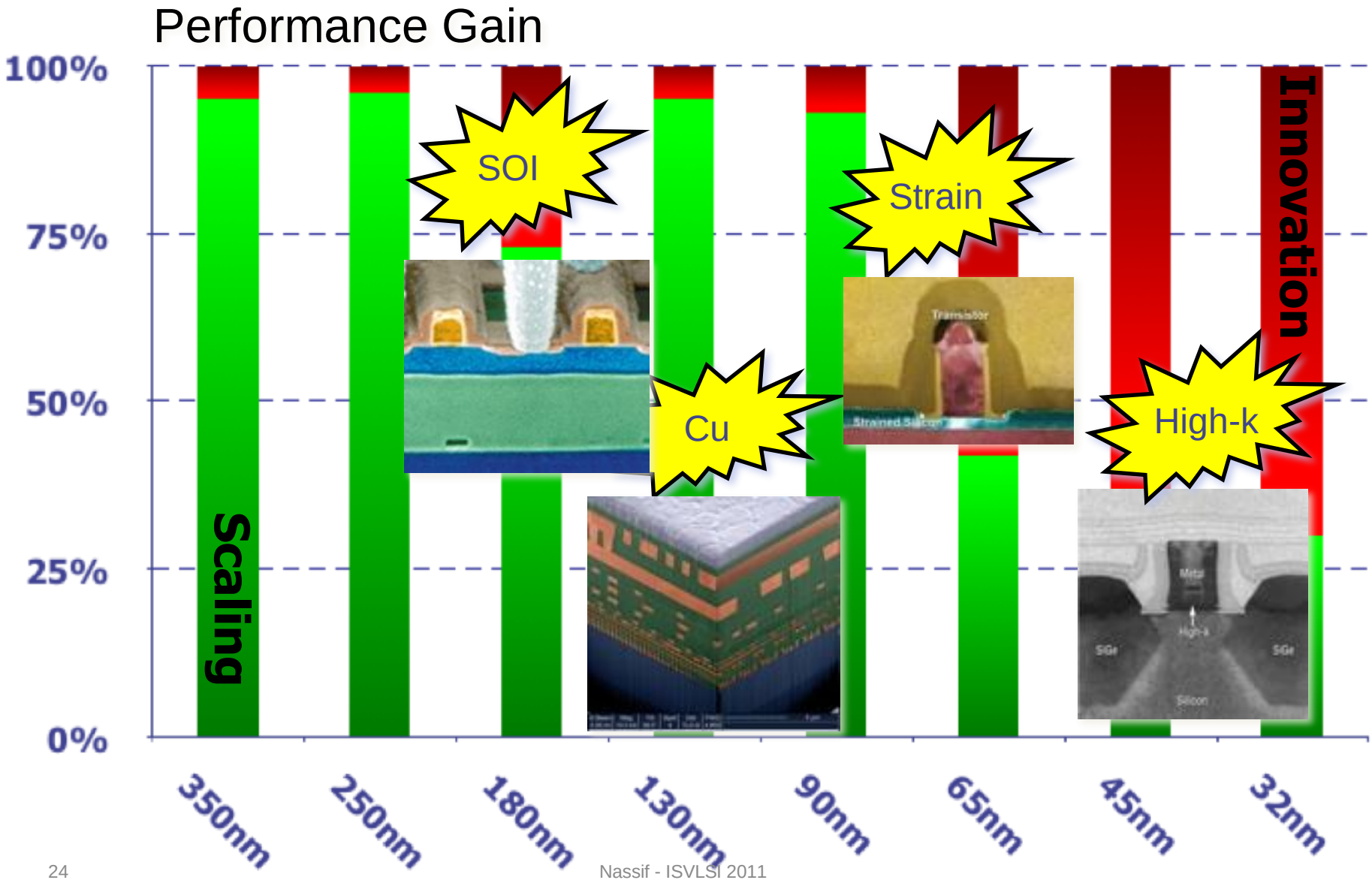
Why So Much Power Uncertainty?

