Design and Other ITRS Technologies: Sharing Brick Walls

IEEE SSCS Kansai Chapter
October 17, 2001

Andrew B. Kahng, UCSD CSE & ECE Departments
e-mail: abk@ucsd.edu
URL: http://vlsicad.ucsd.edu
• Apologies:
  – My slides have too many words
  – My intention was to talk about only parts of the slides, and leave the rest for reading later
  – I am from the Electronic Design Automation (EDA) field, not Solid-State Circuits

• This talk: how semiconductor Design technology and Manufacturing technology must work with each other
• The context for this talk is the Roadmap (ITRS)
• Design brings together all other technologies

• If I go too fast, or speak too fast, please tell me
• Please ask your questions
Outline

1. Background: ITRS and system drivers
2. Design productivity gap
3. Vicious cycle $\rightarrow$ virtuous cycle?
4. Sharing red bricks
5. Design-manufacturing handoff
6. Variability and value
7. Conclusion
MESSAGE 1.

• ITRS = International International Technology Roadmap for Semiconductors (http://public.itrs.net)
• ITRS is like a car
• Before, two drivers (husband = MPU, wife = DRAM)
• But, the drivers looked mostly in the rear-view mirror, they did not touch the steering wheel, and they left the car on cruise control (destination = “Moore’s Law”)
• Problem: many passengers in the car (ASIC, SOC, Analog, Mobile, Low-Power, Networking/Wireless, …) wanted to go different places
• This year:
  – Some passengers became drivers!
  – All drivers must explain more clearly where they are going
Background: ITRS Acceleration and System Drivers

Roadmap Changes Since 2000

• Next “node” = 0.7x half-pitch or minimum feature size
  – → 2x transistors on the same size die
• 90nm node in 2004 (100nm in 2003)
  – 90nm node → physical gate length = 45nm
• MPU/ASIC half-pitch = DRAM half-pitch in 2004
• Psychology: everyone must beat the Roadmap
  – Reasons: density, cost reduction, competitive position
  – TSMC CL010G logic/mixed-signal SOC process: risk production in 4Q02 with multi-Vt, multi-oxide, embedded DRAM and flash, low standby power derivatives, …
System Drivers

- New Chapter in 2001 ITRS
- IC products that drive manufacturing and design technologies
- Overall Roadmap Technology Characteristics + System Drivers = “consistent framework for technology requirements”
- Four system drivers
  - MPU = traditional microprocessor core (large design team, digital CMOS)
  - SOC = three types = three different drivers
    - multi-technology (heterogeneous integration, e.g., analog/mixed-signal)
    - high-performance (high-speed I/O / clock frequencies, e.g., networks)
    - low-cost/low-power (productivity, power)
  - AM/S = four basic circuits (LNA, VCO, PA, ADC) + figures of merit
  - DRAM
MPU Driver

• Two MPU flavors
  – Cost-performance: constant 140 mm² die, “desktop”
  – High-performance: constant 310 mm² die, “server”
  – (Next ITRS: merged desktop-server, mobile flavors)
  – MPU organization: multiple cores, on-board L3 cache
    • More dedicated, less general-purpose logic
    • More cores help power management (lower frequency, lower Vdd, more parallelism → overall power savings)
    • Reuse of cores helps design productivity
    • Redundancy helps yield and fault-tolerance
    • MPU and SOC converge (organization and design methodology)

• Double transistor count each node, not each 18 months
  – “Moore’s Law” may slow down

• No more doubling of clock frequency at each node
• Area of “lead” processor is 2-3X area of “shrink” of previous generation processor
• Performance is only 1.5X better
• “On the wrong side of a square law”
- FO4 INV = inverter driving 4 identical inverters (no interconnect)
- Half of frequency improvement came from reducing logic stages
- Other extra performance came from slower Vdd scaling, but this costs too much power
MPU Supporting Analyses

• Diminishing returns
  – “Pollack’s Rule”: In a given node, new microarchitecture takes 2-3x area of previous generation one, but provides only 50% more performance

• Power knob running out
  – Speed from Power: scale voltage by 0.85x instead of 0.7x per node
  – Large switching currents, large power surges on wakeup, IR drop issues
  – Limited by Assembly and Packaging roadmap (bump pitch, package cost)
  – Limited by cost (e.g., system cost increases by $1 per watt)
  – Power management: 2500% improvement needed by 2016

• Speed knob running out
  – 2x frequency per node: 1.4x from scaling, 1.4x from fewer logic stages
  – But: clocks cannot be generated with period < 6-8 FO4 INV delays
  – Pipelining overhead (1-1.5 FO4 INV delay for pulse-mode latch, 2-3 for FF)
  – 14 - 16 FO4 INV delays limit for clock period in core (L1 cache, 64b add)
  – **Cannot continue 2x frequency per node trend**
## SOC-LP (PDA) Driver - STRJ-WG1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Technology (nm)</td>
<td>130</td>
<td>90</td>
<td>65</td>
<td>45</td>
<td>32</td>
<td>22</td>
</tr>
<tr>
<td>Operation Voltage (V)</td>
<td>1.2</td>
<td>1</td>
<td>0.8</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Clock Frequency (MHz)</td>
<td>150</td>
<td>300</td>
<td>450</td>
<td>600</td>
<td>900</td>
<td>1200</td>
</tr>
<tr>
<td>Application</td>
<td>Still Image Processing</td>
<td>Real Time Video Codec</td>
<td>Real Time Interpretation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(MAX performance required)</td>
<td></td>
<td>(MPEG4/CIF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td>Web Browser</td>
<td>TV Telephone (1:1)</td>
<td>TV Telephone (&gt;3:1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Others)</td>
<td>Electric Mailer</td>
<td>Voice Recognition (Input)</td>
<td>Voice Recognition (Operation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scheduler</td>
<td>Authentication (Crypto Engine)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Processing Perf (GOPS)</td>
<td>0.3</td>
<td>2</td>
<td>15</td>
<td>103</td>
<td>720</td>
<td>5042</td>
</tr>
<tr>
<td>Average Power (W) (req'd)</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Standby power (mW) (req'd)</td>
<td>2.1</td>
<td>2.1</td>
<td>2.1</td>
<td>2.1</td>
<td>2.1</td>
<td>2.1</td>
</tr>
</tbody>
</table>

