How EDA Could Save the World (of Computing)

Todd Austin

University of Michigan

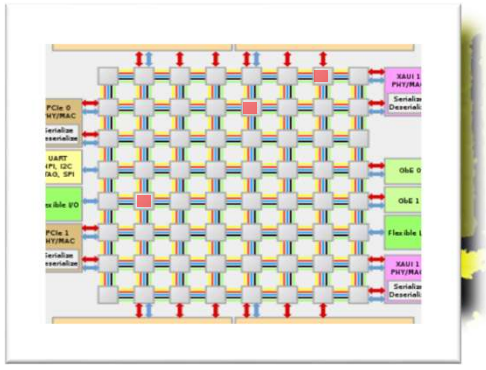
Perspectives on Scaling

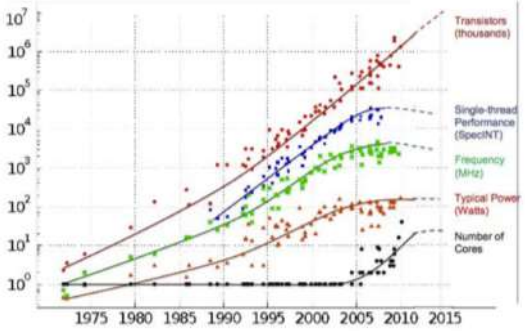
- **C-FAR**: Center for Future Architectures Research
  - Focused on scaling in 2020-2030 silicon
  - Performance, power and cost
  - 27 faculty at 14 universities, 92 students

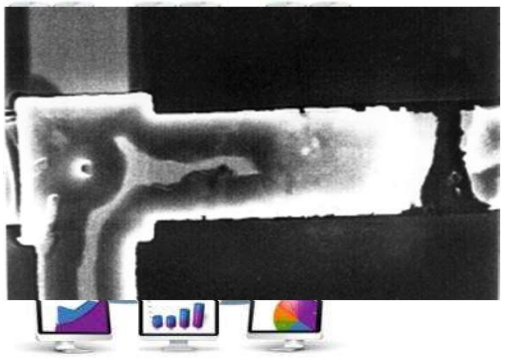
- **Why i**
  - The
- **Why i**
  - The threats… slowing innovation and degrading silicon

All of the work presented in this talk is that of ***C-FAR faculty***.

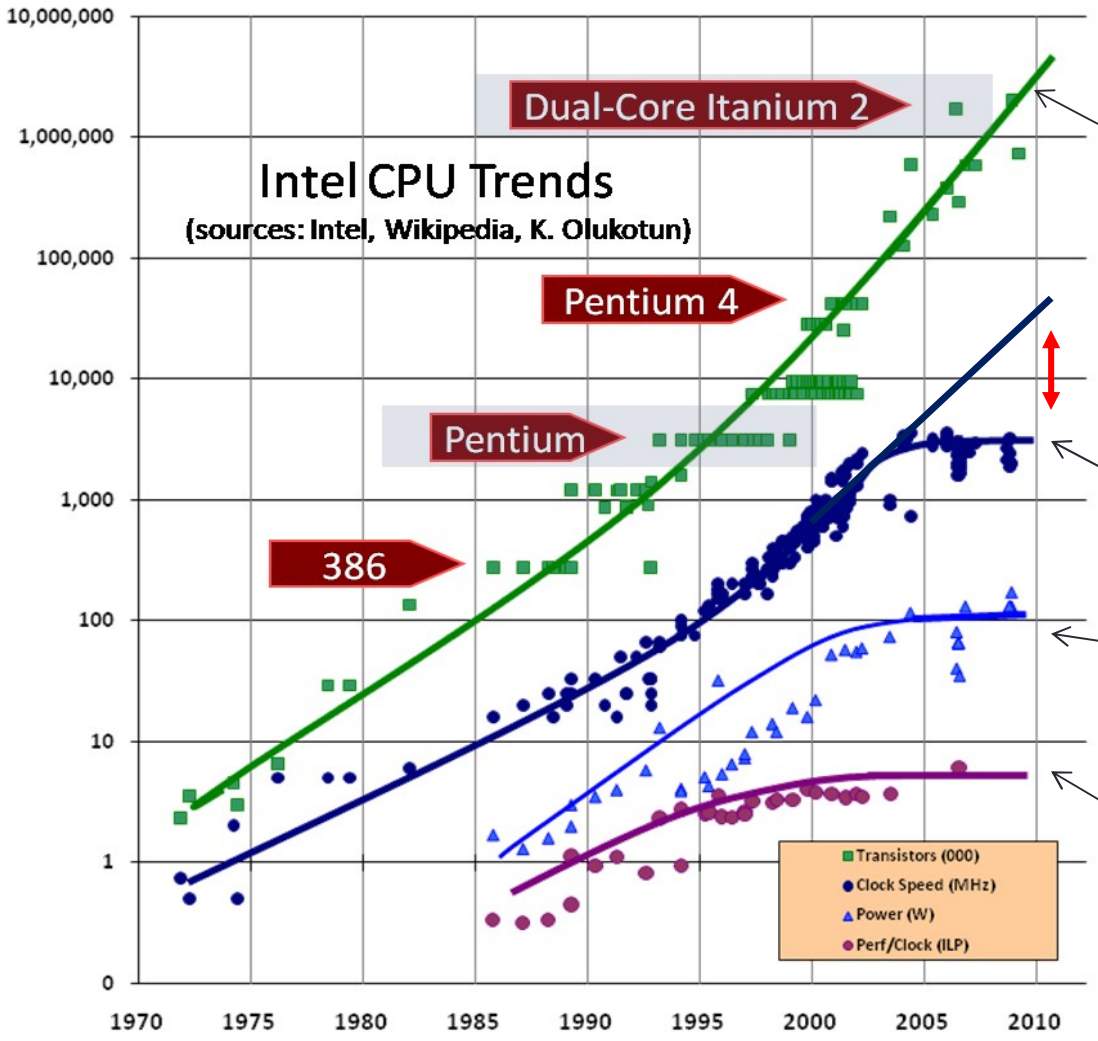
Many-Core Computing | Computer Vision | End of Traditional Scaling | Machine Learning | Big Data Analytics | Silicon Photonics







Moore's Law Performance Gap



When do we worry about this?

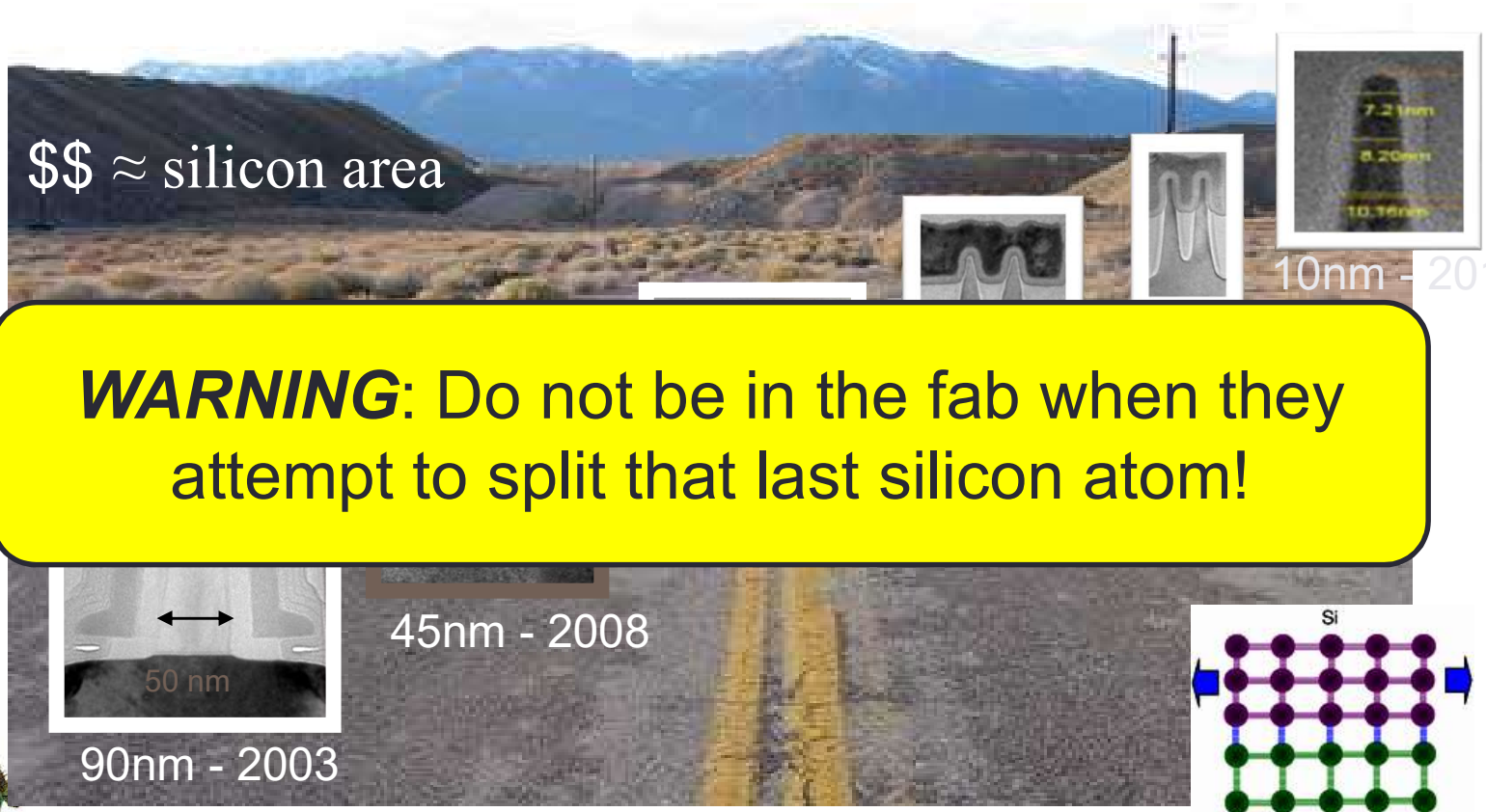
Today, gap is cresting 10x

Lack of perceived value

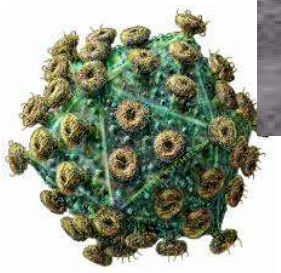
Dark silicon

Diminished ILP

Moore's Law vs. Physics



WARNING: Do not be in the fab when they attempt to split that last silicon atom!



