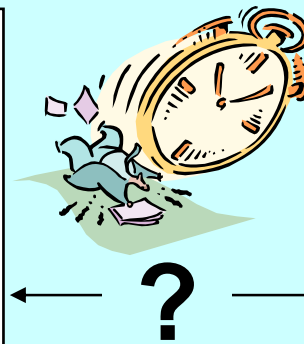


Bridging the Gaps between System Level Complexity and Nano-Scale Uncertainties

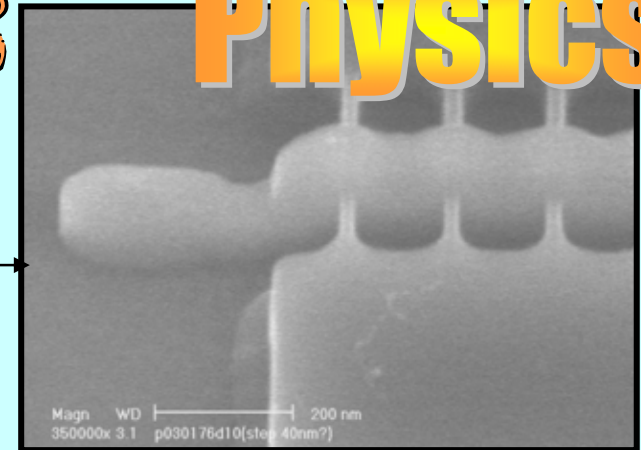
Society



Systems



Physics



Devices

Hugo De Man
Prof. K.U.Leuven
Senior Research Fellow IMEC
deman@imec.be

Design Metrics for Ambient Intelligent Digital Devices (isscc'05, 1.2)

- **Power Efficiency (PE): 10 to 200 GOPS/Watt**
- **Part cost < 10€ ⇔ NRE cost > 50M€**
 - PE two-orders-of-magnitude > GP microprocessor at 1/20 th of part cost
- **Flexible by (downloadable) embedded software**
 - From Hw centric ASIC to **Sw centric PLATFORM**
- **Real-time stream based processing (MM, SDR)**
 - **Memory intensive (> 70% area, power)**
 - **Not min gate delay (ps) BUT max throughput/Watt (Gbit/s;op/s)/Watt**

Min Gate delay vs. Max Throughput

- **Min Tr delay (min $C_g \cdot V_{DD} / I_{on}$) :**

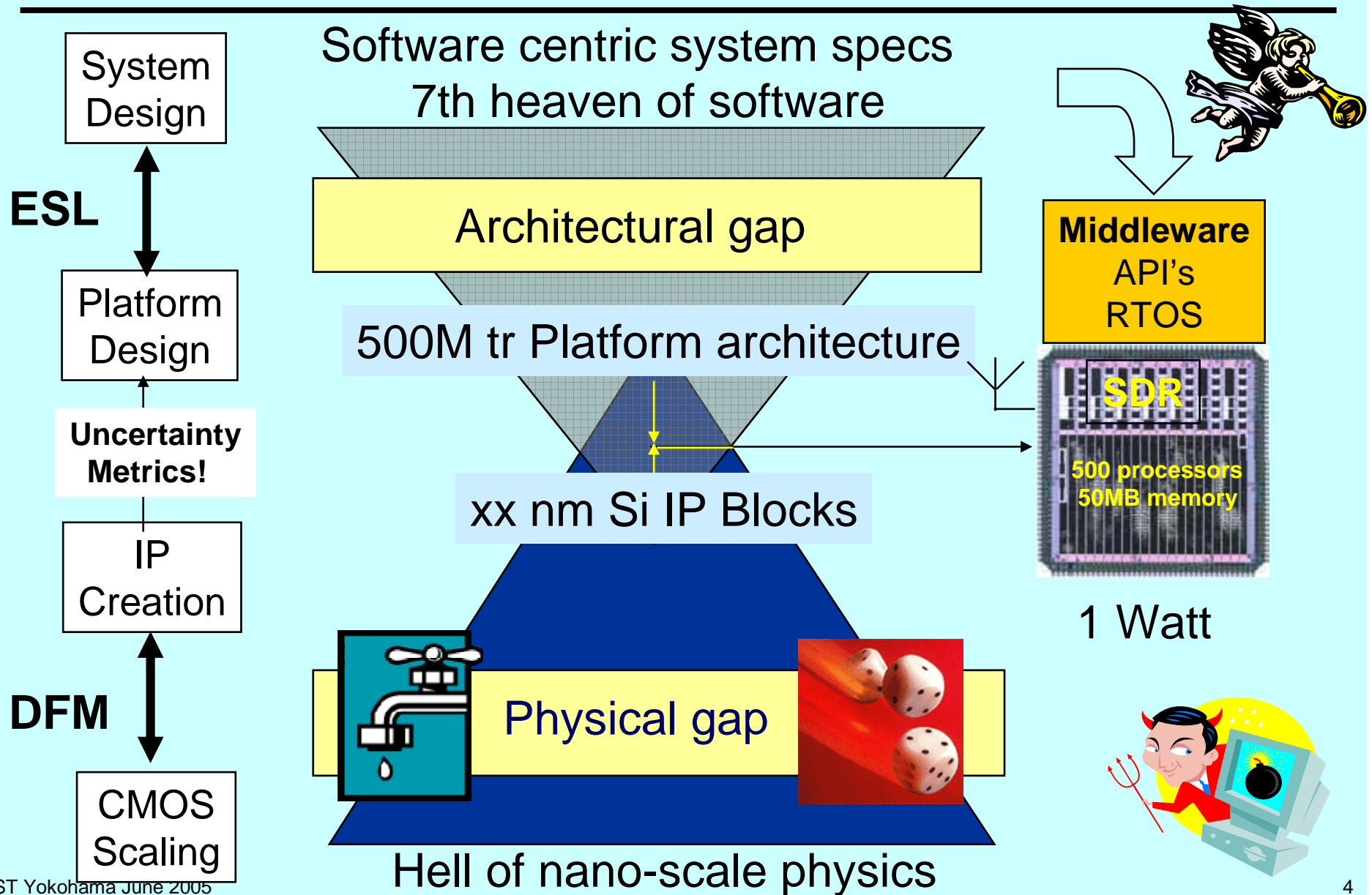
- A **device issue** (scaling, materials, device architecture)
- Scaling justification is in **cost reduction** rather than improvement of 'gate performance'