- Passive power is an increasing proportion of overall power, and has higher sensitivity to manufacturing.

- Added complexity in the manufacturing process introducing new sources of systematic variability.
  - Many impacting power strongly.

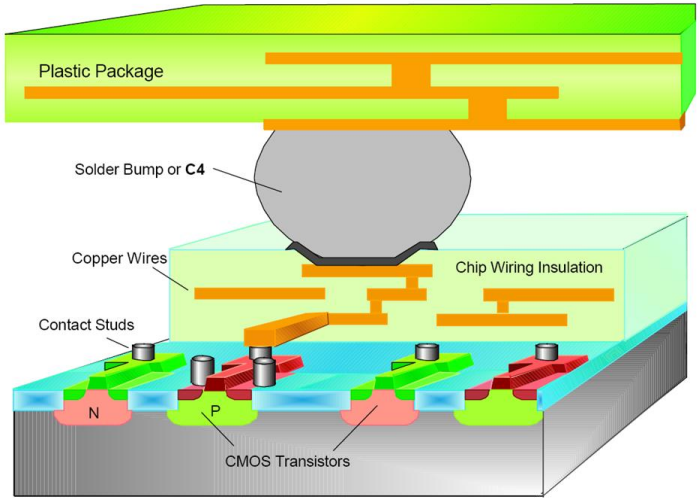


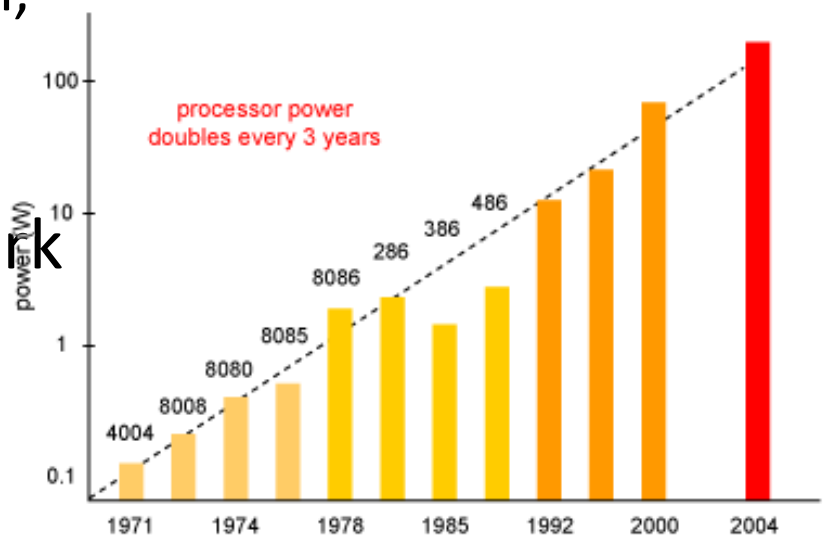
Performance-Driven Complexity



What Should We Do?

- Power should not be a guessing game.
  - Better estimation can drive a variability-aware system-level power reduction methodology.
  - Linkage between high level power and low level electrical (noise, thermal, voltage drop, electro-migration) largely missing!
- One unified stochastic framework to deal with workload and manufacturing variability?