100nm - virus

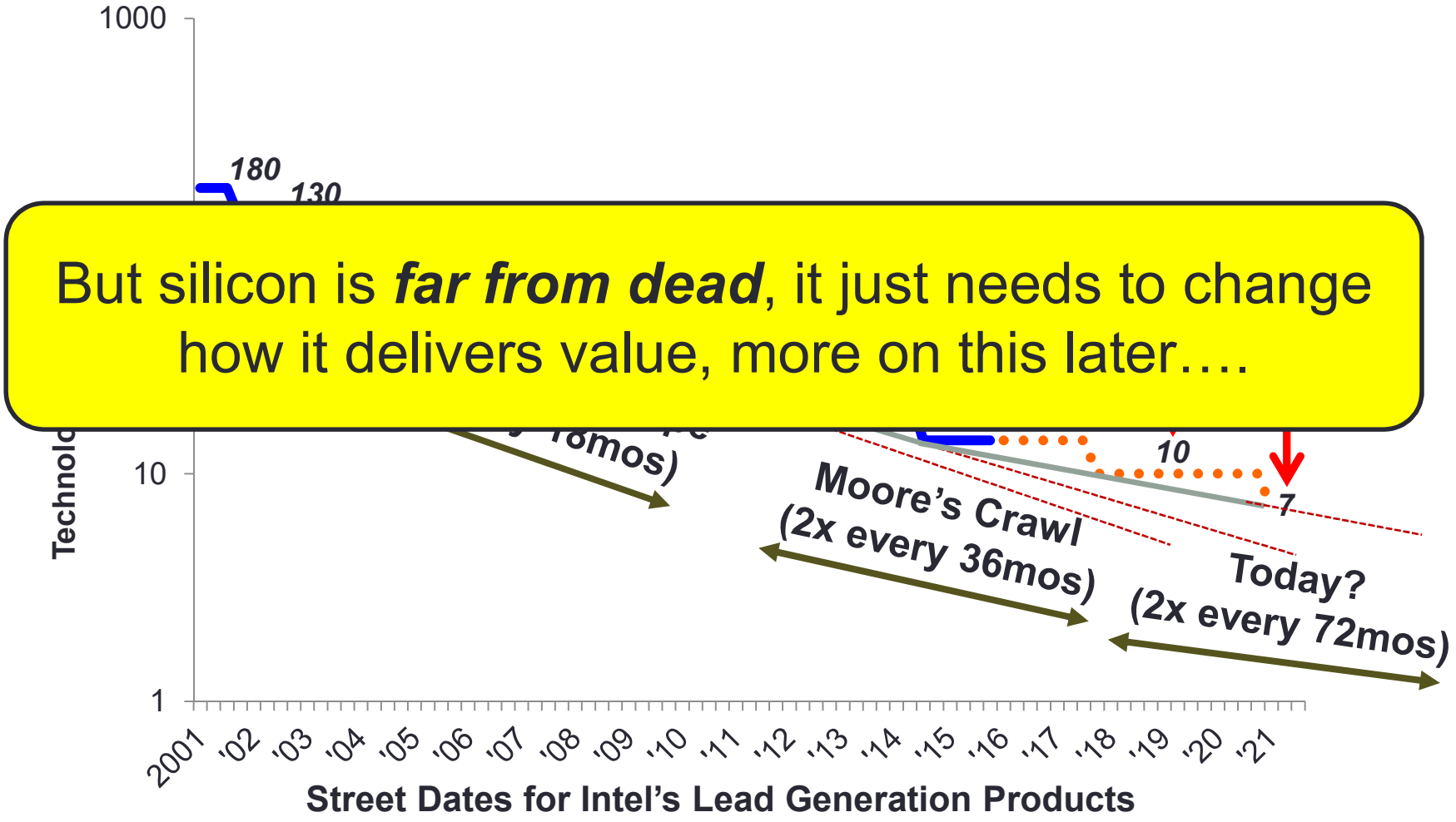
0.57 nm inter-atom spacing for strained silicon!

Moore's Law vs. Economics

Next-generation 10nm fabs will cost in excess of US$12B, few companies will accept this ***liability***!

Sources: Intel Investor Meeting 2014, GlobalFoundries, Intel, TSMC

The Inevitable Becomes Real



But silicon is ***far from dead***, it just needs to change how it delivers value, more on this later….

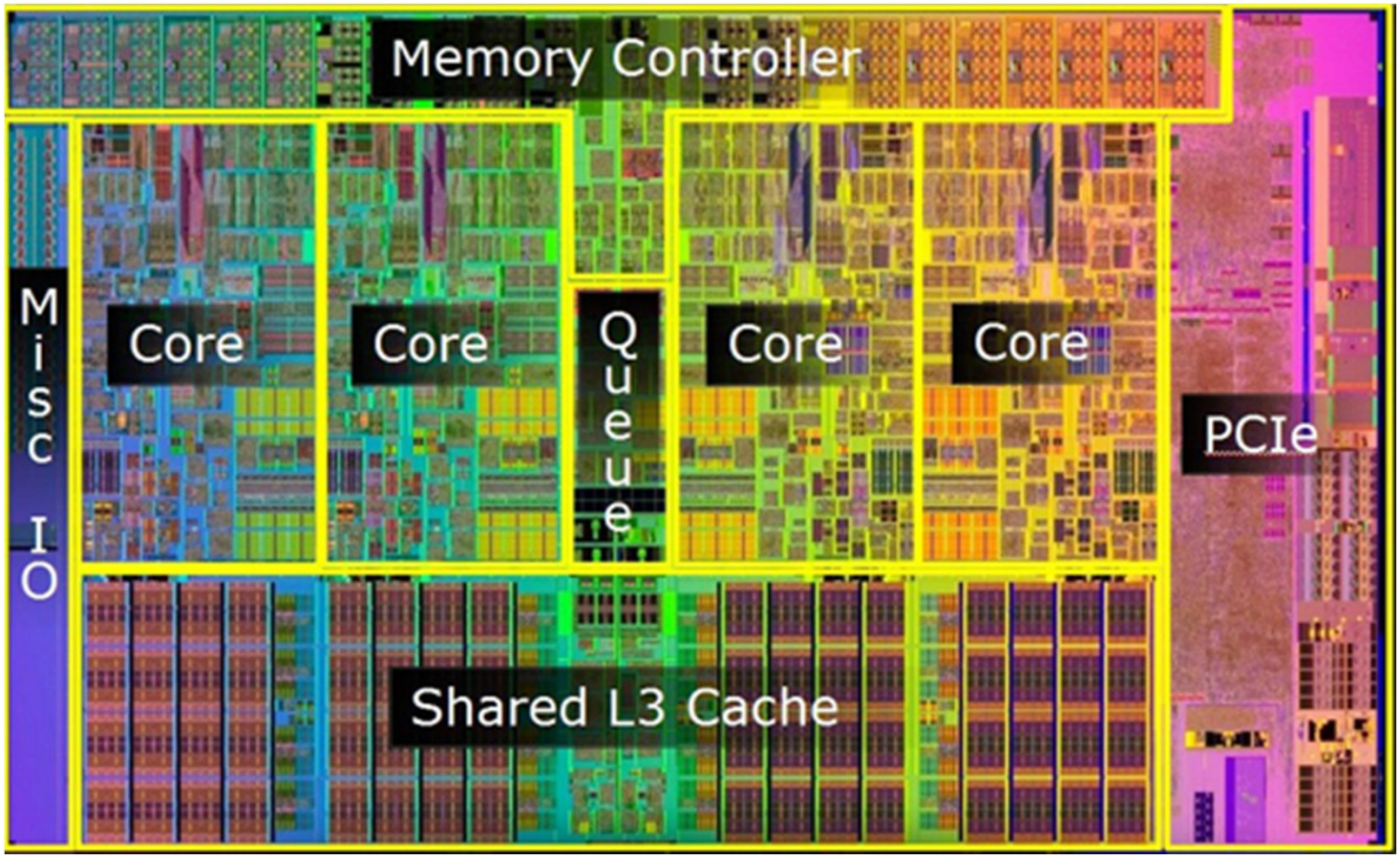
Compiled with David Brooks @ Harvard

What Does This Mean to System Designers?