- **Max Throughput / Watt**

- A **micro-architectural issue** (sw, //, interconnect, storage hierarchy)

Joint research crucial for further progress!

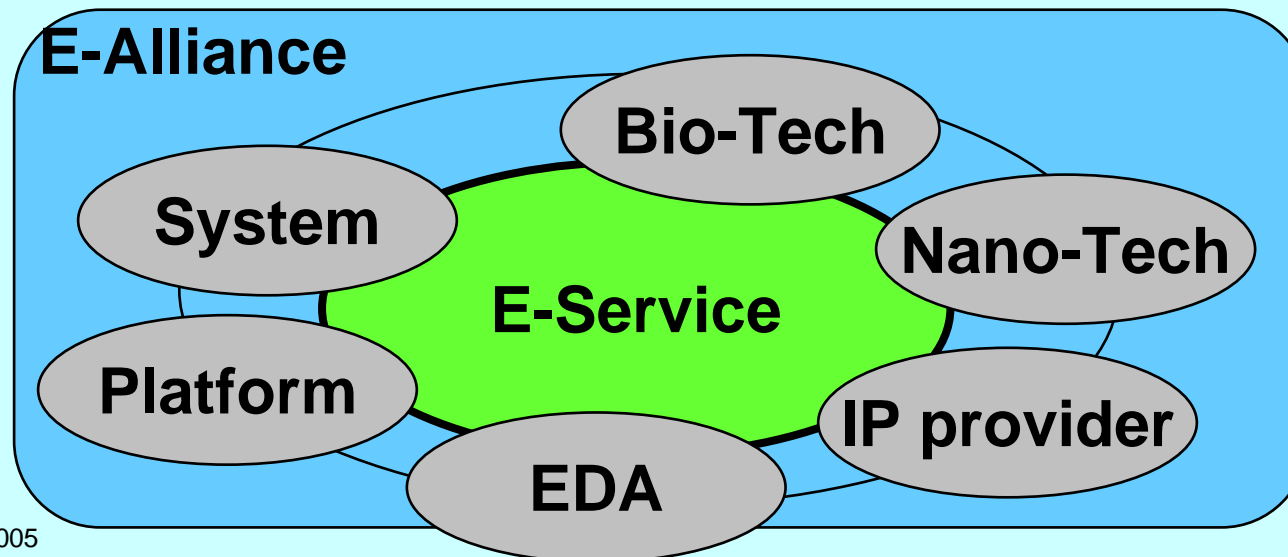
Two Growing Gaps ...



Bottom Line...

We can no longer afford a living-apart-together relationship between system design and process development.