- Driver for power management and low-power device roadmap
- Driver for design productivity and core-based design
  - \(\text{GOPS} / \text{Frequency} = \text{Processing Logic}: \text{increase} 4\times \text{per node}\)
SOC-LP (Low-Power PDA) Driver

• Power management challenge
  – Reduce dynamic and static power to avoid “zero logic content”
  – Necessary tool: low-power process (→ PIDS low-power device roadmap)
  – Slower, less leaky devices: Lgate lags high-performance by 2 years; higher Vth, Vdd, Tox, tau (CV/I) – see next slide
  – Low Operating Power (LOP) and Low Standby Power flavors → design tools handle multi (Vt,Tox,Vdd) (= “unscaled devices” – for analog also)

• Design productivity challenge
  – Processing logic increases 4x per node; die size increases 20% per node

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>½ Pitch</td>
<td>130</td>
<td>90</td>
<td>65</td>
<td>45</td>
<td>32</td>
<td>22</td>
</tr>
<tr>
<td>Logic Mtx per designer-year</td>
<td>1.2</td>
<td>2.6</td>
<td>5.9</td>
<td>13.5</td>
<td>37.4</td>
<td>117.3</td>
</tr>
<tr>
<td>Dynamic power reduction (X)</td>
<td>0</td>
<td>1.5</td>
<td>2.5</td>
<td>4</td>
<td>7</td>
<td>20</td>
</tr>
<tr>
<td>Standby power reduction (X)</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>39</td>
<td>150</td>
<td>800</td>
</tr>
</tbody>
</table>
### LP Device Roadmap

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>99</th>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>10</th>
<th>13</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tox (nm)</td>
<td>MPU</td>
<td>3.00</td>
<td>2.30</td>
<td>2.20</td>
<td>2.20</td>
<td>2.00</td>
<td>1.80</td>
<td>1.70</td>
<td>1.70</td>
<td>1.30</td>
<td>1.10</td>
<td>1.00</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>LOP</td>
<td>3.20</td>
<td>3.00</td>
<td>2.2</td>
<td>2.0</td>
<td>1.8</td>
<td>1.6</td>
<td>1.4</td>
<td>1.3</td>
<td>1.2</td>
<td>1.0</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>LSTP</td>
<td>3.20</td>
<td>3.00</td>
<td>2.6</td>
<td>2.4</td>
<td>2.2</td>
<td>2.0</td>
<td>1.8</td>
<td>1.6</td>
<td>1.4</td>
<td>1.1</td>
<td>1.0</td>
<td>0.9</td>
</tr>
<tr>
<td>Vdd</td>
<td>MPU</td>
<td>1.5</td>
<td>1.3</td>
<td>1.2</td>
<td>1.1</td>
<td>1.0</td>
<td>0.9</td>
<td>0.9</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>LOP</td>
<td>1.3</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
<td>1.0</td>
<td>0.9</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>LSTP</td>
<td>1.3</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>1.1</td>
<td>1.0</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>Vth (V)</td>
<td>MPU</td>
<td>0.21</td>
<td>0.19</td>
<td>0.19</td>
<td>0.15</td>
<td>0.13</td>
<td>0.12</td>
<td>0.09</td>
<td>0.06</td>
<td>0.05</td>
<td>0.021</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>LOP</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
<td>0.35</td>
<td>0.36</td>
<td>0.32</td>
<td>0.33</td>
<td>0.34</td>
<td>0.29</td>
<td>0.29</td>
<td>0.25</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>LSTP</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
<td>0.52</td>
<td>0.53</td>
<td>0.53</td>
<td>0.54</td>
<td>0.55</td>
<td>0.52</td>
<td>0.49</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>Ion (uA/µm)</td>
<td>MPU</td>
<td>1041</td>
<td>1022</td>
<td>926</td>
<td>959</td>
<td>967</td>
<td>954</td>
<td>924</td>
<td>960</td>
<td>1091</td>
<td>1250</td>
<td>1492</td>
<td>1507</td>
</tr>
<tr>
<td></td>
<td>LOP</td>
<td>636</td>
<td>591</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>700</td>
<td>800</td>
<td>900</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LSTP</td>
<td>300</td>
<td>300</td>
<td>300</td>
<td>300</td>
<td>400</td>
<td>400</td>
<td>400</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>600</td>
<td>800</td>
</tr>
<tr>
<td>CV/I (ps)</td>
<td>MPU</td>
<td>2.00</td>
<td>1.64</td>
<td>1.63</td>
<td>1.34</td>
<td>1.16</td>
<td>0.99</td>
<td>0.86</td>
<td>0.79</td>
<td>0.66</td>
<td>0.39</td>
<td>0.23</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>LOP</td>
<td>3.50</td>
<td>2.87</td>
<td>2.55</td>
<td>2.45</td>
<td>2.02</td>
<td>1.84</td>
<td>1.58</td>
<td>1.41</td>
<td>1.14</td>
<td>0.85</td>
<td>0.56</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>LSTP</td>
<td>4.21</td>
<td>3.46</td>
<td>4.61</td>
<td>4.41</td>
<td>2.96</td>
<td>2.68</td>
<td>2.51</td>
<td>2.32</td>
<td>1.81</td>
<td>1.43</td>
<td>0.91</td>
<td>0.57</td>
</tr>
<tr>
<td>Ioff (uA/µm)</td>
<td>MPU</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.07</td>
<td>0.10</td>
<td>0.30</td>
<td>0.70</td>
<td>1.00</td>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>LOP</td>
<td>1e-4</td>
<td>1e-4</td>
<td>1e-4</td>
<td>1e-4</td>
<td>1e-4</td>
<td>1e-4</td>
<td>3e-4</td>
<td>3e-4</td>
<td>7e-4</td>
<td>1e-3</td>
<td>3e-3</td>
<td>1e-2</td>
</tr>
<tr>
<td></td>
<td>LSTP</td>
<td>1e-6</td>
<td>1e-6</td>
<td>1e-6</td>
<td>1e-6</td>
<td>1e-6</td>
<td>1e-6</td>
<td>1e-6</td>
<td>1e-6</td>
<td>3e-6</td>
<td>7e-6</td>
<td>1e-5</td>
<td></td>
</tr>
<tr>
<td>Gate L (nm)</td>
<td>MPU</td>
<td>100</td>
<td>70</td>
<td>65</td>
<td>53</td>
<td>45</td>
<td>37</td>
<td>32</td>
<td>30</td>
<td>25</td>
<td>18</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>L(*)P</td>
<td>110</td>
<td>100</td>
<td>90</td>
<td>80</td>
<td>65</td>
<td>53</td>
<td>45</td>
<td>37</td>
<td>32</td>
<td>22</td>
<td>16</td>
<td>11</td>
</tr>
</tbody>
</table>
Outline