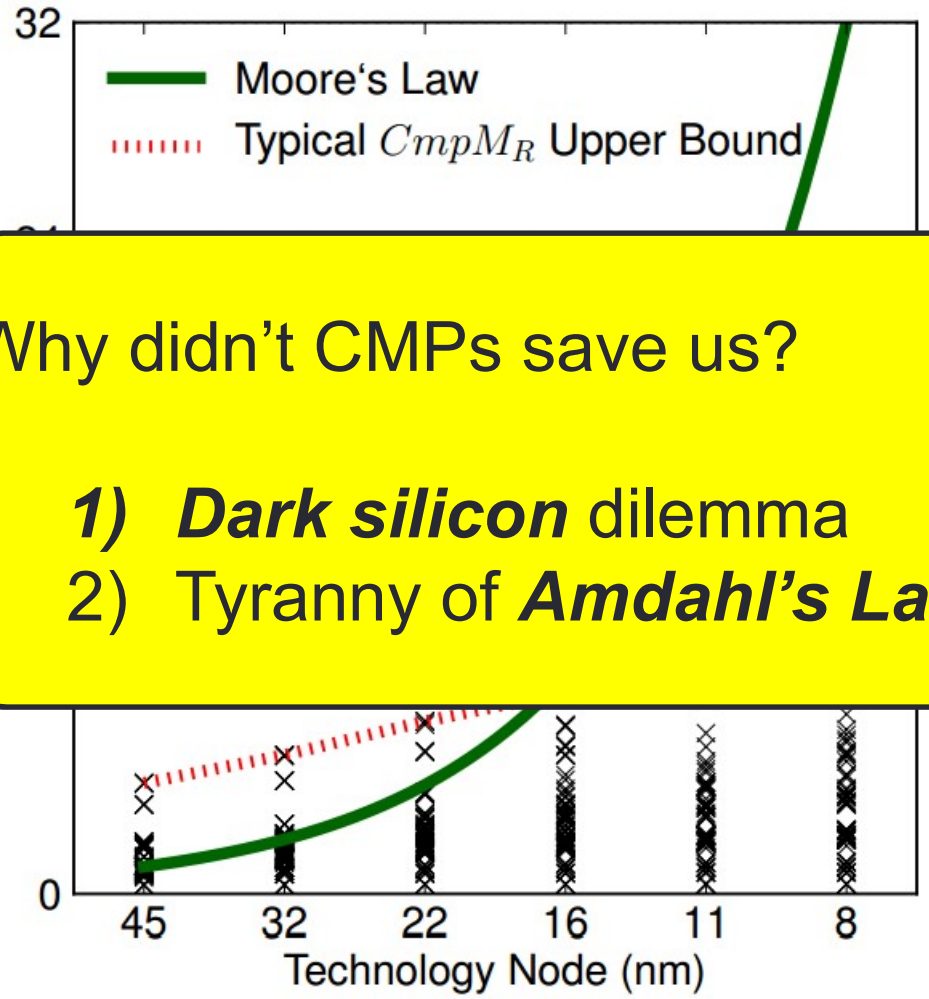
Today, value = scalability (performance, power, cost).

But, the ***Dennard scaling*** has left us, and ***silicon dimensional scaling*** scaling is slowing.

Remedy #1: Chip Multiprocessors



CMP Performance Scaling for the Highly Parallel PARSEC Benchmarks

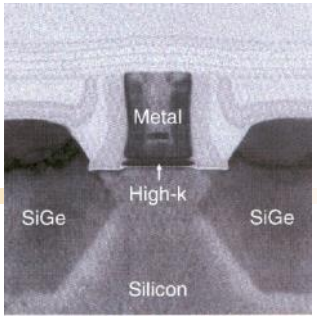


Why didn't CMPs save us?

1) ***Dark silicon*** dilemma
2) Tyranny of ***Amdahl's Law***

From "Dark Silicon and the End of Multicore Scaling," by Esmaeilzadeh *et al*.

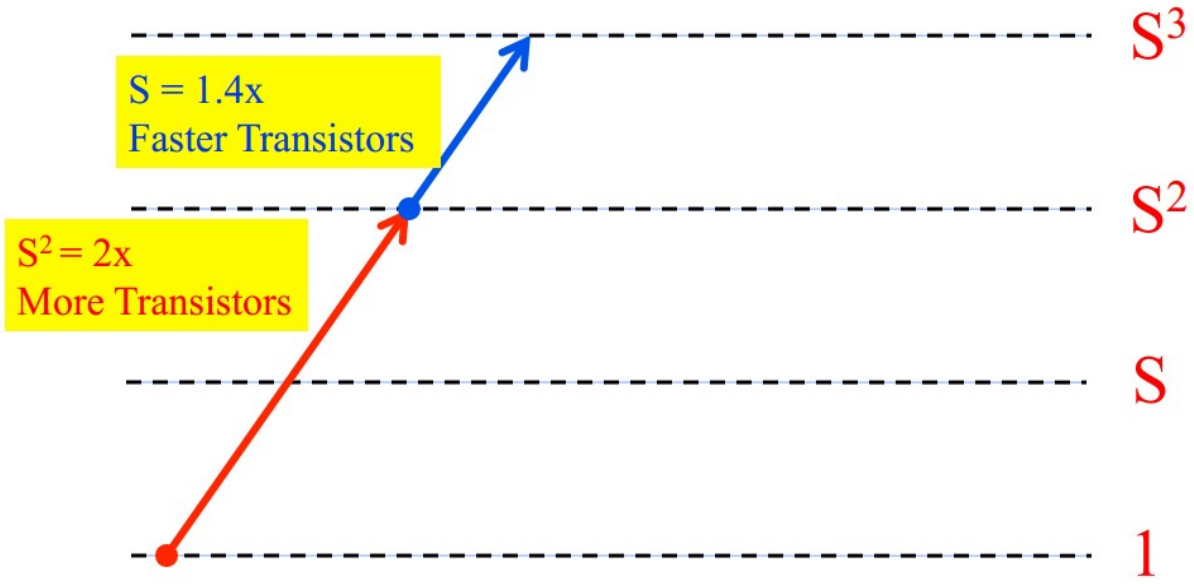
The Dark Silicon Dilemma



Advanced Scaling:

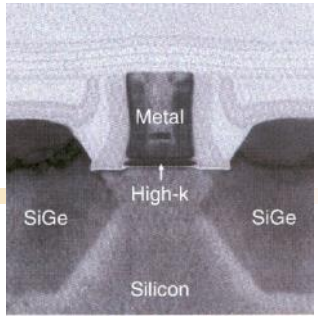
Dennard: "Computing Capabilities Scale by $S^3 = 2.8x$"

If S=1.4x …

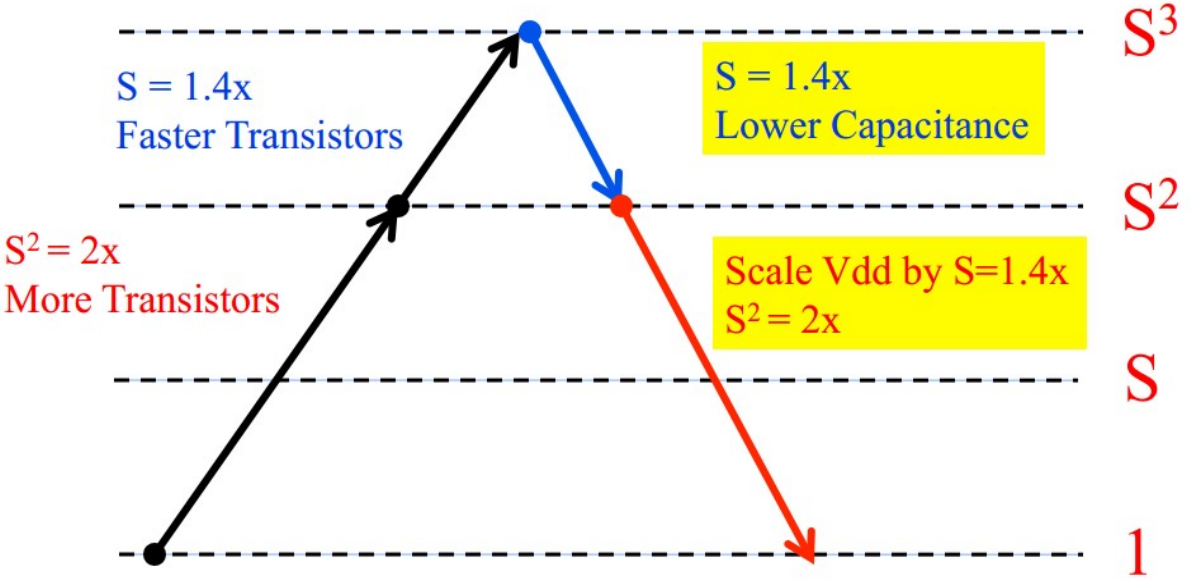


Courtesy Michael Taylor @ UWash

The Dark Silicon Dilemma

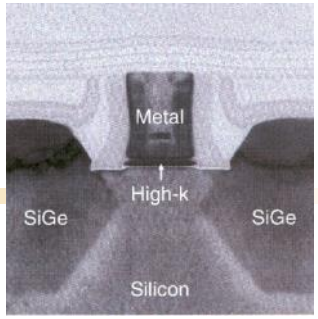


Dennard:
"We can keep power consumption constant"



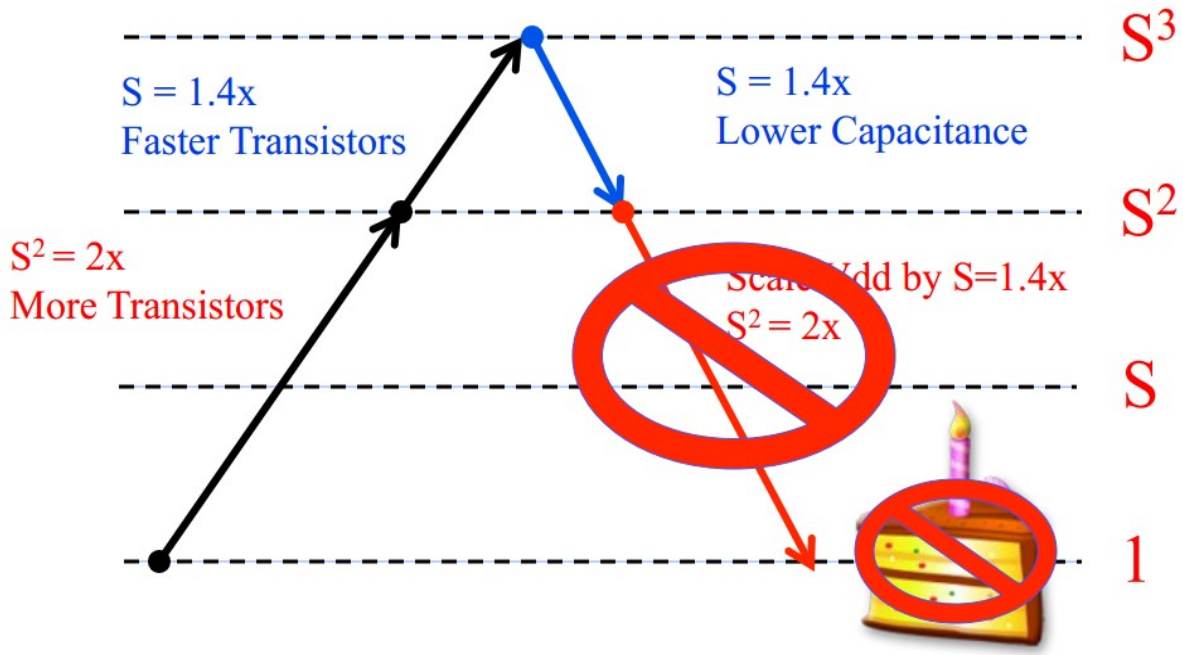
Courtesy Michael Taylor @ UWash

The Dark Silicon Dilemma



Fast forward to 2005:

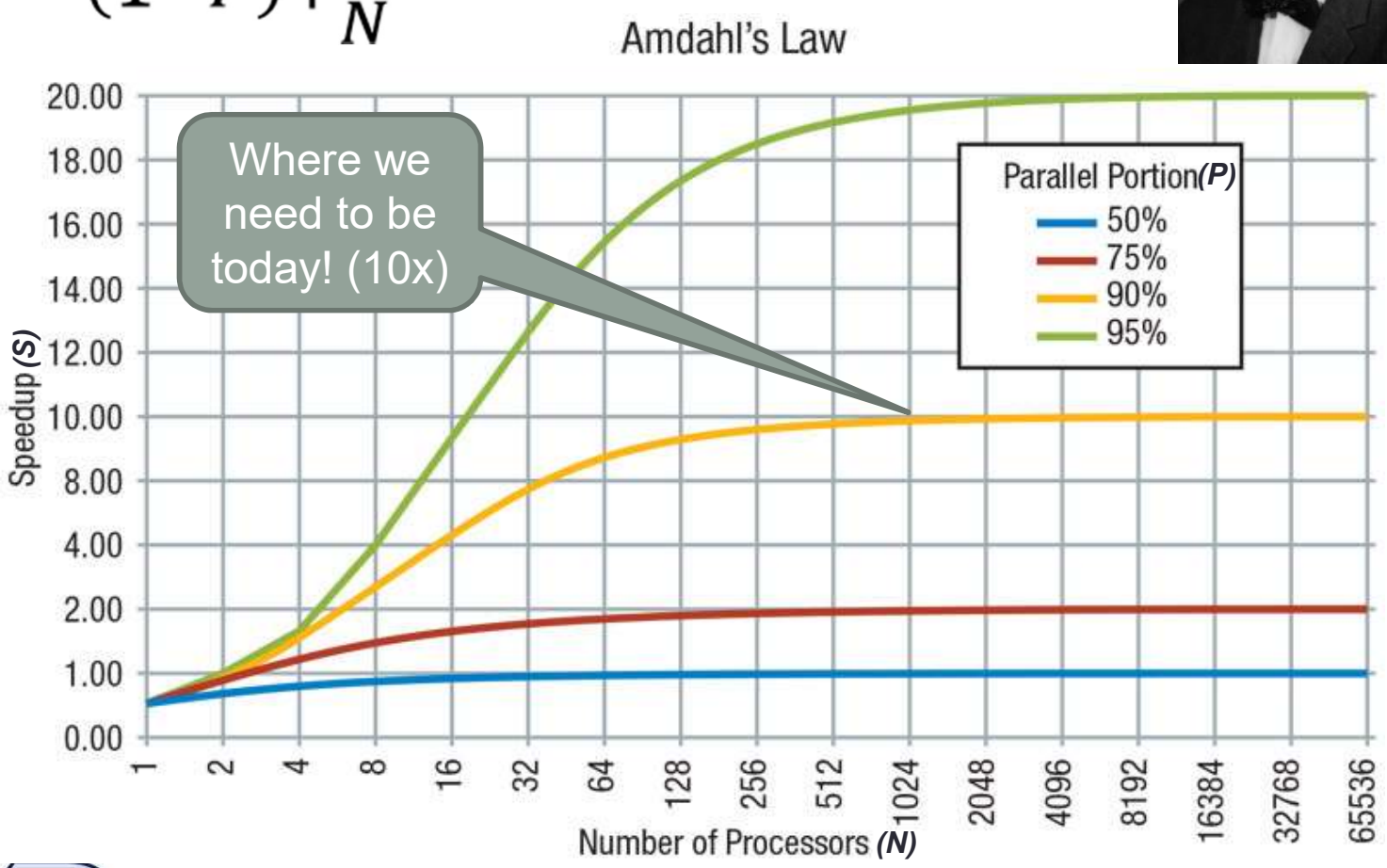
Threshold Scaling Problems due to Leakage Prevents Us From Scaling Voltage