More Moore will depend on forming **VIRTUAL VERTICALLY INTEGRATED R&D ALLIANCES** rallied around strategically important e-services for society



1. The Architectural Gap

“Managing Giga-Complexity ”

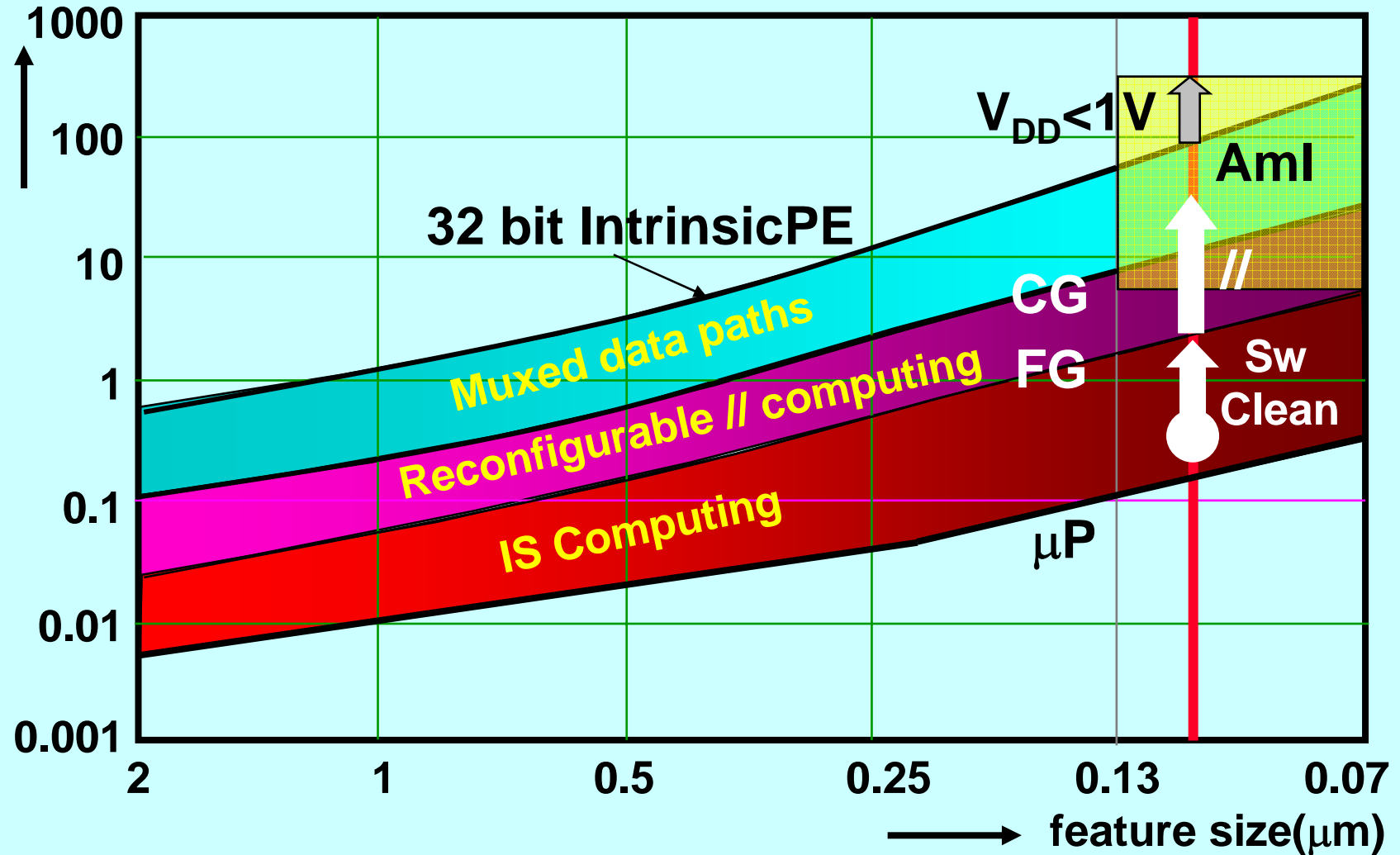
Create SL methods, tools and **skills** for designing flexible, yet power-efficient platforms
= 1 B uncertain devices + >10M lines of code
for mapping real-time **MULTIPLE S/W** applications on them **at lowest NRE cost**

Greatest Challenge:
Overcome Power-Flexibility Conflict
While coping with CMOS (r)evolution

Power-Flexibility Conflict

Power Efficiency (GOPS/Watt)

[T.Claasen isscc'99]



How to Reconcile PE and Flexibility?

- **Cleaning software for low power**

- Minimize # memory transfers, operations = energy
- Regularize, vectorize code, min. data-size
- Factor 5 to 10 lower power but **tools and skills** needed