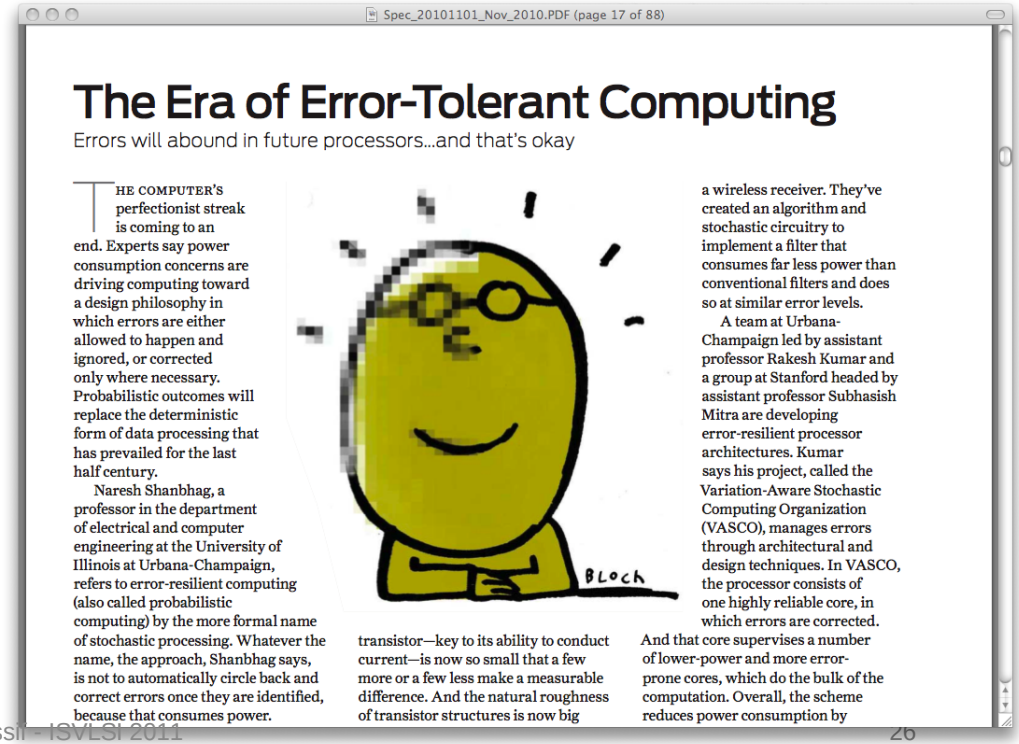




Resilience and Failure?

- As systems become more complex, failure becomes more and more probable.
  - Failure can be because of external factors (noise), aging (metal fatigue) or design (a bug).

- With continued scaling we are at the threshold of a new regime in intrinsic failure rates for semiconductors!



Spec_20101101_Nov_2010.PDF (page 17 of 88)

The Era of Error-Tolerant Computing

Errors will abound in future processors...and that's okay

THE COMPUTER'S perfectionist streak is coming to an end. Experts say power consumption concerns are driving computing toward a design philosophy in which errors are either allowed to happen and ignored, or corrected only where necessary. Probabilistic outcomes will replace the deterministic form of data processing that has prevailed for the last half century.

Naresh Shanbhag, a professor in the department of electrical and computer engineering at the University of Illinois at Urbana-Champaign, refers to error-resilient computing (also called probabilistic computing) by the more formal name of stochastic processing. Whatever the name, the approach, Shanbhag says, is not to automatically circle back and correct errors once they are identified, because that consumes power.

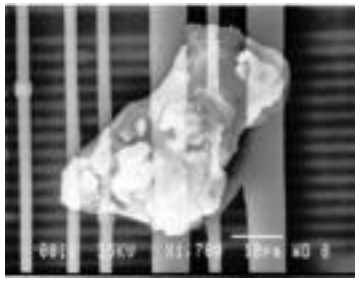
transistor—key to its ability to conduct current—is now so small that a few more or a few less make a measurable difference. And the natural roughness of transistor structures is now big

a wireless receiver. They've created an algorithm and stochastic circuitry to implement a filter that consumes far less power than conventional filters and does so at similar error levels.