Courtesy Michael Taylor @ UWash

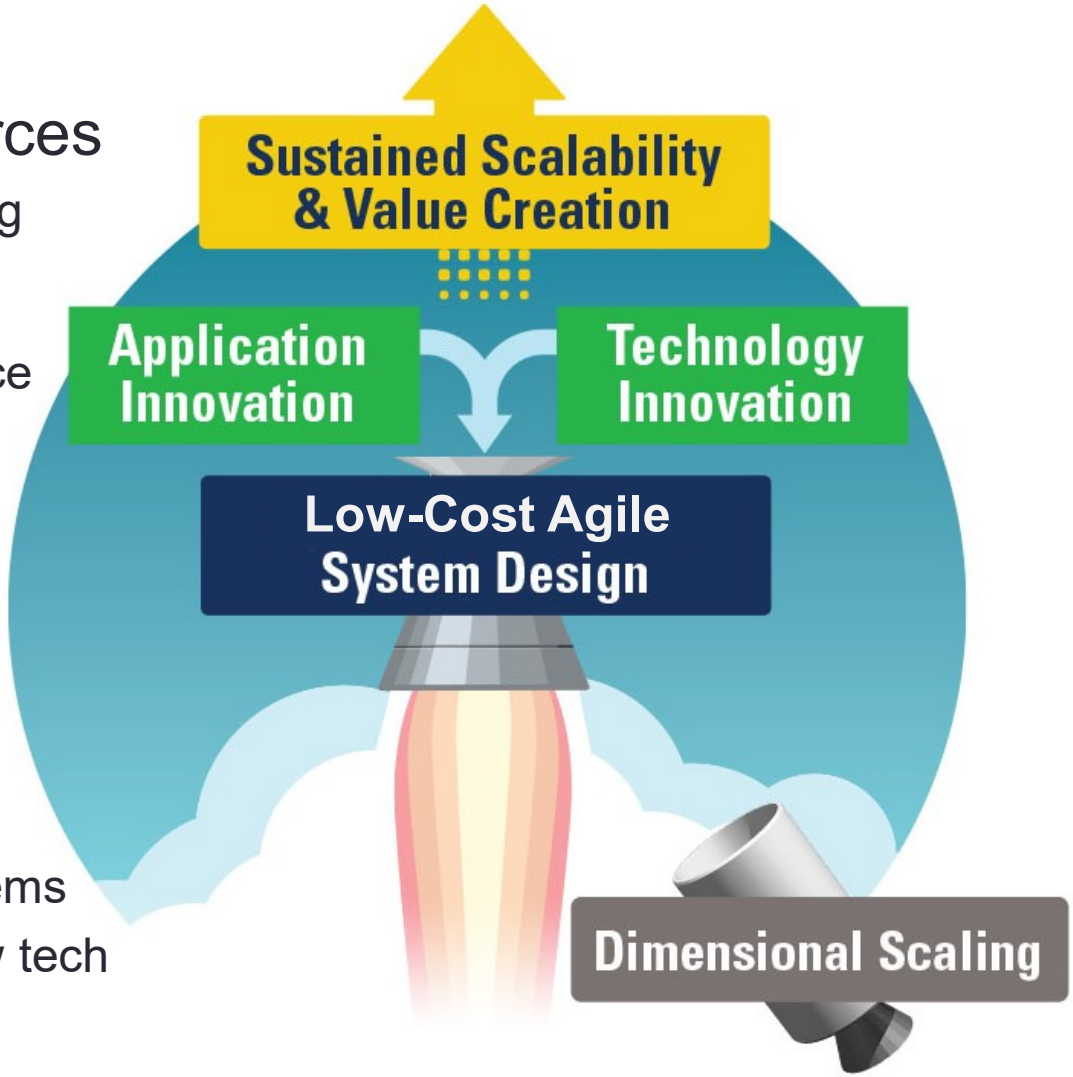
The Tyranny of Amdahl's Law

$$S(N) = \frac{1}{(1-P)+\frac{P}{N}}$$

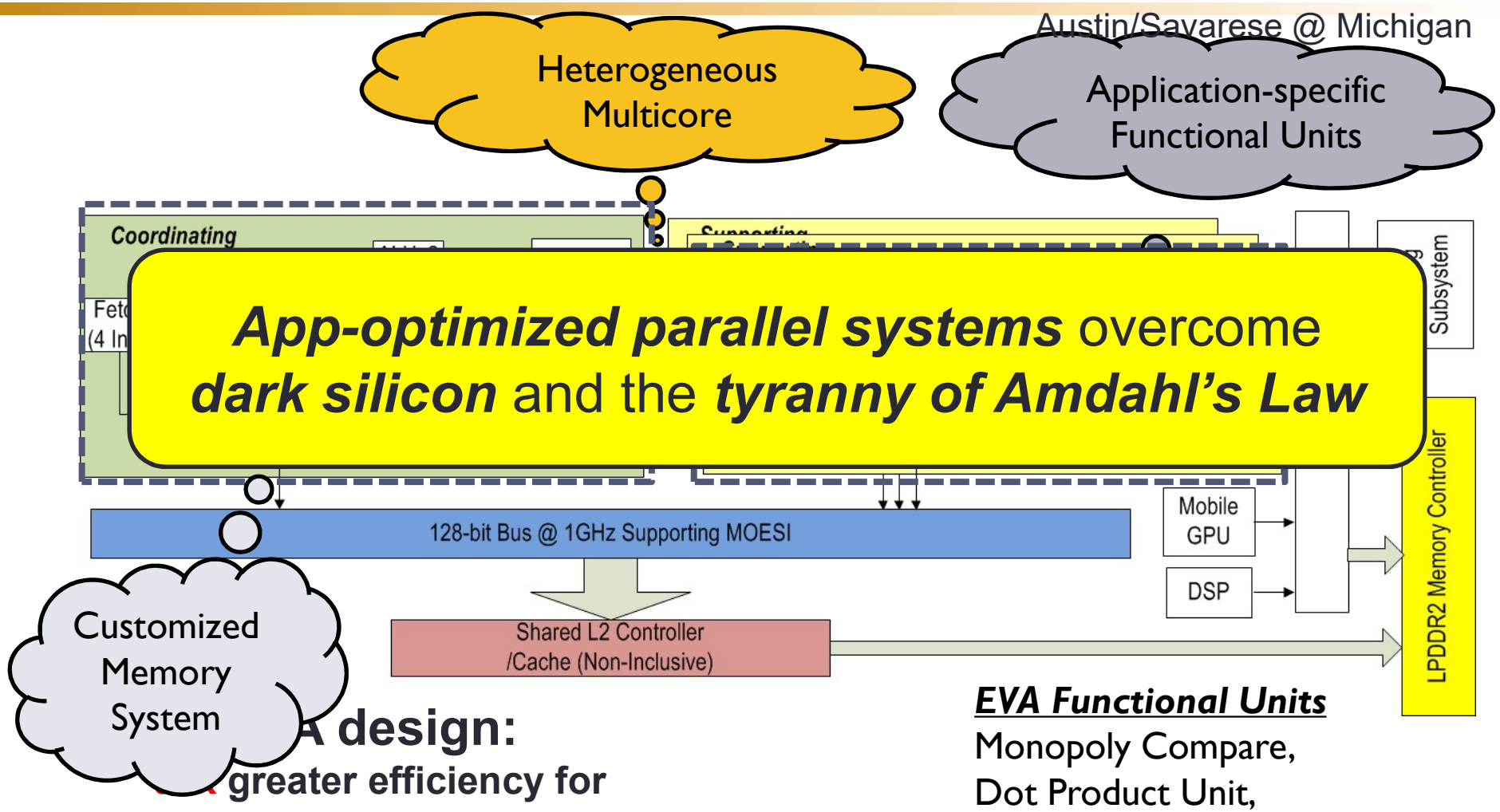


Where Should We Go Next?

- Value creation requires sustained innovation sources
  - Traditionally dimensional scaling
- New source: applications
  - Rely on fast-changing app space to deliver perf/power/cost value
- New source: silicon tech
  - Rely on structure, materials, and device advances in silicon
- Both sources necessitate low-cost agile design
  - Cheaply translate apps to systems
  - Effectively connect apps to new tech



Application Innovation: EVA Embedded Vision Architecture



App-optimized parallel systems overcome ***dark silicon*** and the ***tyranny of Amdahl's Law***

A design:

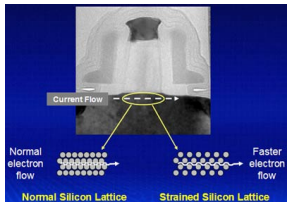
- greater efficiency for computer vision algorithms

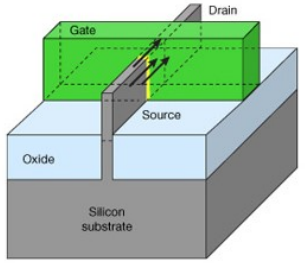
EVA Functional Units

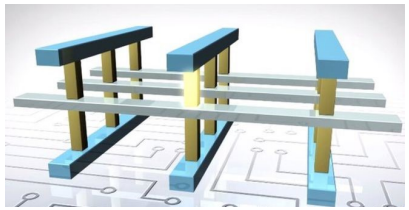
Monopoly Compare,
Dot Product Unit,
Vector Max,
Decision Tree Compare

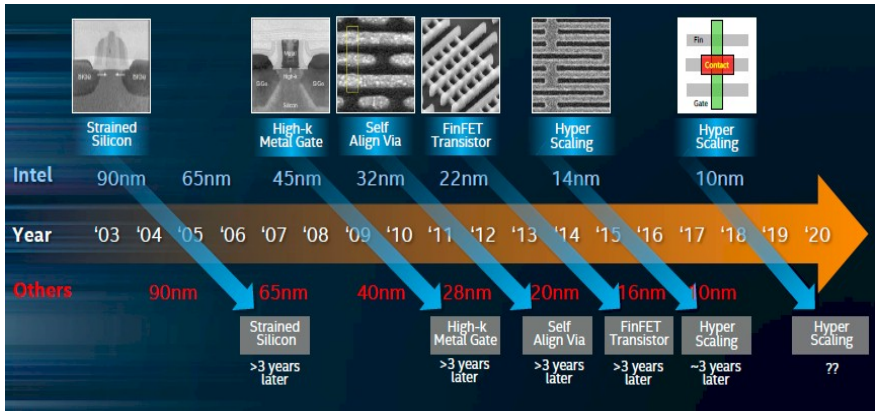
Technology Innovation: Beyond dimensional scaling

- For quite some time, silicon advances have been supported by technology advances
  - Strained silicon, High-k materials, FinFETs
- More coming, and the pace of this innovation will likely increase
  - Intel/Micron's XPoint memory technology
  - Intel's "hyperscaling" technologies
- Many additional opportunities possible
  - TFETs, CNTs, spintronics, novel materials, analog accelerators, etc…
- ***Key question***: will silicon scaling cost advantages persist?
- ***Key challenge***: driving tech advances for maximum system impact









Why These Ideas Will Likely Fail, Unless We Make a Change…

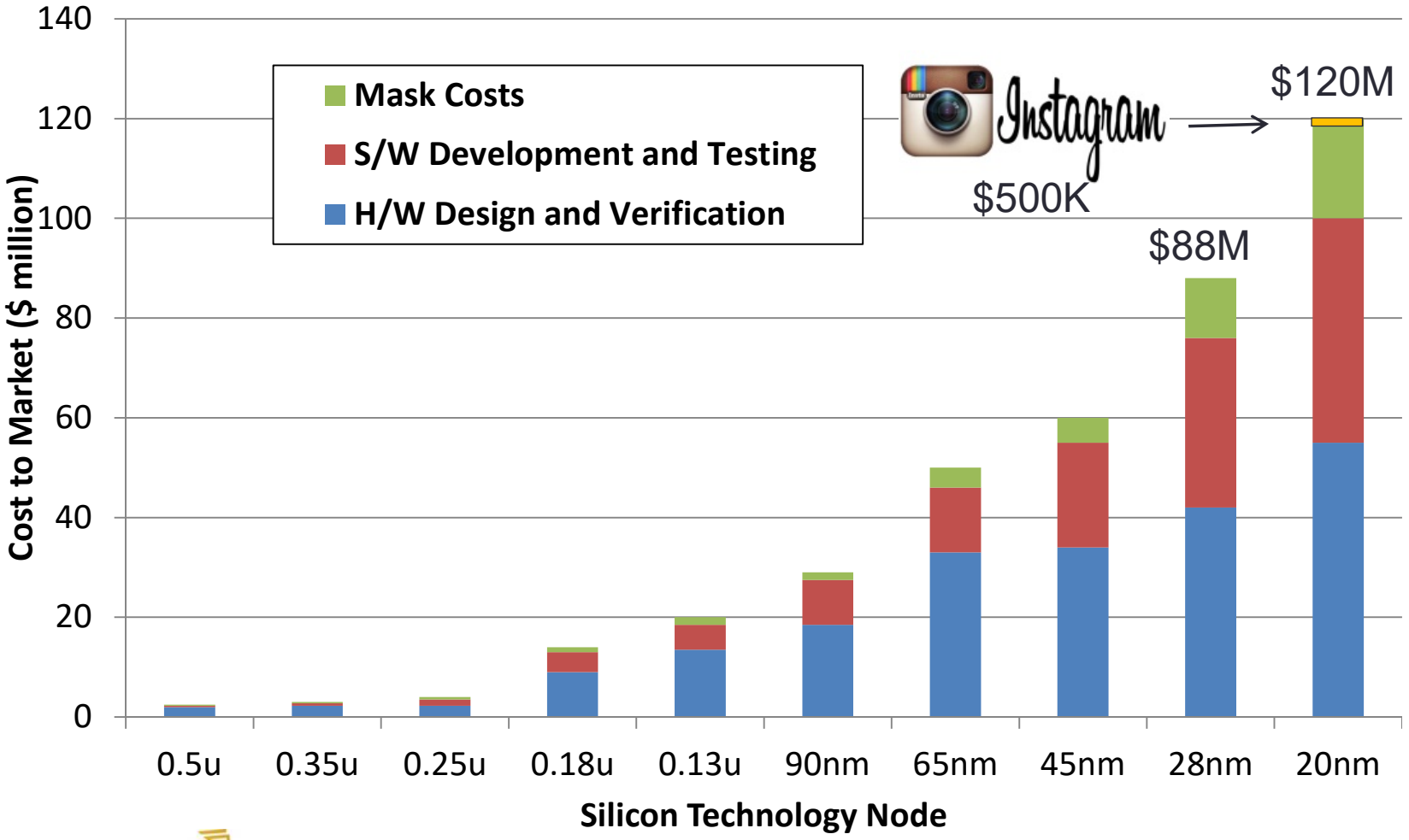
- ***The Good***: App/tech optimized systems can close the Moore's Law gap
- ***The Bad***: Dennard scaling has stopped, Moore's Law is slowing, leaving a growing scalability gap
- ***The Ugly***: The app/tech optimized designs needed to close the gap will be ***too expensive to afford***
  - We must work to make design much ***cheaper***!

What I Want You to Remember

- Successfully bridging the Moore's Law performance gap is less about "***How***" to do it and more about "***How Much***" does it cost!