- **Exploit parallelism to reduce power**

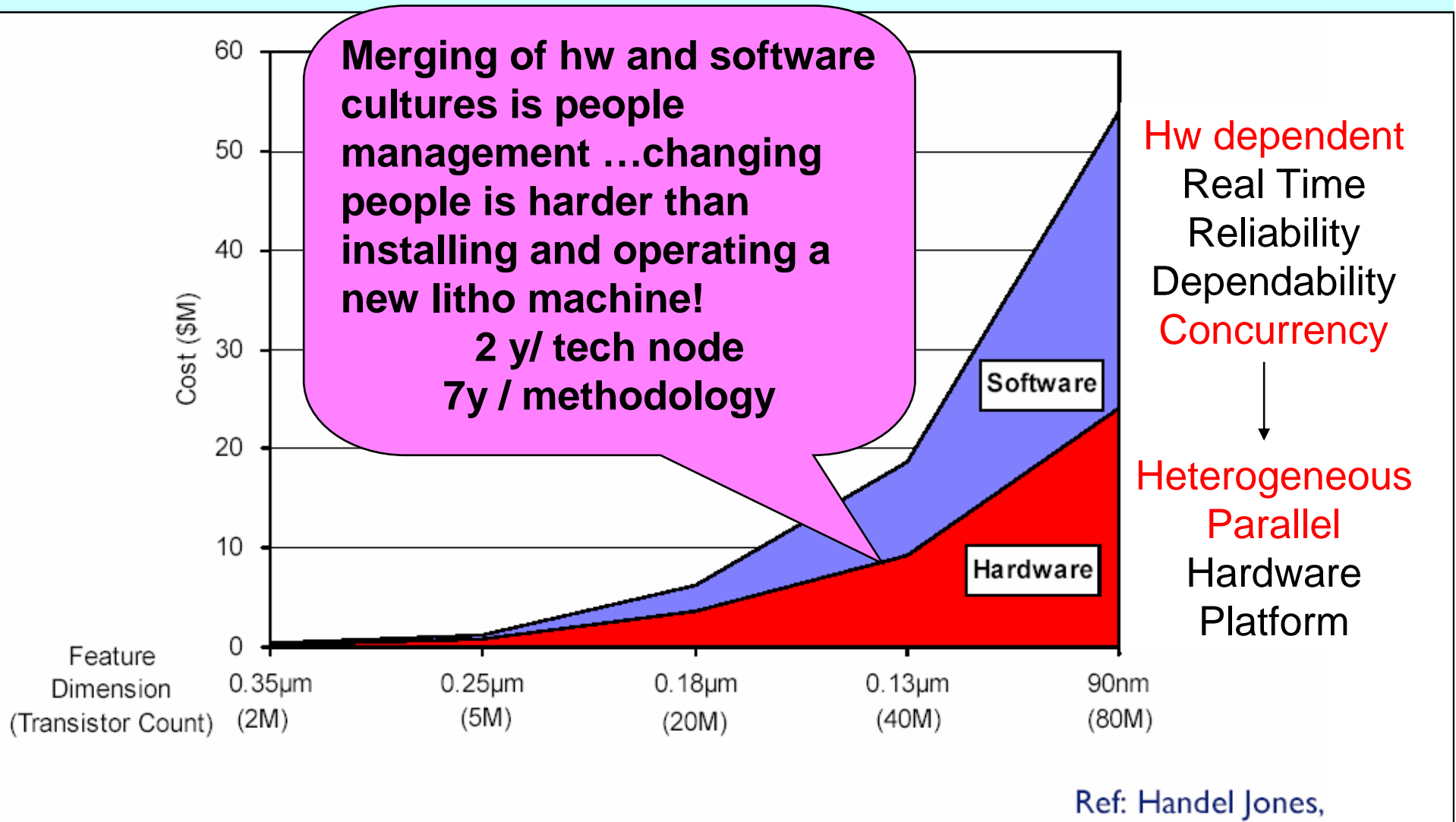
- Better to use 20 processors @ 100MHz than 1 @ 2Ghz
- But urgent need for design **methods, tools and skills** to support Hw/**Sw** co-design (devil is in the software!)

- **Exploit task level dynamism**

- Just enough V_{DD} , F_{cl} , I_L

The Devil is in the Software and the NRE cost...

Software productivity increase 5.8%/y \leftrightarrow 58%/y Moore



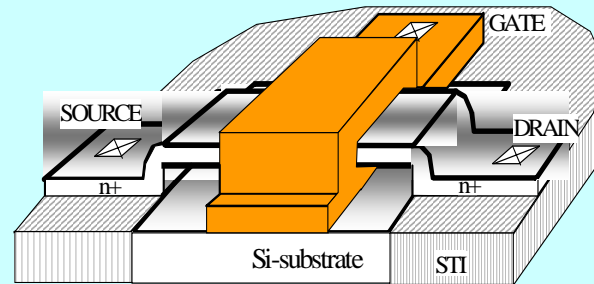
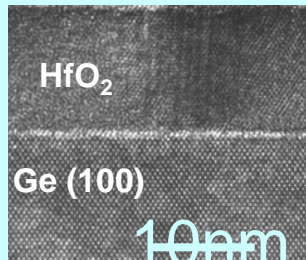
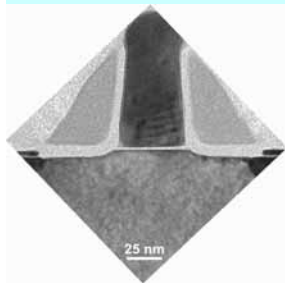
2. Hell of Nano-Scale Physics