1. Background: ITRS and system drivers
2. Design productivity gap
3. Vicious cycle $\rightarrow$ virtuous cycle?
4. Sharing red bricks
5. Design-manufacturing handoff
6. Variability and value
7. Conclusion
MESSAGE 2.

• “Design Productivity Gap” = “failure of Design Technology”

• Number of available transistors grows faster than designer ability to design them well
  – → Increased design effort, risk, turnaround time (TAT)
  – → fewer designs are worth trying

• Manufacturing non-recurring engineering (NRE) cost also increasing (mask set)
  – → fewer designs are worth trying

• “Workarounds” sacrifice quality, value of designs
  – → even with workarounds, fewer designs worth trying

• This is a semiconductor industry problem, not an EDA problem
## Productivity Gap (1994)

### Potential Design Complexity and Designer Productivity

#### Equivalent Added Complexity
- 68% per year compounded
- Complexity growth rate: 21% per year compounded

#### Logic Transistor per Chip

<table>
<thead>
<tr>
<th>Year</th>
<th>Technology</th>
<th>Chip Complexity</th>
<th>Frequency</th>
<th>Staff</th>
<th>Staff Cost*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>250 nm</td>
<td>13 M Tr.</td>
<td>400 MHz</td>
<td>210</td>
<td>90 M</td>
</tr>
<tr>
<td>1998</td>
<td>250 nm</td>
<td>20 M Tr.</td>
<td>500</td>
<td>270</td>
<td>120 M</td>
</tr>
<tr>
<td>1999</td>
<td>180 nm</td>
<td>32 M Tr.</td>
<td>600</td>
<td>360</td>
<td>160 M</td>
</tr>
<tr>
<td>2002</td>
<td>130 nm</td>
<td>130 M Tr.</td>
<td>800</td>
<td>800</td>
<td>360 M</td>
</tr>
</tbody>
</table>

* @ $150 k / Staff Yr. (In 1997 Dollars)

Source: SEMATECH

Andrew Kahng – October 2001
“$1M (= 10^8 Yen) mask set” in 100nm, but average only 500 wafers per set
The Implementation Gap

System Complexity: Need to raise the handoff level to improve productivity
Silicon Complexity: More nanometer implementation details

source: MARCO GSRC
Andrew Kahng – October 2001
Closing the Implementation Gap: How?

Effort/Value

Level of Abstraction

Application

Design Entry Level

Hand-off “platform”

SW/HW

RTL

Mask

source: MARCO GSRC

Andrew Kahng – October 2001
### Low-Value Designs?

Percent of die area that must be occupied by memory to maintain SOC design productivity (STRJ-WG1 scenario published in ITRS-2000 update)

An all-memory design is probably a low-value design.
Reduced Back-End Effort?

Example: regular shielded wiring fabric pattern at minimum pitch

- Eliminates signal integrity, delay uncertainty concerns
- But has at least 60% - 80% density cost
Improved Reuse Productivity?

Example: “communication-based design”

Pearls (the IP Processes)
MicroShells (the IP Requirements)
MacroShells (the Protocol Interface)
Communication Channels

source: MARCO GSRC
Andrew Kahng – October 2001
But: Quality Trades Off With Flexibility

Source: Prof. Jan Rabaey, UC Berkeley
“What If Design Technology Fails?”