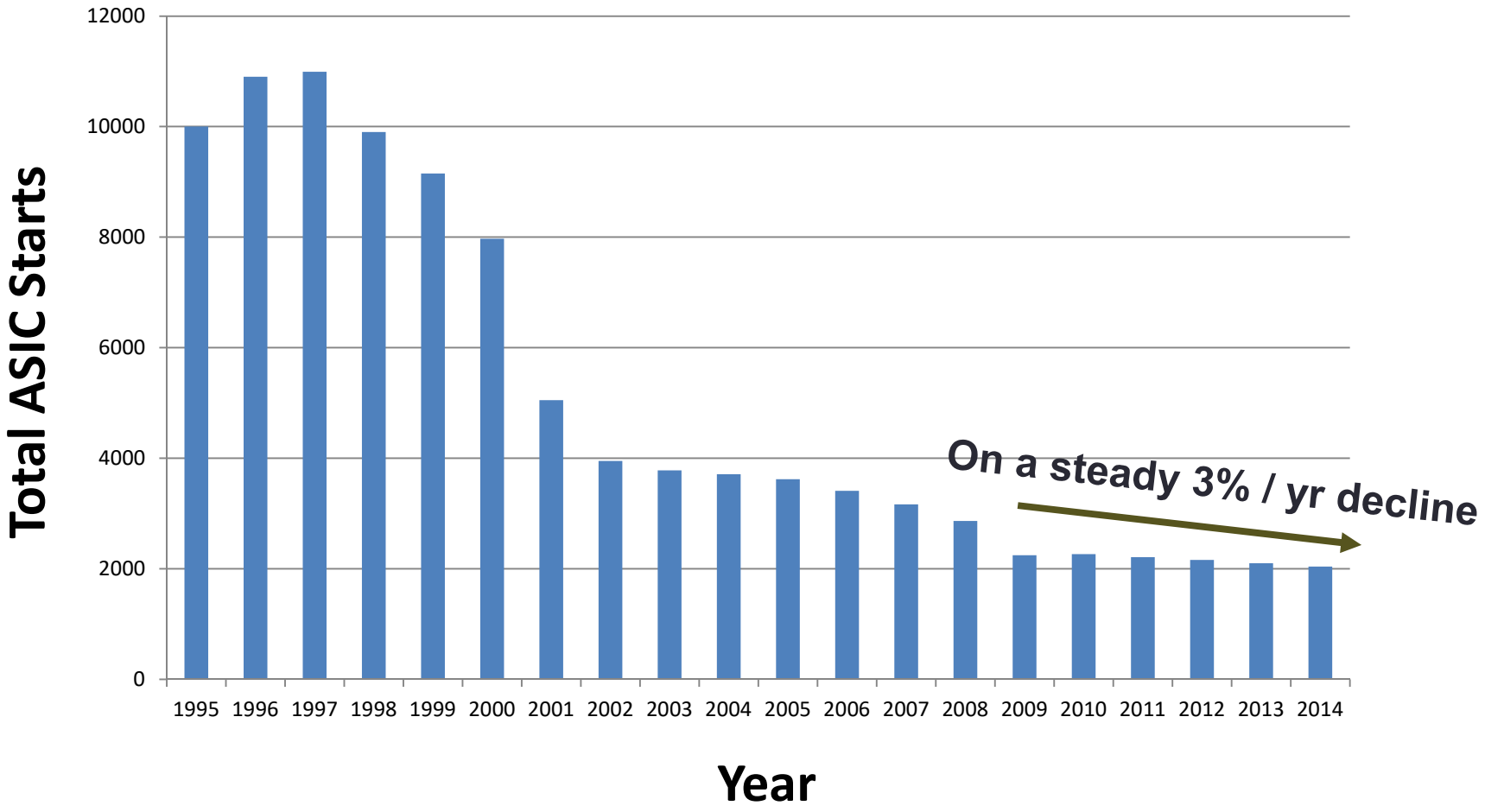
- **My claim:** if we can effect a ***100x reduction*** in the cost to bring a design to market, ***innovation will flourish*** and ***scaling challenges will be overcome***.

Design Costs Are Skyrocketing



Source: International Business Strategies

Outcome: "Nanodiversity" is Dwindling



Source: Gartner Group

The Remedy: Scale Innovation

- Ultimate goal: ***accelerate system innovation*** and make it sufficiently inexpensive that ***anyone can do it anywhere***
  - $1M for idea-to-market
  - Requires a 100x decrease in design costs

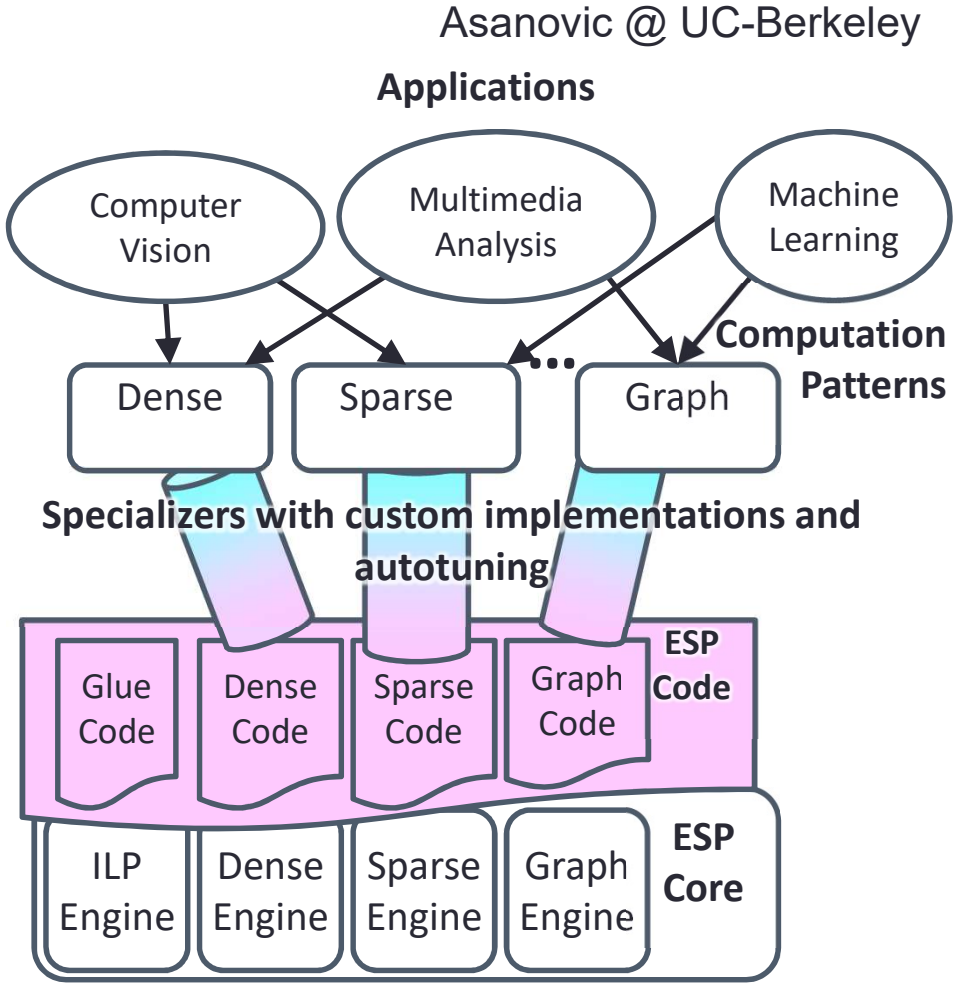
- Things you ***cannot*** do (and still meet the $1M design point):
  - Design new accelerators
  - Specify the design in low-level RTL
  - Write low-level software
  - Verify everything
  - Fabricate a chip (in full production)
  - License today's expensive 3rd part IP

The Remedy: Scale Innovation

- Ultimate goal: ***accelerate system innovation*** and make it sufficiently inexpensive that ***anyone can do it anywhere***
- Remedy #1: Embrace system-level design with reusable algorithm-inspired components

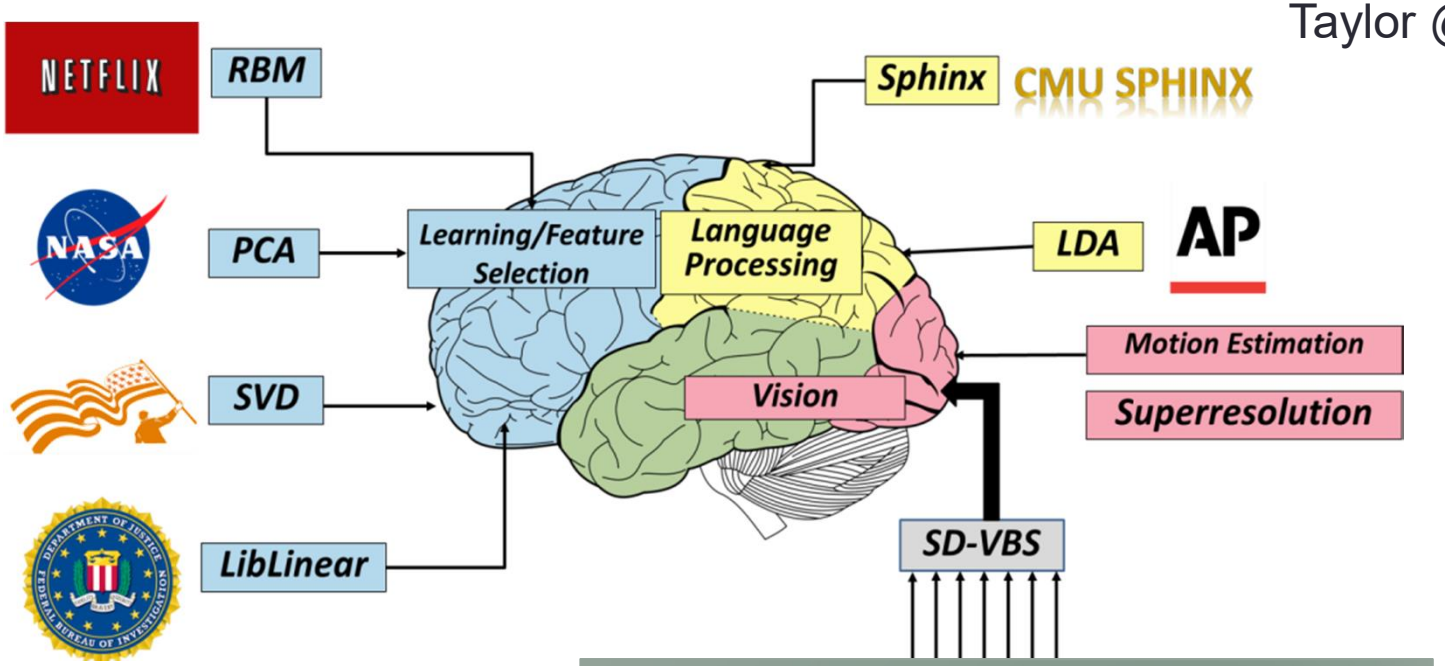
1) Embrace system-level design with reusable algorithm-inspired components

- System-level design delivers higher efficiency and productivity
- Example: Ensembles of Specialized Processors (ESP)
- Reusability achieved by targeting algorithms rather than apps
  - All apps built on a few dozen algs
  - Dense linear algebra vs. GPU
- Grand challenges remain: ***what are the components*** and ***how are they connected***?