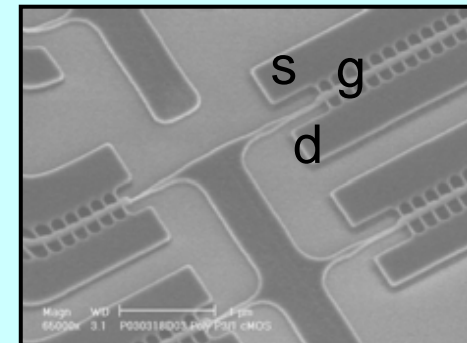
- **Leakage power starts to dominate**
 - larger gate delay than scaling for performance predicts
- **Voltage headroom shrinks**
 - makes A&RF guys deeply worried (sub 1 Volt circuits)
- **Interconnect claims the first role...**
 - challenging timing, power, synchronism, signal integrity
- **Increasing device variability**
 - jeopardizes predictability and yield, affects design process
- **New device architectures needed <45nm**
 - Impact layout style, need for DFM and IP library design

“NewCMOS” Device Architectures below 65nm

- Gate leakage: **Hi-k, metal (fusi) gates**
- SD leakage: **Hi VT, M-VT \Rightarrow Hi-VDD**
 \Leftrightarrow dyn power, variability
- Mobility keeper: **Strained Si, (Ge...)**
- Ion boost: **Multi-Gate, FINFET, DGFD SOI**



Emerald DGFD SOI
STM



Finfets, IMEC

2010

2005

BUT CMOS diversity must percolate to system level...

Different CMOS for Different People...

For Microprocessor, Servers

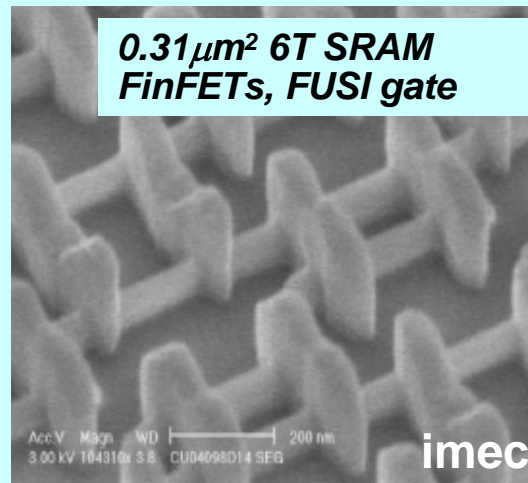
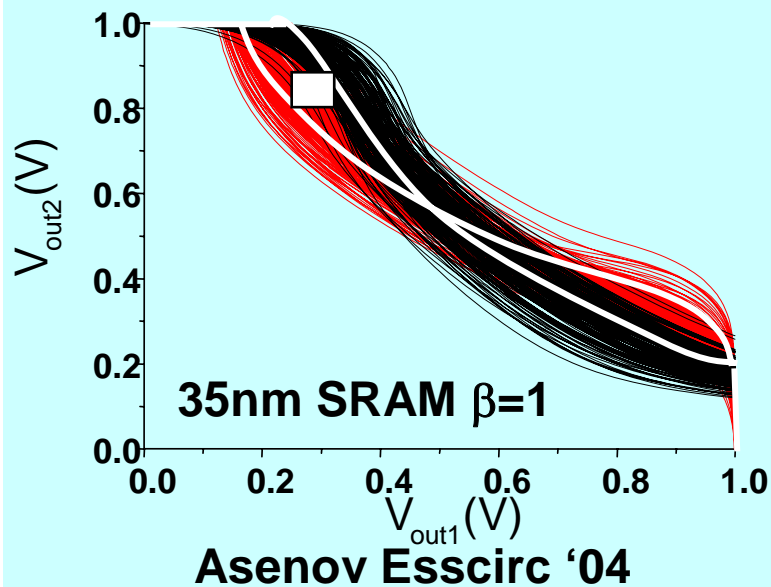
@ 100 Watt active power, 20W leakage acceptable
⇒ low V_{TH} , high I_{on} , ABB to manage leakage yield

BUT for Aml...