• Role of Design Technology: “Fill the fab”
  – keep manufacturing facilities fully utilized with high-volume parts, high-value (= high-margin) parts

• “When design technology fails”
  – not enough high-value designs
  – semiconductor industry looks for a “workaround”
    • reconfigurable logic
    • platform-based design
    • extract value somewhere other than silicon differentiation

• What about:
  – Electronics industry looks for a “workaround”?  
    • extract value somewhere other than silicon?
Design and Manufacturing In Same Boat

• Design productivity gap
  – Threatens design quality
  – This is really a design technology productivity gap

• Design starts, ASIC business models at risk
  – More reprogrammable, platform-based “workarounds”
  – More software workarounds
  – → Why retool?

2001 ITRS: “Cost of design is the greatest threat to continuation of the semiconductor roadmap.”
Outline

• 1. Background: ITRS and system drivers
• 2. Design productivity gap
• 3. Vicious cycle → virtuous cycle?
• 4. Sharing red bricks
• 5. Design-manufacturing handoff
• 6. Variability and value
• 7. Conclusion
MESSAGE 3.

• Fact 1. Design is the bottleneck
• Fact 2. Investment in Design Technology is low
  – We may think “things are okay”
  – However, there are many crises in 2001
• Why this contradiction?

• How can we prove that Design Technology merits investment?
Mystery

• Fact 1. Design technology is a bottleneck for the semiconductor industry.

• Fact 2. Investment in process technology is much greater than investment in design technology.

• Good News: Progress in design technology continues
Design Cost of SOC-LP PDA Driver

SOC Design Cost Model

- RTL Methodology Only
- With all Future Improvements

Year
- 1985
- 1990
- 1995
- 2000
- 2005
- 2010
- 2015
- 2020

Total Design Cost (log scale)
- $100,000,000,000
- $10,000,000,000
- $1,000,000,000
- $100,000,000
- $10,000,000
- $1,000,000
- $100,000
- $10,000
- $1,000
- $100
- $10
- $1

Andrew Kahng – October 2001
Design Cost Model (ITRS-2001)

- Engineer cost per year increases 5% per year ($181,568 in 1990)
- EDA tool cost per year (per engineer) increases 3.9% per year ($99,301 in 1990) (+ separate term for interoperability)
- Productivity due to 8 major Design Technology innovations (3.5 of which are still unavailable): RTL methodology; In-house P&R; Tall-thin engineer; Small-block reuse; Large-block reuse; IC implementation suite; Intelligent testbench; Electronic System-level methodology
- Matched up against SOC-LP PDA content:
  - SOC-LP PDA design cost = $15M (= 1.5B Yen) in 2001
  - Would have been $342M without EDA innovations and the resulting improvements in design productivity
Mystery

• Fact 1. Design technology is a bottleneck for the semiconductor industry.

• Fact 2. Investment in process technology is much greater than investment in design technology.

• Bad News: In 2001, many design technology gaps have become crises
• 2-3X more verification engineers than designers on microprocessor teams
• Software = 80% of system development cost (and Analog design hasn’t scaled)
• Design NRE > 10’s of $M (B’s of Yen) ↔ manufacturing NRE $1M (100M Y)
• Design TAT = months or years ↔ manufacturing TAT = weeks
• Test cost per transistor grows exponentially relative to mfg cost
Mystery

• Fact 1. Design technology is a bottleneck for the semiconductor industry.

• Fact 2. Investment in process technology is much greater than investment in design technology.

• Why this contradiction?
Hold These Thoughts...

• ITRS is created by worldwide semi/system houses
  – EDA’s star customers
• EDA in the big picture
  – Has one chapter out of 12 in ITRS
  – Is just one part of SISA (semiconductor industry supplier association
  – Is small: 6000 R&D worldwide, $4B (400B Yen) total market
• EDA growth
  – Dataquest: 3.9% annual growth in tools $ spent per designer
  – integration costs > tool costs
• Hold these thoughts:
  – “A small industry with poor perceived ROI will stay small” is a “vicious cycle”
  – How do we turn a vicious cycle into a “virtuous cycle”?  

Andrew Kahng – October 2001
How to Achieve the Virtuous Cycle?

• Passive / Negative Approach  (NO !!!)
  – (senior manager at major EDA company, IEEE CANDE Workshop, 9/2001):
    “Rising NRE will force semiconductor manufacturers to produce primarily high-volume, general purpose components such as memory, FPGAs, and standard processors. New EDA tools will then have an impact on only a smaller fraction of the semiconductor industry, and research funding will evaporate, leaving only the service and support functions, which don’t need to be centralized. Prediction: EDA industry is reduced to a service role as semiconductor design starts decline.”
  – ICCAD, DAC, etc. panels: “Why doesn’t EDA get any respect?”

• Active / Positive Approach  (YES !!!)
  – Understand cost and value of Design Technology
  – Prove EDA ROI
Outline

• 1. Background: ITRS and system drivers
• 2. Design productivity gap
• 3. Vicious cycle \rightarrow virtuous cycle?
• 4. Sharing red bricks
• 5. Design-manufacturing handoff
• 6. Variability and value
• 7. Conclusion
MESSAGE 4.

- ITRS technologies are like parts of the car
- Every one takes the “engine” point of view when it defines its requirements

- All parts must work together to make the car go smoothly

- But: “The Squeaky Wheel Gets The Grease”
  – (Design Technology has never squeaked loudly…)

- Need “global optimization” of requirements
What Is A “Red Brick”?