Application-Driven Design Requires Expertise with Emerging Applications

Taylor @ UWash



CortexSuite

Image Segmentation | Disparity Map | Robot Localization

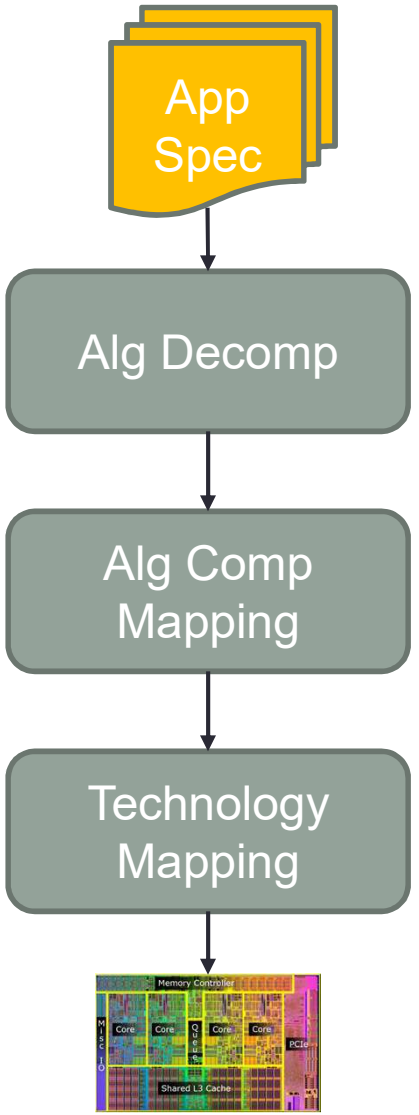
Support Vector Machines | Texture Synthesis | Feature Tracking | Image Stitch | SIFT

The Remedy: Scale Innovation

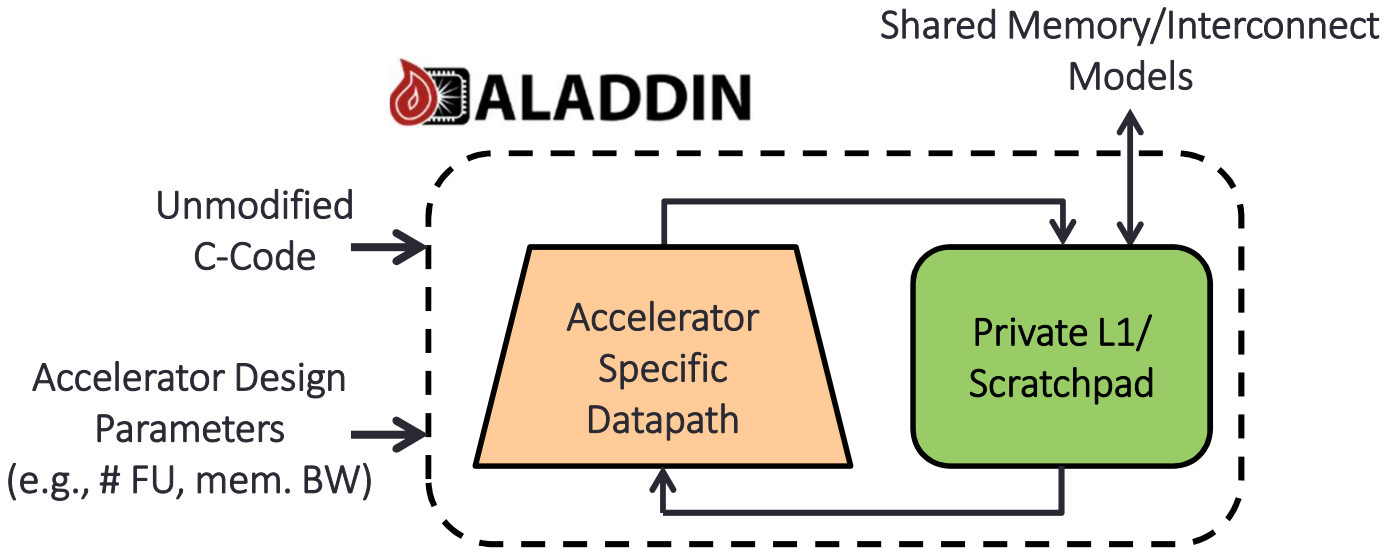
- Ultimate goal: ***accelerate system innovation*** and make it sufficiently inexpensive that ***anyone can do it anywhere***
- Remedy #1: Embrace system-level design with reusable algorithm-inspired components
- Remedy #2: Evolve EDA into language-to-package correct-by-construction design flows

2) Evolve EDA into language-to-package correct-by-construction design flows

- Today, EDA focuses on RTL-to-technology mapping
- Tomorrow, EDA must evolve to app-to-technology mapping
  - DSL or language level app spec, to…
  - Decomposed app, matching reusable algorithmic components, to…
  - Programmed and balanced algorithmic components with appropriate interfaces, to…
  - Technology-mapped application
- Would greatly boost productivity
- Tool verification would ensure much of the design is correct-by-construction



Aladdin Pre-RTL Analysis Tool



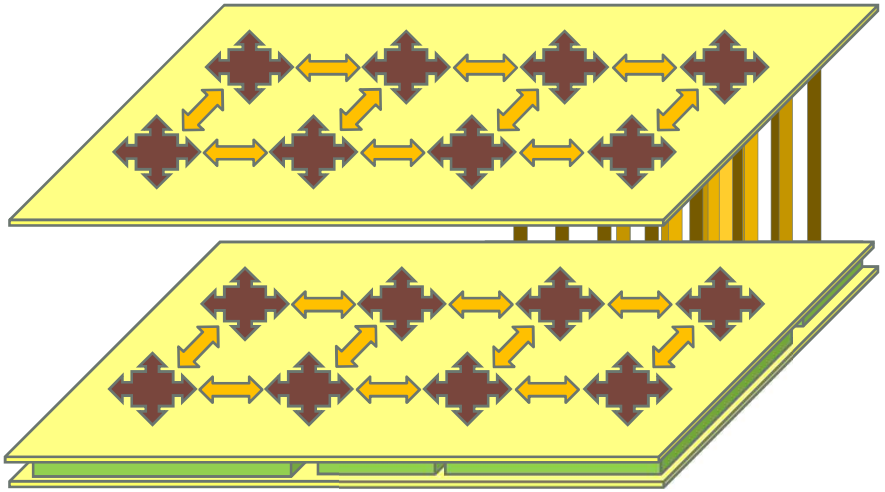
- Aladdin high-level synthesis tool
  - Analyzes ***language-level specification*** of accelerator under designer constraints
  - Utilizes ***design space exploration*** to efficiently identify performance, power, and area optimized designs
  - Considers both ***computational and memory components***

The Remedy: Scale Innovation

- Ultimate goal: ***accelerate system innovation*** and make it sufficiently inexpensive that ***anyone can do it anywhere***
- Remedy #1: Embrace system-level design with reusable algorithm-inspired components
- Remedy #2: Evolve EDA into language-to-package correct-by-construction design flows
- Remedy #3: Mature system-in-package technologies to enable package-time custom ICs with low NRE costs

3) Mature system-in-package technologies to enable package-time custom ICs with low NRE costs

- ***System-in-package*** technologies enable ***assembly-time customization***, *i.e.*, MCMs + 3D + reconfigurable interconnect



Brick-and-mortar silicon design flow:
1) Assemble brick layer
2) Connect with mortar layer
3) Package assembly
4) Deploy software

- Customized design made from physical alg components
- Component NRE costs amortized across all designs
- Key hurdle: ***placement and connection tolerant integration***

The Remedy: Scale Innovation

- Ultimate goal: ***accelerate system innovation*** and make it sufficiently inexpensive that ***anyone can do it anywhere***
- Remedy #1: Embrace system-level design with reusable algorithm-inspired components
- Remedy #2: Evolve EDA into language-to-package correct-by-construction design flows
- Remedy #3: Mature system-in-package technologies to enable package-time custom ICs with low NRE costs
- Remedy #4: Let system demands drive silicon technology advances