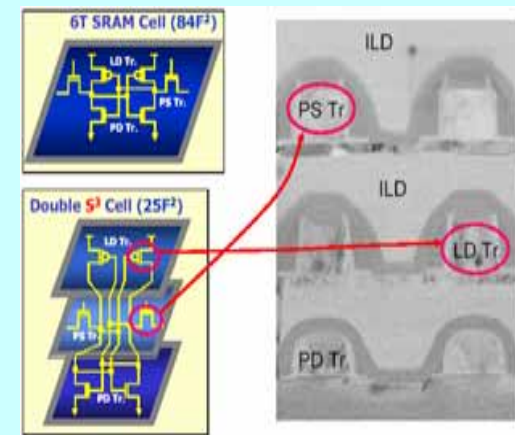
Leakage mWatt nodes <10mW, μ Watt nodes <10 μ W
⇒ Scaling for low leakage at expense of gate delay
⇒ More GOPS from more transistors \Leftrightarrow faster ones
⇒ Wires more important anyway...change layout style

Variability Impact

- Adds a new dimension to design space (P-A-T-Y)
- Statistical design, device sizing for yield crucial
- SRAM first victim...VDD scaling problematic
- Below 65nm new device architectures needed



45nm FinFet
SRAM IEDM'04



3D SRAM
Samsung isscc'05




But System Design must change...

Better than worst case design else lose
advantage of scaling

Design reliable systems from uncertain
components

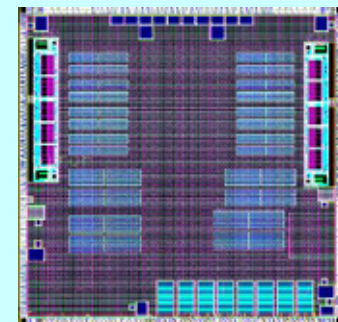
Allow errors to happen!





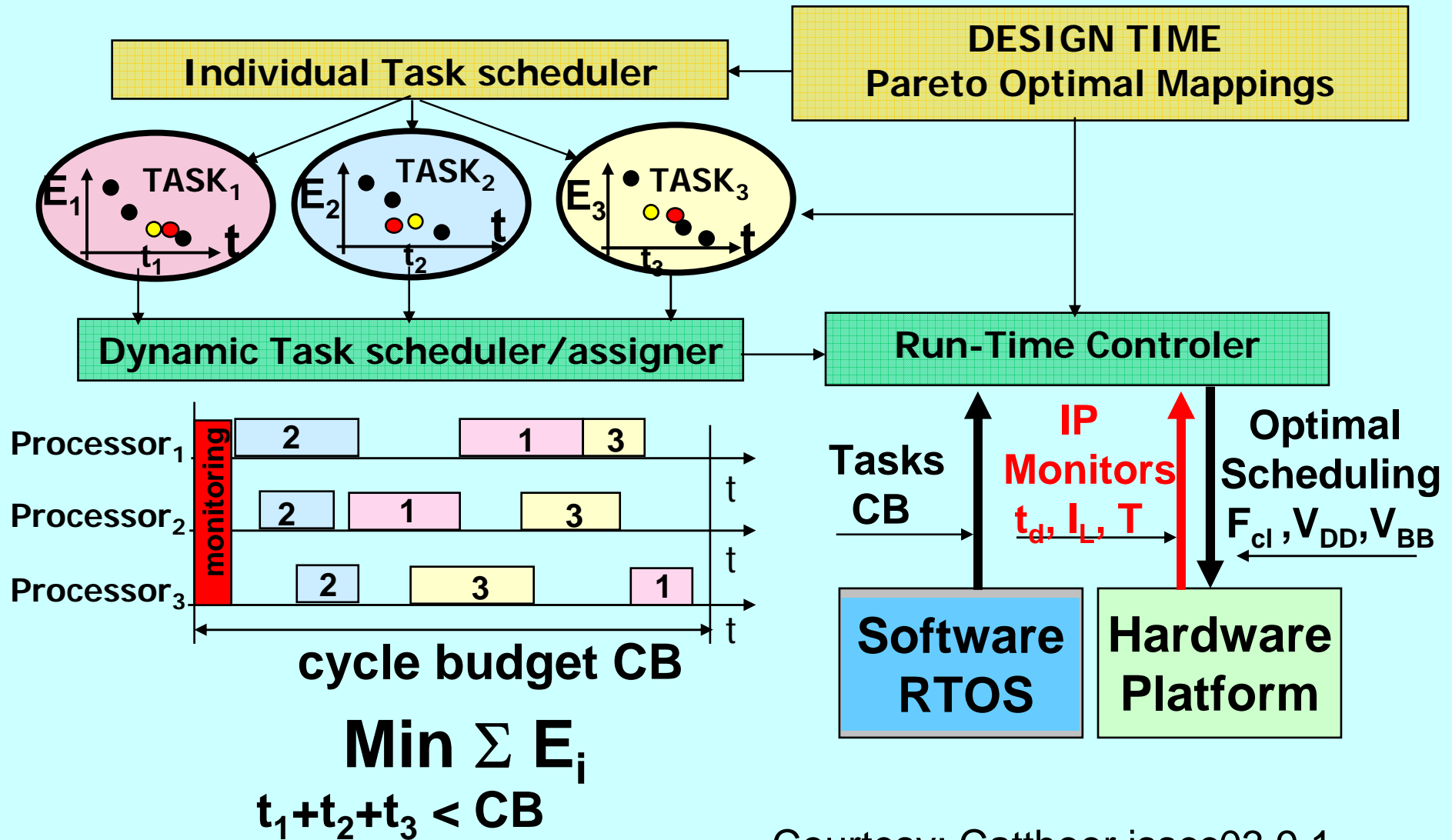
Ways to Cope...