A team at Urbana-Champaign led by assistant professor Rakesh Kumar and a group at Stanford headed by assistant professor Subhasish Mitra are developing error-resilient processor architectures. Kumar says his project, called the Variation-Aware Stochastic Computing Organization (VASCO), manages errors through architectural and design techniques. In VASCO, the processor consists of one highly reliable core, in which errors are corrected. And that core supervises a number of lower-power and more error-prone cores, which do the bulk of the computation. Overall, the scheme reduces power consumption by

Failures Types

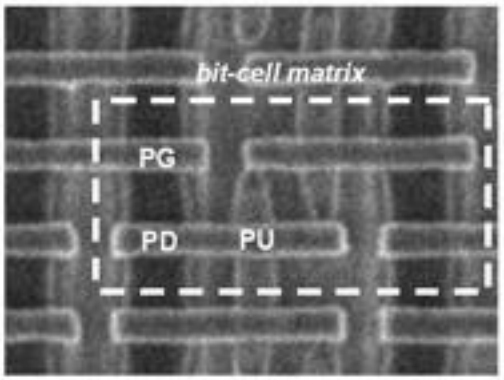
- Failures in integrated circuits can be of two varieties:
- HARD failures are characterized by a permanent mismatch between expected and observed behavior.
  - An output is stuck at 0 or 1!
  - Cause: often a topological change.

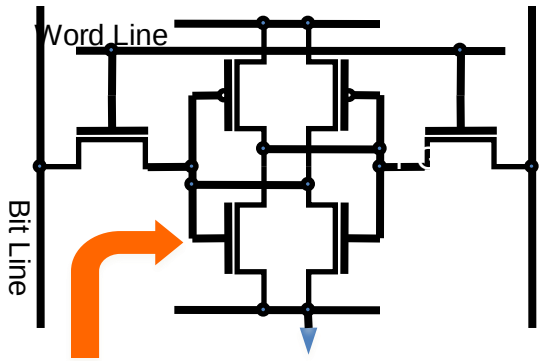


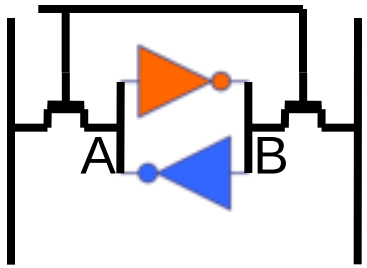
- SOFT failures occur occasionally or under a specific set of conditions.
  - Excessive power at high T.
  - Cause: often a mismatch between models and reality.

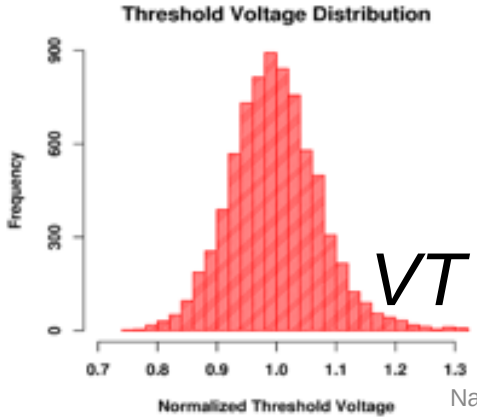
The Merging of Soft and Hard

- For the smallest devices in extremely scaled designs, hard and soft failures are indistinguishable!
  - Post child: SRAM!