• Red Brick = ITRS Technology Requirement with no known solution

• Alternate definition: Red Brick = something that REQUIRES billions of dollars ($1B = 10^{11} \text{ Yen}) in R&D investment

• Observation: Design Technology “is different”, and has never stated any meaningful red bricks in the ITRS
### Table **High Frequency Serial Communications Test Requirements—Near Term**

<table>
<thead>
<tr>
<th>Year of Production</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>Driver</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DRAM ½ Pitch (Sc. 2.0)</strong></td>
<td>130</td>
<td>115</td>
<td>100</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td><strong>MPU ½ Pitch (Sc. 3.7)</strong></td>
<td>150</td>
<td>130</td>
<td>105</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td><strong>MPU Printed Gate Length (Sc. 3.7)</strong></td>
<td>90</td>
<td>75</td>
<td>65</td>
<td>53</td>
<td>45</td>
<td>40</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td><strong>MPU Physical Gate Length (Sc. 3.7)</strong></td>
<td>65</td>
<td>53</td>
<td>45</td>
<td>37</td>
<td>32</td>
<td>30</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>

#### High-performance-level serial transceivers

| Serial data rate (Gbits/s)       | 10   | 10   | 40   | 40   | 40   | 40   | 40   |        |
| Maximum Reference Clock Speed (MHz) | 667  | 667  | 2500 | 2500 | 2500 | 2500 | 2500 |        |

#### High-integration-level backplane and computer I/O

| Serial data rate (Gbits/s)       | 2.5  | 3.125 | 3.125 | 10   | 10   | 40   | 40   |        |
| Port count                       | 20   | 100   | 200   | 100  | 200  | 100  | 200  |        |
| Maximum Reference Clock Speed (MHz) | 166  | 166   | 166   | 667  | 667  | 2500 | 2500 | *      |

*White—Manufacturable Solutions Exist, and Are Being Optimized*

*Yellow—Manufacturable Solutions are Known*

*Red—Manufacturable Solutions are NOT Known*
2001 Big Picture = Big Opportunity

• Why ITRS has “red brick” problems
  – “Wrong” Moore’s Law
    • Frequency and bits are not the same as efficiency and utility
    • No awareness of applications or architectures (only Design is aware)
  – Independent, “linear” technological advances don’t work
    • Car has more drivers (mixed-signal, mobile, etc. applications)
    • Every car part thinks that it is the engine → too many red bricks
  – No clear ground rules
    • Is cost a consideration? Is the Roadmap only for planar CMOS?

• New in 2001: Everyone asks “Can Design help us?”
  – Process Integration, Devices & Structures (PIDS): 17%/year improvement in CV/I metric → sacrifice Ioff, Rds, …analog, LOP, LSTP, … many flavors
  – Assembly and Packaging: cost limits → keep bump pitches high → sacrifice IR drop, signal integrity (impacts Test as well)
  – Interconnect, Lithography, PIDS/Front-End Processes: What variability can Designers tolerate? 10%? 15%? 25%?
“Design-Manufacturing Integration”

• 2001 ITRS Design Chapter: “Manufacturing Integration” = one of five Cross-Cutting Challenges

• Goal: **share** red bricks with other ITRS technologies
  – Lithography CD variability requirement → new Design techniques that can better handle variability
  – Mask data volume requirement → solved by Design-Mfg interfaces and flows that pass functional requirements, verification knowledge to mask writing and inspection
  – ATE cost and speed red bricks → solved by DFT, BIST/BOST techniques for high-speed I/O, signal integrity, analog/MS
  – Does “X initiative” have as much impact as copper?
<table>
<thead>
<tr>
<th>YEAR</th>
<th>TECHNOLOGY NODE</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DRAM ½ PITCH (nm) (Sc. 2.0)</td>
<td>130</td>
<td>115</td>
<td>100</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>MPU/ASIC ½ PITCH (nm) (Sc. 3.7)</td>
<td>150</td>
<td>130</td>
<td>107</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>MPU PRINTED GATE LENGTH (nm) (Sc. 3.7)</td>
<td>90</td>
<td>75</td>
<td>65</td>
<td>53</td>
<td>45</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>MPU PHYSICAL GATE LENGTH (nm) (Sc. 3.7)</td>
<td>65</td>
<td>53</td>
<td>45</td>
<td>37</td>
<td>32</td>
<td>28</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Conductor effective resistivity (µΩ-cm) Cu intermediate wiring*</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>Barrier/cladding thickness (for Cu intermediate wiring) (nm)</td>
<td>18</td>
<td>15</td>
<td>13</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Interlevel metal insulator —effective dielectric constant (κ)</td>
<td>3.0–3.7</td>
<td>3.0–3.7</td>
<td>2.9–3.5</td>
<td>2.5–3.0</td>
<td>2.5–3.0</td>
<td>2.5–3.0</td>
<td>2.0–2.5</td>
</tr>
<tr>
<td></td>
<td>Interlevel metal insulator (minimum expected) —bulk dielectric constant (κ)</td>
<td>2.7</td>
<td>2.7</td>
<td>2.7</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Bulk and effective dielectric constants
Porous low-k requires alternative planarization solutions
Cu at all nodes - conformal barriers

Do we really need this?

C. Case, BOC Edwards – ITRS-2001 preliminary
Conductor resistivity increases expected to appear around 100 nm linewidth - will impact intermediate wiring first - ~ 2006

Courtesy of SEMATECH

Andrew Kahng – October 2001

C. Case, BOC Edwards – ITRS-2001 preliminary
Cost of Manufacturing Test

Is this better solved with Automated Test Equipment technology, or with Design (for Test, Built-In Self-Test)?