1. Adapt V_{DD} , f_{cl} , I_L at run time to whatever mother nature provides and to dynamic nature of software tasks
2. Use tile based **GALS** architectures, **networks-on-chip**, and **error correcting** on-chip communication techniques
3. **Back to regularity** (Litho, context)



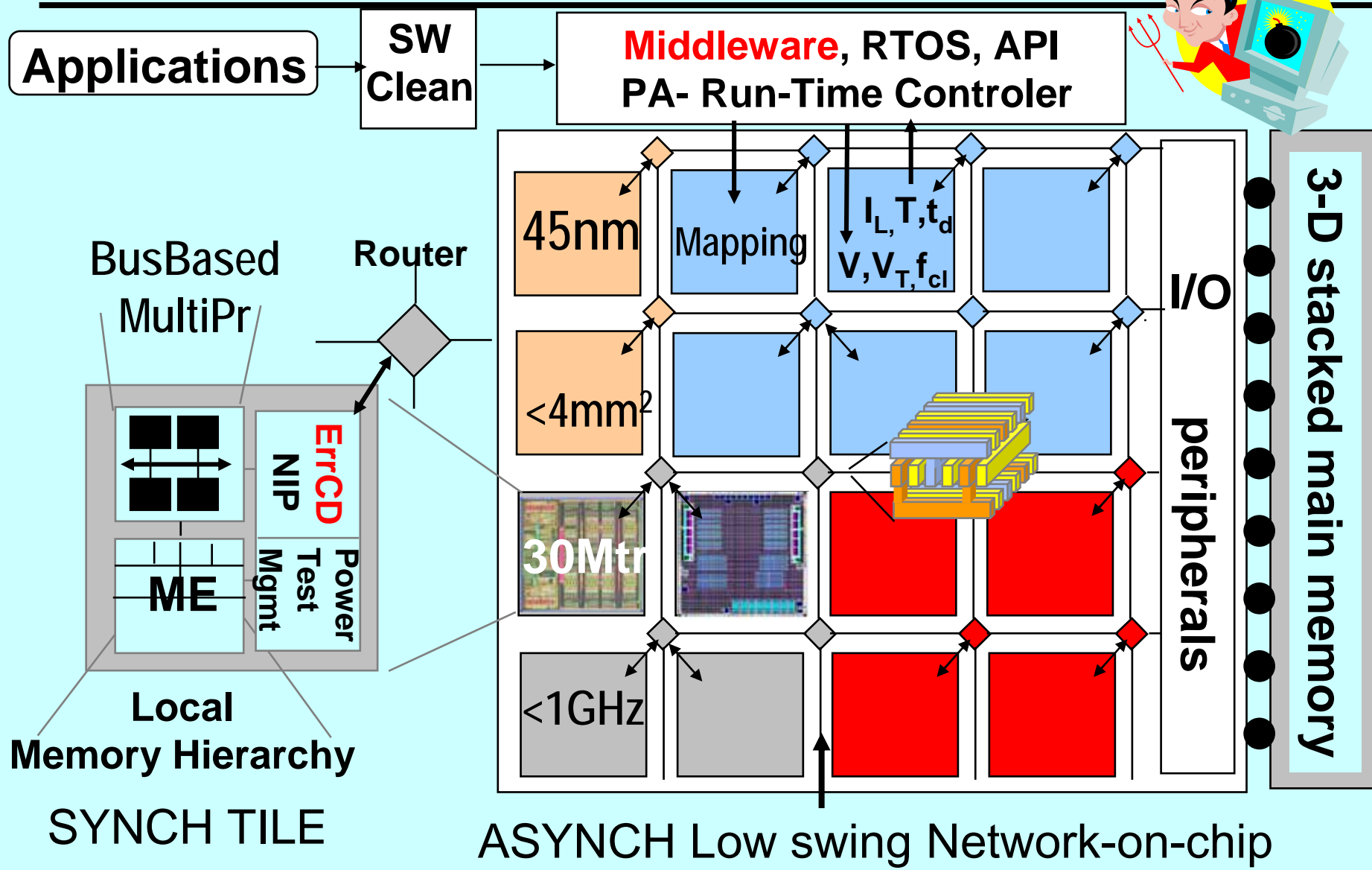
RapidChip

Run Time Adaptive Design

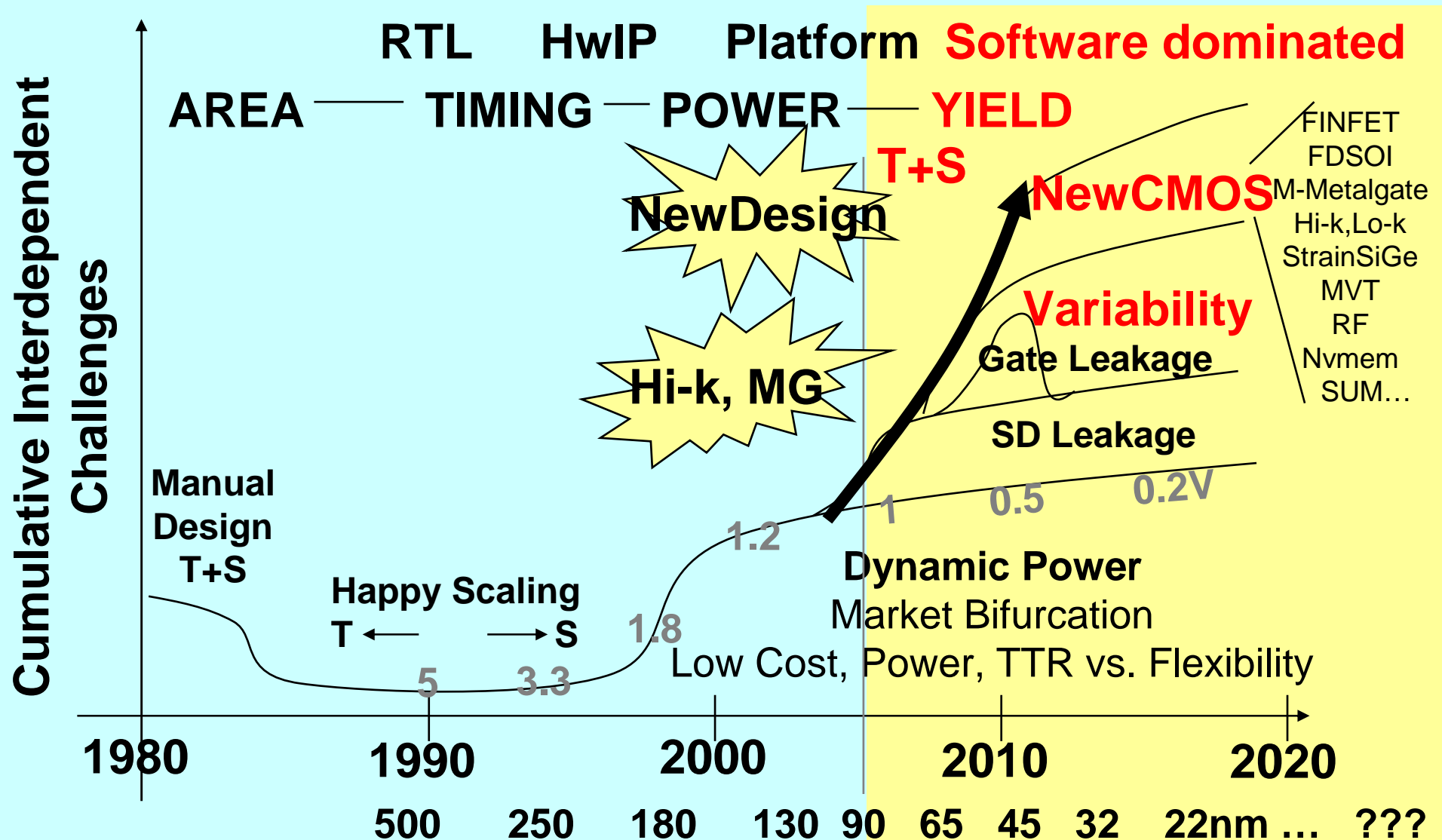


Courtesy: Cattoor isscc03 9.1

GALS Platform of Future + NOC



Interesting Times Ahead...



More Moore = more than more of the same...

- **Greater challenges than ever before:**
 - Managing giga-complexity faced with CMOS diversification
 - New design methods, tools and skills **urgently** needed
- **To reconcile the hell of physics with heaven of Aml dreams**
 - But easier to install equipment than to change people's mind
 - New processing nodes 2 years, **new design tools take 7 years,**
- **Need for sharing multi-disciplinary R&D cost**
 - From material science to computer science and societal engineering...Open innovation alliances from software to process development
- **Not for the faint of heart:**
 - few platforms by industry-R&D alliances and disciplined armies of engineers...

SEEDS FOR
TOMORROW'S
WORLD

THANK YOU!



expansion forces now!