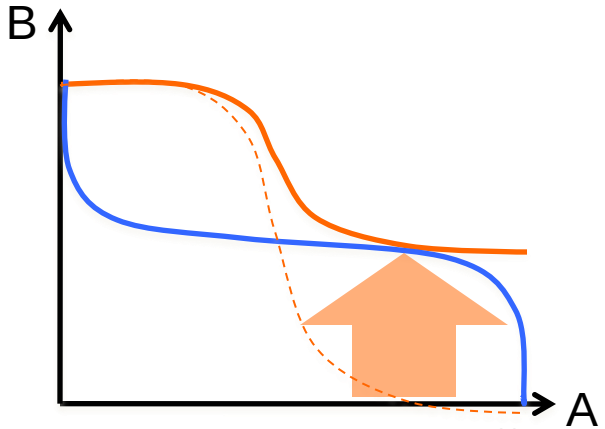




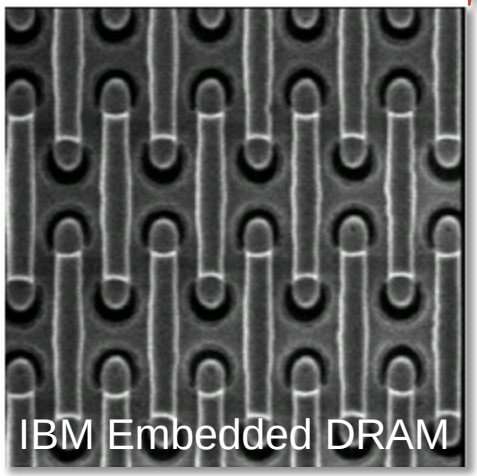




VT Variability



Memory Observations

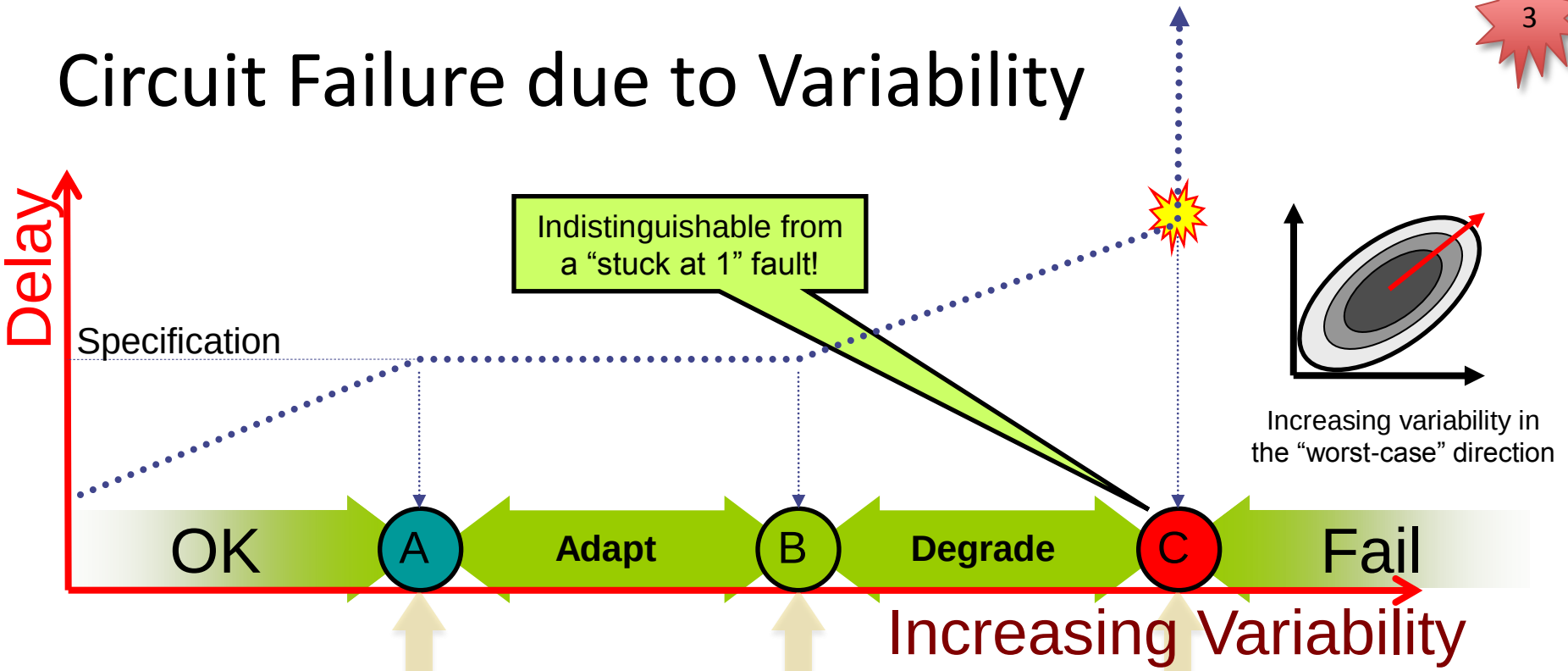


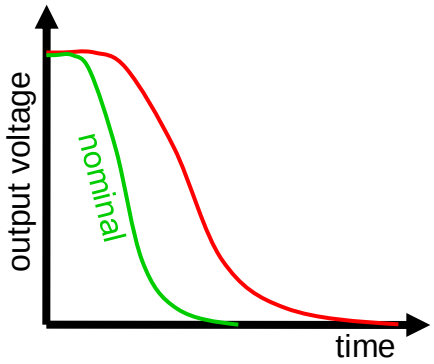
- Since Memory is a uniform fabric, it can "push" the design rules for density.
  - The benefit of regularity and engineering.
- The price for higher density is higher variability.
  - So high, that a memory cells can have "catastrophic" failure due to parametric variability, e.g. cell loses data when we read it.
- But there are solutions:
  - Rely on the fact that memory is in an array!
  - Add redundancy and error correction at the architecture level.
  - Aggressively predict and manage the electrical design margins for SRAM at the circuit/device levels.