Is this even solvable with ATE technology alone?
PIDS (Devices/Structures)

- CV/I trend (17% per year improvement) = “constraint”
- Huge increase in subthreshold $I_{\text{off}}$
  - Room temperature: increases from 0.01 uA/um in 2001 to 10 uA/um at end of ITRS (22nm node)
    - At operating temperatures (100 – 125 deg C), increase by 15 - 40x
  - Standby power challenge
    - Manage multi-$V_t$, multi-$V_{dd}$, multi-Tox in same core
    - Aggressive substrate biasing
    - Constant-throughput power minimization
    - Modeling and controls passed to operating system and applications

- Aggressive reduction of Tox
  - Physical Tox thickness < 1.4nm (down to 1.0nm) starting in 2001, even if high-k gate dielectrics arrive in 2004
  - Variability challenge: “10%” < one atomic monolayer
Assembly and Packaging

- **Goal**: cost control ($0.07/pin, $2 package, …)
- “Grand Challenge” for A&P: work with Design to develop die-package co-analysis, co-optimization tools
- **Bump/pad counts scale with chip area only**
  - Effective bump pitch roughly constant at 300um
  - MPU pad counts flat from 2001-2005, but chip current draw increases 64%
- **IR drop control challenge**
  - Metal requirements explode with $I_{chip}$ and wiring resistance
- **Power challenge**
  - 50 W/cm² limit for forced-air cooling; MPU area becomes flat because power budget is flat
  - More control (e.g., dynamic frequency and supply scaling) given to OS and application
  - Long-term: Peltier-type thermoelectric cooling, …→ design must know
Manufacturing Test

• High-speed interfaces (networking, memory I/O)
  – Frequencies on same scale as overall tester timing accuracy

• Heterogeneous SOC design
  – Test reuse
  – Integration of distinct test technologies within single device
  – Analog/mixed-signal test

• Reliability screens failing
  – Burn-in screening not practical with lower Vdd, higher power budgets → overkill impact on yield

• Design challenges: DFT, BIST
  – Analog/mixed-signal
  – Signal integrity and advanced fault models
  – BIST for single-event upsets (in logic as well as memory)
  – Reliability-related fault tolerance
Lithography

• 10% CD uniformity is a red brick today
• 10% < 1 atomic monolayer at end of ITRS
• This year: Lithography, PIDS, FEP agreed to raise CD uniformity requirement to 15% (but still a red brick)

• Design for variability
  – Novel circuit topologies
  – Circuit optimization (conflict between slack minimization and guardbanding of quadratically increasing delay sensitivity)
  – Centering and design for $/wafer

• Design for when devices, interconnects no longer 100% guaranteed correct?
  – Potentially huge savings in manufacturing, verification, test costs
How to Share Red Bricks

- **Cost** is the biggest missing link within the ITRS
  - Manufacturing cost (silicon cost per transistor)
  - Manufacturing NRE cost (mask, probe card, …)
  - Design NRE cost (engineers, tools, integration, …)
  - Test cost
  - **Technology development cost** → who should solve a given red brick wall?

- Return On Investment (ROI) = \( \frac{\text{Value}}{\text{Cost}} \)
  - Value needs to be defined ("design quality", "time-to-market")

- Understanding **cost** and ROI allows sensible sharing of red bricks across industries
Outline

• 1. Background: ITRS and system drivers
• 2. Design productivity gap
• 3. Vicious cycle → virtuous cycle?
• 4. Sharing red bricks
• 5. Design-manufacturing handoff
• 6. Variability and value
• 7. Conclusion
MESSAGE 5.

• Manufacturing handoff (to mask flow) is complicated and expensive because of “reticle enhancement techniques” (RET)

• RET examples: Optical Proximity Correction (OPC), Phase-Shifting Masks (PSM)

• To reduce mask complexity, write time, and verification time (= mask NRE cost), we need smarter handoff from design to manufacturing

• Other manufacturing interfaces (process models, libraries, etc.) are also critical, but not discussed
Subwavelength Optical Lithography

- WYSIWYG (layout = mask = wafer) failed starting with 350nm generation
- Optical lithography: feature size limited by diffraction
- Available knobs
  - aperture: OPC
  - phase: PSM
Optical Proximity Correction (OPC)

- Aperture changes to improve process control
  - improve yield (process window)
  - improve device performance
OPC Terminology

- Serif
- Hammerhead
- Gate Assist Features
- Gate Biasing
- Diffusion
- Polysilicon
Phase Shifting Masks (PSM)

**Conventional Mask**
- Glass
- Chrome
- Phase shifter
- 0 E at mask 0
- 0 E at wafer 0
- 0 L at wafer 0

**Phase Shifting Mask**
- Glass
- Chrome
- Phase shifter
- 0 E at mask 0
- 0 E at wafer 0
- 0 L at wafer 0
Many Other Optical Litho Issues

- Example: **Field-dependent aberrations cause placement errors and distortions.**

\[
\text{CELL}_A(X_1, Y_1) \neq \text{CELL}_A(X_0, Y_0) \neq \text{CELL}_A(X_2, Y_2)
\]
RET Roadmap

Rule-based OPC
Model-based OPC
Scattering Bars
AA-PSM
Weak PSM
Rule-based Tiling
Optimization-driven MB Tiling

Number Of Affected Layers Increases / Generation

248 nm
248/193 nm
193 nm

W. Grobman, Motorola – DAC-2001
Optical Lithography Becomes Harder

• Process window and yield enhancement
  – Forbidden width-spacing combinations (defocus window sensitivities)
  – Complex “local DRCs”

• Lithography equipment choices (e.g., off-axis illumination)
  – Forbidden configurations (wrong-way critical-width doglegs, or diagonal features)

• OPC subresolution assist features (scattering bars)
  – Notch rules, critical-feature rules on local metal
Context-Dependent Fracturing

Same pattern, different fracture
ITRS Maximum Single Layer File Size

ITRS 2000 prediction for maximum single layer data volume in uncompressed MEBES format.