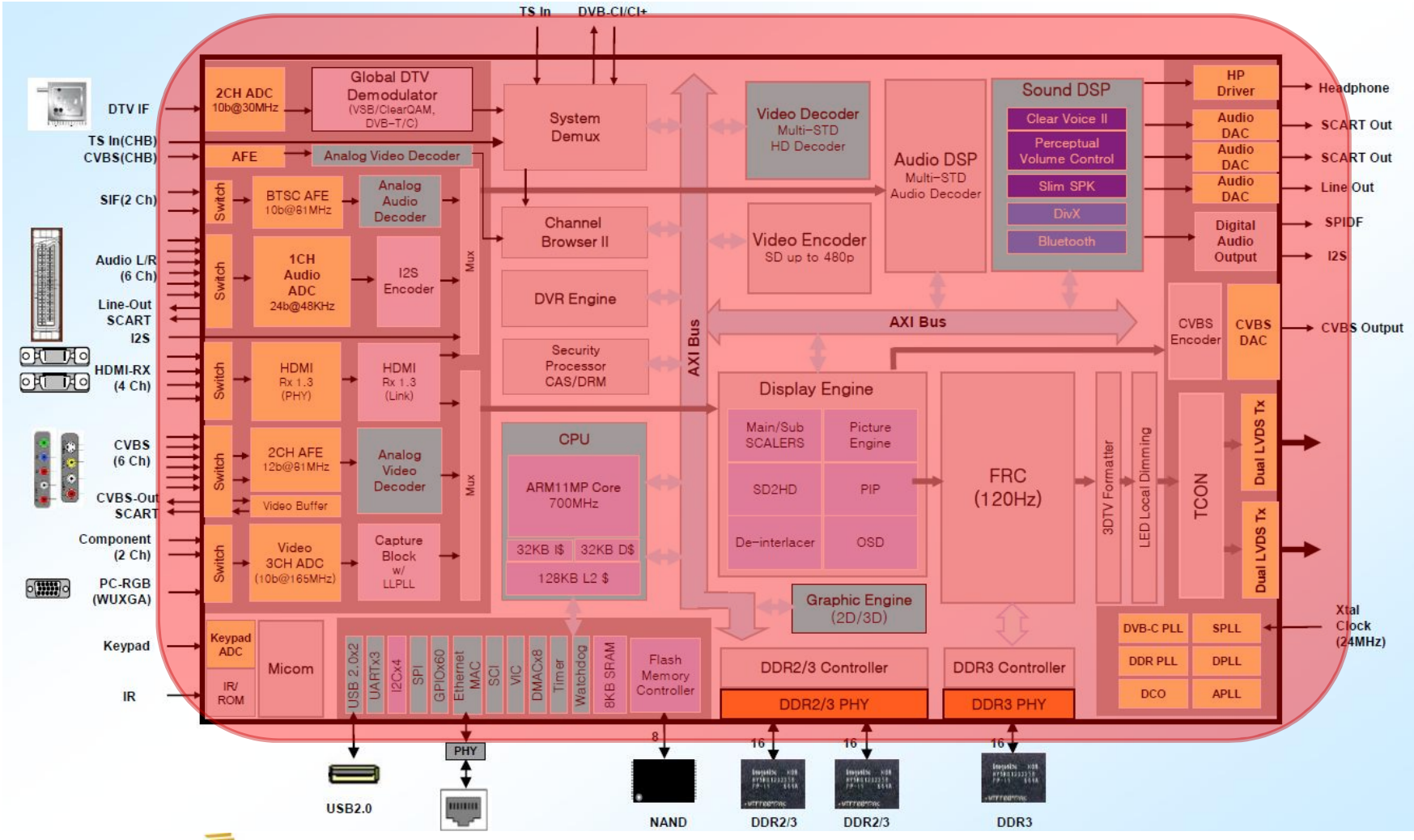
4) Let system demands drive silicon technology advances

- Reduced leakage for computation/memory
- More efficient communication w/o more latency
- More efficient computation that is dense and cheap
- Could silicon advances have ***bigger impacts***?
- Controllable and recognizable analog functions
  - Allows computation to be replaced with fast and efficient analog compute
- Emerging tech that delivers value at low fault rates
  - Many low-cost fault tolerance techs exist today, limit faults to < 0.1%
- Fast, efficient, non-destructive writes for NVM
  - Gives simpler, denser, more efficient mem, enables ultra-low power
- Computation w/o power/electrical signature
  - Systems are riddled with side channels, needed to establish H/W trust

The Remedy: Scale Innovation

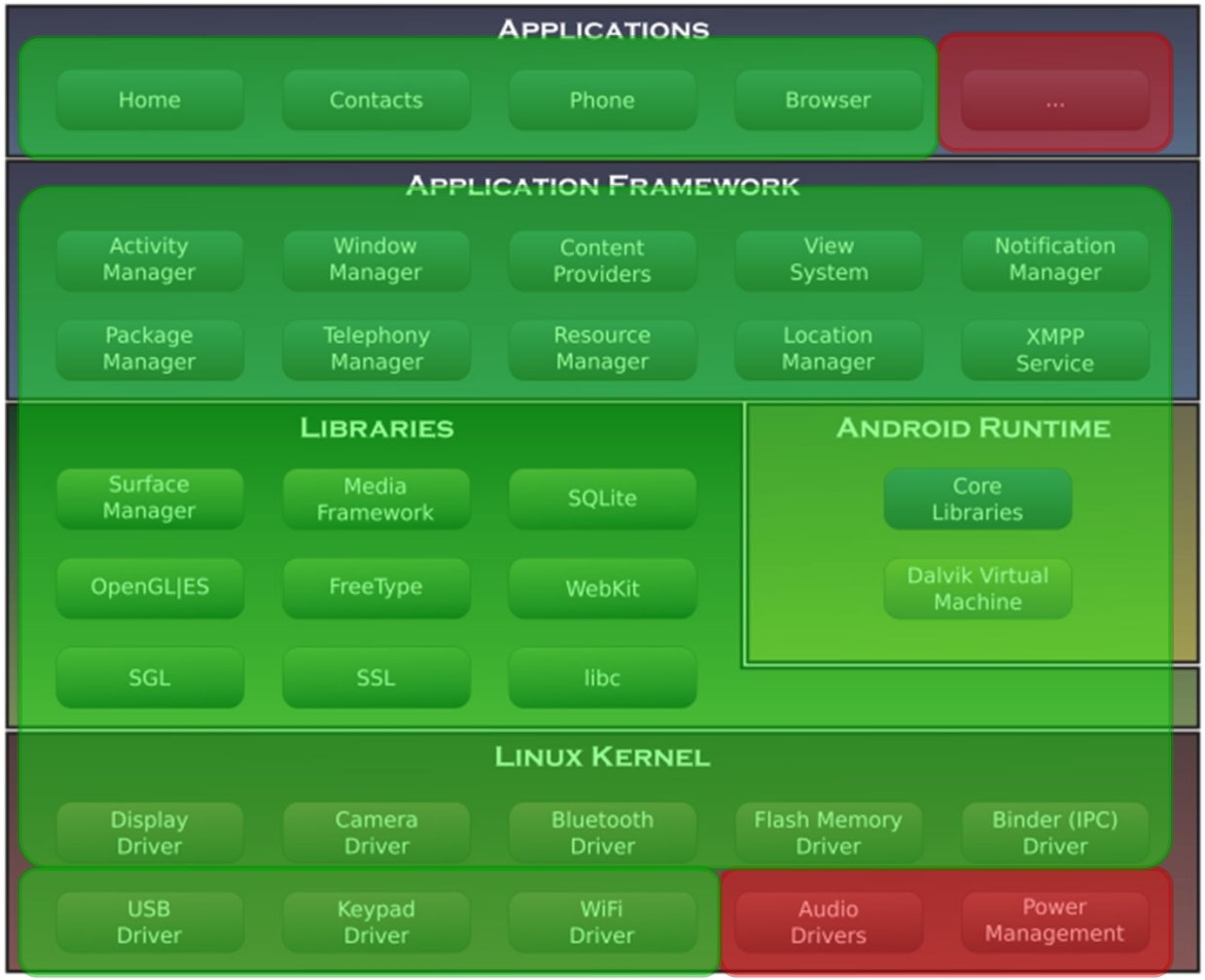
- Ultimate goal: ***accelerate system innovation*** and make it sufficiently inexpensive that ***anyone can do it anywhere***
- Remedy #1: Embrace system-level design with reusable algorithm-inspired components
- Remedy #2: Evolve EDA into language-to-package correct-by-construction design flows
- Remedy #3: Mature system-in-package technologies to enable package-time custom ICs with low NRE costs
- Remedy #4: Let system demands drive silicon technology advances
- Remedy #5: Embrace open-source H/W to focus industry investment

5) Embrace open-source H/W to focus industry investment



Red = non-free IP, **Green** = free IP

5) Embrace open-source H/W to focus industry investment



Red = non-free IP, **Green** = free IP

Does Open-Source H/W Mean an End to Profits?

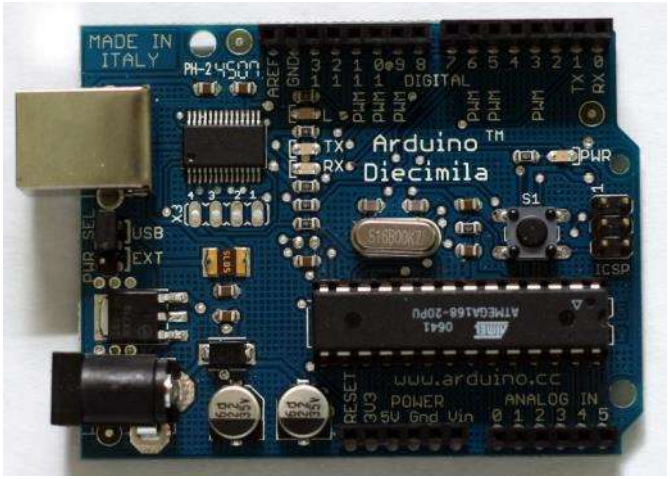
- No, not in the way I am suggesting we utilize open source…
- We need to decide as a community what IP is no longer worth investing in its *closed-source development*

Open source helps to ensure that every $ invested creates ***customer-perceived value***.

Instead, ***collaborate to maintain open-source IP*** and ***focus investment on developing new closed-source IP***

- ***Over time the availability of open-source IP grows***, and the cost of bringing new IP to the market is further mitigated
  - Consider: How did Instagram come to market for $500k?

Good News: Open-Source H/W is Growing



RISC·V

OpenPiton

CFAR

Conclusions

- App and technology optimized design could continue Moore's Law scaling benefits
  - But, it requires a diverse hardware ecosystem with affordable customization

- Effective and affordable customization won't happen without our help
  1. Embrace system-level innovation w/ reusable parts
  2. Evolve EDA into language-to-package flows
  3. Mature system-in-package technologies
  4. Let system demands drive silicon technology
  5. Embrace open-source H/W

- Increasing "nanodiversity" is a good thing
  - More jobs, companies, and students
  - More competition and ***scalable innovation***

Questions