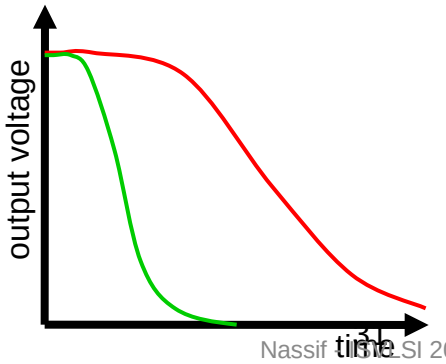
Current Work on Manufacturing Variability

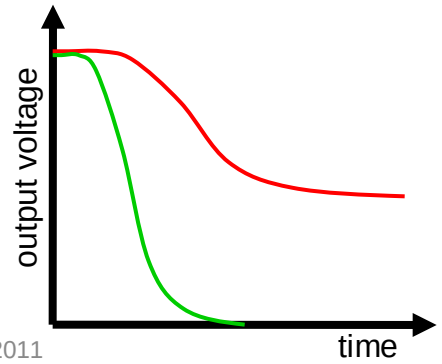
- Much current work is focused on the short term impact.
  - Examples: SSTA, SRAM analysis, Analog circuit yields, etc…
- Impact of increasing variability will change in character!
  - And will need to be handled at higher levels of design!
- 1st order impact: delay and power variability (at 90nm).
  - Widely published, numerous EDA tools available.
- 2nd order impact: reliability of devices and wires (at 45nm).
  - Less widely published, more open problems (e.g. EM).
- 3rd order impact: resilience (at and beyond 22nm).
  - Emerging currently!