Year

MEBES Data Volume (GB)

180 nm
130 nm
90 nm
ALTA-3500 Mask Write Time

Write Time (Reformat + Print) (Hrs)

ABF Data Volume (MB)

P. Buck, Dupont Photomasks – ISMT Mask-EDA Workshop July 2001
Mask Data and $1M (= 10^8$ Yen) Mask NRE

- Too many data formats
  - Most tools have unique data format
  - Raster to variable shaped-beam conversion is inefficient
  - Real-time manufacturing tool switch, multiple qualified tools → duplicate fractures to avoid delays if tool switch required

- Data volume
  - OPC increases figure count acceleration
  - MEBES format is flat
  - ALTA machines (mask writers) slow down with > 1GB data
  - Data volume strains distributed manufacturing resources

- Refracturing mask data
  - Before: mask industry never touched mask data (risky, no good reason)
  - Today: 90% of mask data files manipulated or refractured: process bias sizing (iso-dense, loading effects, linearity, ...), mask write optimization, multiple tool formats, ...
Shared Red Bricks for Mask Handoff

- WYSIWYG broken \(\rightarrow\) (mask) verification bottleneck
- Need function- and cost-aware OPC, PSM, dummy fill
  - Real goal = predictable circuit performance and function
  - Therefore, tools must understand functional intent
    - make only corrections that gain $$$, reduce performance variation
    - make only corrections that can be manufactured and verified (including mask inspection)
    - understand (data volume, verification) costs of breaking hierarchy
  - Understand flow issues
    - e.g., avoid making same corrections 3x (library, router, PV tool)
- Need much more than GDSII in manufacturing interface
  - Includes sensitivities to patterning variation / error
  - Guided by models of manufacturing equipment
  - Mask verification needs to know same function, sensitivity info
- Manufacturing NRE vital to mask, ASIC industries
Outline

• 1. Background: ITRS and system drivers
• 2. Design productivity gap
• 3. Vicious cycle → virtuous cycle?
• 4. Sharing red bricks
• 5. Design-manufacturing handoff
• 6. Variability and value
• 7. Conclusion
MESSAGE 6.

- Design Technology must be able to measure its value

- One example measure of value is $ per wafer

- To measure this, we need (1) detailed models of process variability, and (2) models of how chip parameters (frequency, testability, etc.) affect value
**Process Variation Sources**

- **Design** → (manufacturing variability) → **Value**
  - **Intrinsic variations**
    - *Systematic*: due to predictable sources, can be compensated during design stage
    - *Random*: inherently unpredictable fluctuations and cannot be compensated
  - **Dynamic variations**
    - Stem from circuit operation, including supply voltage and temperature fluctuations
    - Depend on circuit activity and hard to be compensated
  - **Correlations**
    - $V_{th0} = V_{fb} + 2\varphi_B + \frac{|Q_{dep}|}{\varepsilon_{ox}} \cdot T_{ox}$
    - Line width and spacing are anti-correlated by one; ILD and interconnect thickness also anti-correlated
# Technology Trend Over Generations

<table>
<thead>
<tr>
<th>Technology</th>
<th>180nm</th>
<th>130nm</th>
<th>100nm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Device</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leff (µm)</td>
<td>0.10 ±15%</td>
<td>0.12 ±15%</td>
<td>0.09 ±15%</td>
</tr>
<tr>
<td>Tox (nm)</td>
<td>40 ± 4%</td>
<td>42 ± 4%</td>
<td>33 ± 4%</td>
</tr>
<tr>
<td>Vth0 (V)</td>
<td>0.40 ±12.5%</td>
<td>-0.42 ±12.5%</td>
<td>0.27 ±15.5%</td>
</tr>
<tr>
<td>Rsw (Ω/ )</td>
<td>250 ±10%</td>
<td>450 ±10%</td>
<td>200 ±10%</td>
</tr>
<tr>
<td><strong>Interconnect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ε</td>
<td>3.5 ± 3%</td>
<td>3.2 ± 5%</td>
<td>2.8 ± 5%</td>
</tr>
<tr>
<td>w (µm)</td>
<td>0.28 ± 20%</td>
<td>0.80 ± 20%</td>
<td>0.20 ± 20%</td>
</tr>
<tr>
<td>s (µm)</td>
<td>0.28 ± 20%</td>
<td>0.80 ± 20%</td>
<td>0.20 ± 20%</td>
</tr>
<tr>
<td>t (µm)</td>
<td>0.45 ± 10%</td>
<td>1.25 ± 10%</td>
<td>0.45 ± 10%</td>
</tr>
<tr>
<td>ILDh (µm)</td>
<td>0.65 ± 15%</td>
<td>1.80 ± 15%</td>
<td>0.45 ± 15%</td>
</tr>
<tr>
<td>Rsw (Ω)</td>
<td>46 ± 20%</td>
<td>50 ± 20%</td>
<td>54 ± 20%</td>
</tr>
<tr>
<td>Length (µm)</td>
<td>61.01</td>
<td>1061</td>
<td>45.19</td>
</tr>
<tr>
<td>Wn/Ln (µm)</td>
<td>1.26/0.18</td>
<td>20/0.18</td>
<td>0.91/0.13</td>
</tr>
<tr>
<td><strong>Dynamic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temp (°C)</td>
<td>25-100</td>
<td></td>
<td>25/100</td>
</tr>
<tr>
<td>Vdd (V)</td>
<td>1.8 ± 10%</td>
<td></td>
<td>1.5 ± 10%</td>
</tr>
<tr>
<td>Tr (ps)</td>
<td>160</td>
<td></td>
<td>95</td>
</tr>
</tbody>
</table>

- Values are from ITRS, BPTM, and industry; red is $3\sigma$
- From ongoing work at UCSD/UCB/Michigan; some values are wrong (e.g., Rsw)

Andrew Kahng – October 2001
## Copper CMP Variability in Near Term

### Annual Technology Node Summary

<table>
<thead>
<tr>
<th>Year Technology Node</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRAM ½ Pitch (nm) (Sc. 2.0)</td>
<td>130</td>
<td>115</td>
<td>100</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td>65</td>
</tr>
<tr>
<td>MPU/ASIC ½ Pitch (nm) (Sc. 3.7)</td>
<td>150</td>
<td>130</td>
<td>107</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td>65</td>
</tr>
<tr>
<td>MPU Printed Gate Length (nm) (Sc. 3.7)</td>
<td>90</td>
<td>75</td>
<td>65</td>
<td>53</td>
<td>45</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td>MPU Physical Gate Length (nm) (Sc. 3.7)</td>
<td>65</td>
<td>53</td>
<td>45</td>
<td>37</td>
<td>32</td>
<td>28</td>
<td>25</td>
</tr>
</tbody>
</table>

### Copper CMP Variability

- **Cu thinning at minimum pitch due to erosion (nm), 10% X height, 50% areal density, 500 µm square array**
- **Cu thinning at minimum intermediate pitch due to erosion (nm), 10% X height, 50% areal density, 500 µm square array**
- **Cu thinning global wiring due to dishing and erosion (nm), 10% X height, 80% areal density, 15 micron wide wire**
- **Cu thinning global wiring due to dishing (nm), 100 micron wide feature**

### Combined dishing/erosion metric for global wires

- **Cu thinning due to dishing for isolated lines/pads**
- **No significant dishing at local levels - thinning due to erosion over large areas (50% areal coverage)**

C. Case, BOC Edwards – ITRS-2001 preliminary
Variation Sensitivities: Local Stage

- Sensitivity evaluated by the percentage change in performance when there is $3\sigma$ variation at the parameter
- For local stage, device variations have larger impact on line delay and interconnect variations have stronger impact on crosstalk noise
Mapping Design to Value (1)

Across-Wafer Frequency Variation
Mapping Design to Value (2)

AMD Processors

Goal: combine (1) and (2), drive Design optimizations
Conclusions

- ITRS-2001: Too many independent red bricks
- Design Technology must actively share red bricks from other technology areas
  - Many possibilities
- Design Technology community must measure itself
  - Value of designs, design tools, design processes
  - Design NRE cost: TAT/TTM, tools, integration, …
  - Return On Investment = Value / Cost
- Virtuous cycle: DT gives better ROI, enables silicon-based product differentiation, achieves higher value
Thank you for your attention!
SPARE / HIDDEN SLIDES
Silicon Complexity Challenges

- Silicon Complexity = impact of process scaling, new materials, new device/interconnect architectures
- Non-ideal scaling (leakage, power management, circuit/device innovation, current delivery)
- Coupled high-frequency devices and interconnects (signal integrity analysis and management)
- Manufacturing variability (library characterization, analog and digital circuit performance, error-tolerant design, layout reusability, static performance verification methodology/tools)
- Scaling of global interconnect performance (communication, synchronization)
- Decreased reliability (SEU, gate insulator tunneling and breakdown, joule heating and electromigration)
- Complexity of manufacturing handoff (reticle enhancement and mask writing/inspection flow, manufacturing NRE cost)
System Complexity Challenges

- System Complexity = exponentially increasing transistor counts, with increased diversity (mixed-signal SOC, …)
- **Reuse** (hierarchical design support, heterogeneous SOC integration, reuse of verification/test/IP)
- **Verification and test** (specification capture, design for verifiability, verification reuse, system-level and software verification, AMS self-test, noise-delay fault tests, test reuse)
- **Cost-driven design optimization** (manufacturing cost modeling and analysis, quality metrics, die-package co-optimization, …)
- **Embedded software design** (platform-based system design methodologies, software verification/analysis, codesign w/HW)
- **Reliable implementation platforms** (predictable chip implementation onto multiple fabrics, higher-level handoff)
- **Design process management** (design team size and geographic distribution, data management, collaborative design support, systematic process improvement)
Cross-Cutting Design Challenges

• Productivity
• Power
• Manufacturing Integration
• Interference
• Error-Tolerance
What does EDA know about process?

Clean abstraction!

TCAD

Process
Develop.:
• Lithography
• Device

Device models
Design rules

Design

GDSII

ECAD
Developmental Fab in Tight Loop

- **Process Requirements**
  - Process Develop.
    - Lithography
    - Device
  - TCAD
  - Semi suppliers
- **Device models**
- **Design rules**
- **Design**
- **Mask**
- **Production Fab**
- **ECAD**

- GDSII, tolerances,...
- tolerances...

Andrew Kahng – October 2001
Density Control for CMP

• Chemical-mechanical planarization (CMP)
  – applied to interlayer dielectrics (ILD) and inlaid metals
  – polishing pad wear, slurry composition, pad elasticity make this a very difficult process step

• Cause of CMP variability
  – pad deforms over metal feature
  – greater ILD thickness over dense regions of layout
  – “dishing” in sparse regions of layout
  – huge part of chip variability budget used up (e.g., 4000Å ILD variation across-die)

• Relationship between layout density, ILD thickness

• Variation controlled by insertion of dummy features into layout