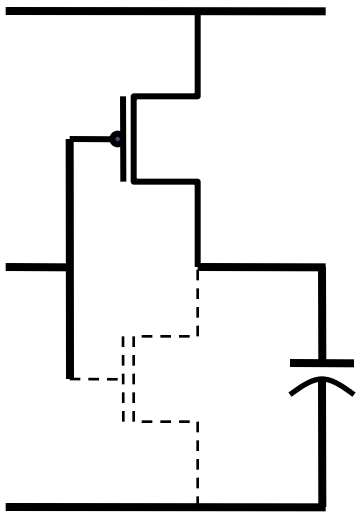
Circuit Failure due to Variability



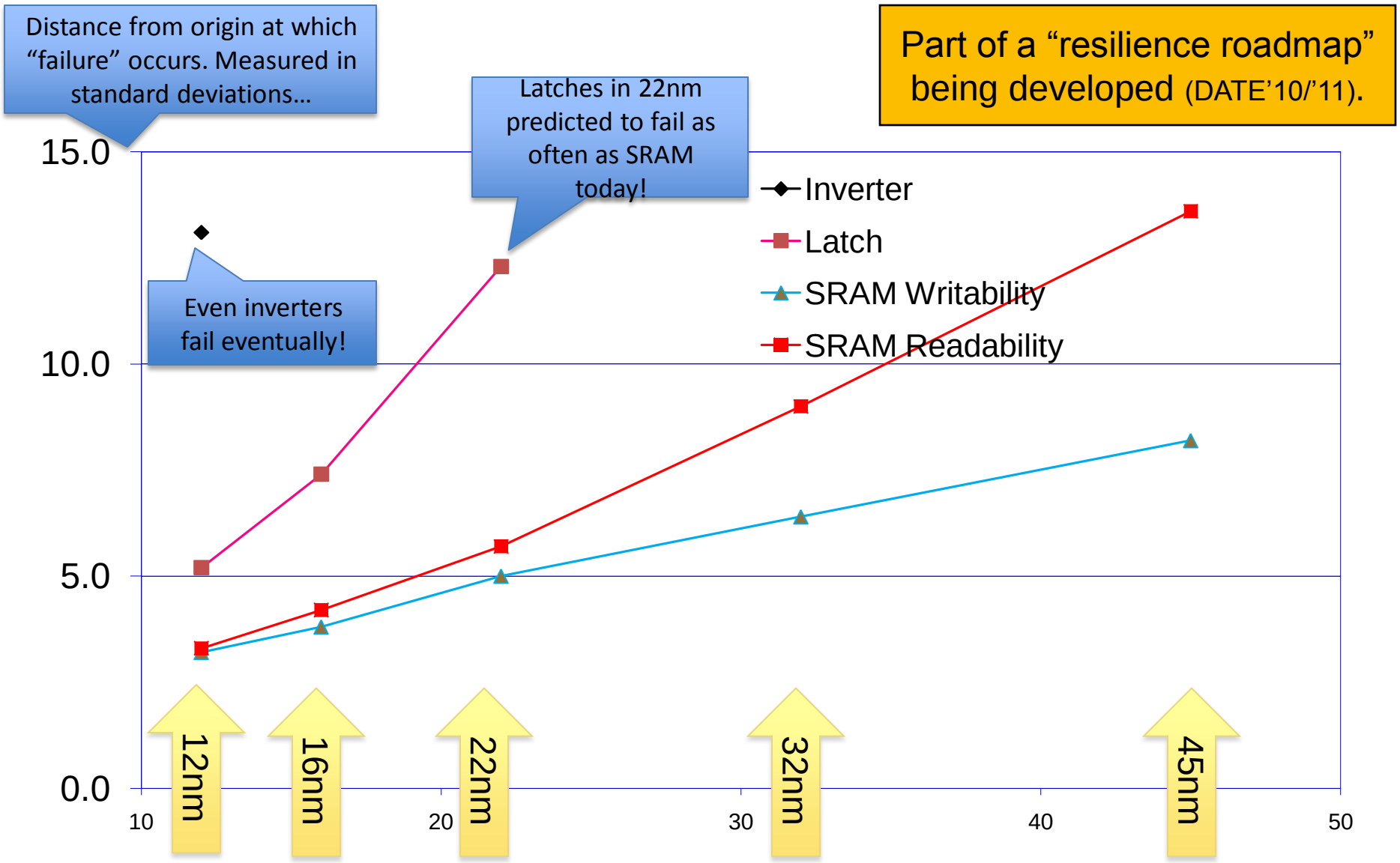






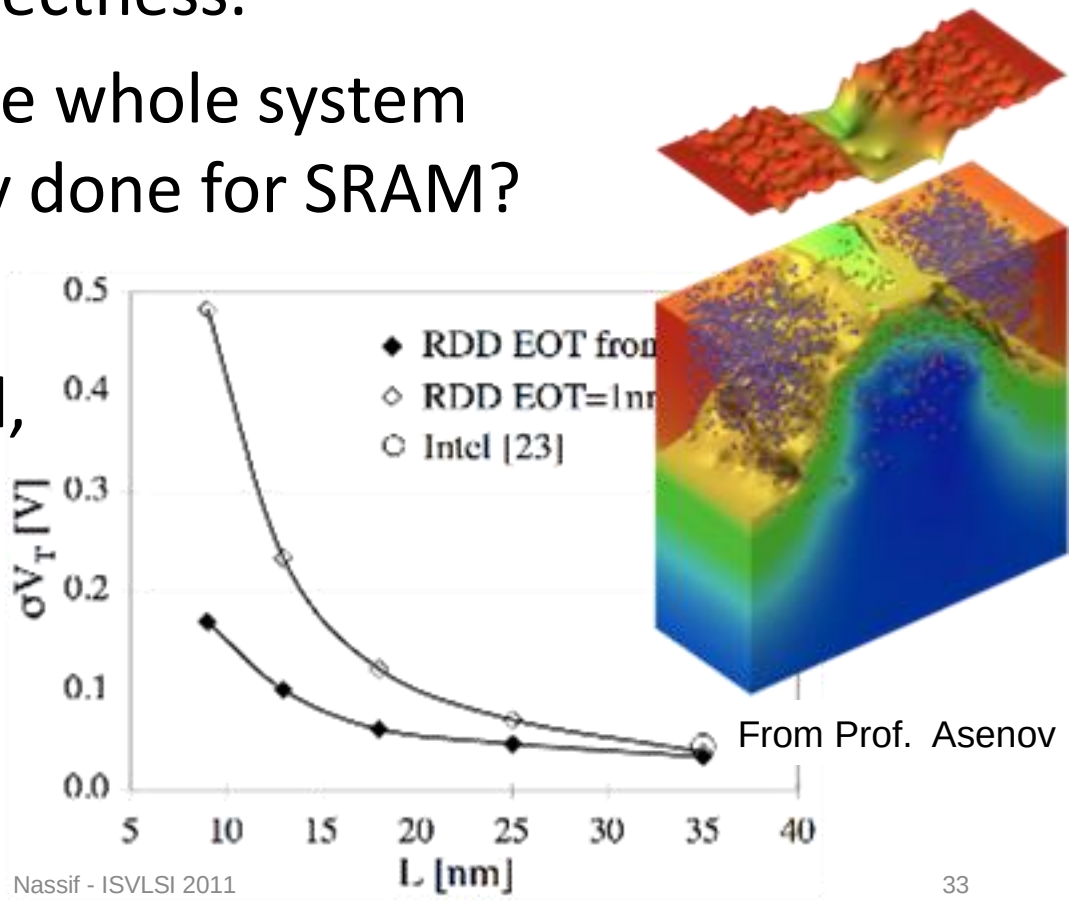


Recent Results (part of ITRS'10)



What Should We Do?

- There are be some technology fixes (like FD-SOI) that can help a little, but increasing variability will impact circuit operation/correctness.
- How can we do for the whole system what we have already done for SRAM?
- It must be automated, and it must be verifiable!



From Prof. Asenov

Review and Summary

- Three problem areas that need attention:

Lithography.

  - We must rationalize the design/manufacturing interface, and create design implementation methods which are inherently manufacturable and verifiable..

Power.

  - Design tools and methodologies need to evolve to treat power as the first class and possibly only constraint (as opposed to performance).

Resilience.

  - We must apply to logic and control circuits what we've done for memory in order to deal with widespread and frequent device failures.

What is at the end of the Tunnel?

- We want to make sure that Silicon Technology will continue to amaze and deliver for at least until I can retire .
- There are numerous research and development problems to tackle.
  - Silicon careers still matter!
- Silicon will be the substrate on which any future post-Si technology will be built.
  - We must sustain Si until an alternative emerges.