Trends and Challenges in High-Performance Microprocessor Design

Stefan Rusu
Senior Principal Engineer
Intel Corporation

stefan.rusu@intel.com
Agenda

- Microprocessor Design Trends
- Process Technology Directions
- Active Power Management
- Leakage Reduction Techniques
- Thermal Modeling
- Call to Action
- Summary
## Microprocessor Evolution

<table>
<thead>
<tr>
<th></th>
<th>4004 Processor</th>
<th>Itanium® 2 Processor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year</strong></td>
<td>1971</td>
<td>2004</td>
</tr>
<tr>
<td><strong>Transistors</strong></td>
<td>2300</td>
<td>592 M</td>
</tr>
<tr>
<td><strong>Process</strong></td>
<td>10 um</td>
<td>0.13 um</td>
</tr>
<tr>
<td><strong>Die size</strong></td>
<td>12 mm²</td>
<td>432 mm²</td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td>108 kHz</td>
<td>1.7 GHz</td>
</tr>
</tbody>
</table>
Moore’s Law - 1965

The experts look ahead

Cramming more components onto integrated circuits

With unit cost falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 85,000 components on a single silicon chip

By Gordon E. Moore

The future of integrated electronics is the future of electronics itself. The advantages of integration will bring about a proliferation of electronics, pushing this science into many new areas.

Integrated circuits will lead to such wonders as 'home computers' -- as long as multi-bit visual displays are central computer peripherals. Automatic controls for automobiles, and personal portable communication equipment. The electronic world will no longer be a display to be found only in a laboratory.

But the biggest 'pot of gold' is the production of large systems. In telephone communications, integrated circuits, digital filters will separate channels on multiple equipment. Integrated circuits will also switch telephone channels and perform data processing.

Companies will be more mobile, and will be organized in completely different ways. For example, numerous small integrated electronics may be distributed throughout the machine instead of being concentrated in a central unit. In addition, the improved reliability made possible by integrated circuits will allow the construction of large processing units. Machines similar to those in existence today will be built at lower cost and with faster turn-around.

Present and future

By integrated electronics, I mean all the various sub-systems which are referred to as microelectronic systems today as well as any additional ones that result in electronic function supplied to the user in a form suitable for use. These sub-systems were first originated in the late 1960s. The object was to minimize electronic equipment to include the minimum complete electronic function in limited space with minimum weight. Several approaches were evaluated including inter-assembly techniques for individual components, thin-film structures and semiconductor integrated circuits.

Each approach evolved rapidly and encouraged new that might have been predicted before. Many remember the moves by the computer to be a combination of the various approaches.

The advantages of semiconductor integrated circuits are already using the increased characterizing of the dimensions by applying such film directly into active semiconductors. Extending this, the processes of active semiconductor devices will produce film on film.

Both approaches have worked well and are being used in equipment today.

Electronics, April 1965

©2004 Intel Corp.
Past Forecasts

“Heavier-than-air flying machines are not possible”
Lord Kelvin, 1895

“I think there is a world market for maybe five computers”
IBM Chairman Thomas Watson, 1943

“640,000 bytes of memory ought to be enough for anybody”
Bill Gates, 1981

“The Internet will catastrophically collapse in 1996”
Robert Metcalfe
Moore’s Law Continues

Heading Toward 1 Billion Transistors By 2005
Processor Frequency Trend

- Frequency doubles each generation
- Number of gates per clock reduces by 25%
• Lead processor power increases every generation
• Process scaling provides higher performance at lower power
Voltage Scaling Is Slowing Down

~0.7X Scaling

~0.9X Scaling
Power Density Trend

- Assumptions: 15mm die, 1.5x frequency increase per generation
Active and Leakage Power Trends

![Graph showing Active and Leakage Power Trends from 1990 to 2004. The graph plots Power [W] on a logarithmic scale against Year. The graph shows a significant increase in Leakage Power and Active Power over time, with Leakage Power growing much faster than Active Power.](image-url)
Memory and I/O Bandwidth Are Essential For High Performance
Agenda

- Microprocessor Design Trends
- Process Technology Directions
- Active Power Management
- Leakage Reduction Techniques
- Thermal Modeling
- Call to Action
- Summary
Transistor Physical Gate Length

New Process Generation Every 2 Years

Source: Robert Chau
Planar CMOS Transistor Scaling

- Intel R&D groups are exploring aggressive scaling of conventional planar CMOS transistors

Source: Robert Chau
Depleted Substrate Transistor

Single-gate DST

Silicide

$T_{Si} \approx 18 \text{ nm}$

$L_g = 60 \text{ nm}$

Epi Raised S-D

Tri-gate DST

Source: Robert Chau
Extreme Ultraviolet Lithography

- EUV lithography uses extremely short wavelength light (20x shorter than today’s lithography processes)
  - Visible light – 400 to 700 nm
  - DUV lithography – 193 and 248 nm
  - EUV lithography – 13 nm
Process Fluctuations

Die-to-Die Fluctuations
- Resist Thickness

Within-Die Fluctuations
- Systematic
- Lens Aberrations
- Random
- Random Placement of Dopant Atoms

Source: K. Bowman, et.al., ISSCC’2001

©2004 Intel Corp.
P, V, T Variations

- **Process**
  - Die-to-die variation
  - Within-die variation
  - Static for each die

- **Voltage**
  - Chip activity change
  - Current delivery—RLC
  - Dynamic: ns to 10-100us
  - Within-die variation

- **Temperature**
  - Activity & ambient change
  - Dynamic: 100-1000us
  - Within-die variation
Impact on Design Methodology

Due to variations in: Vdd, Vt, and Temp

Major paradigm shift from deterministic design to probabilistic / statistical design
Metal Layers

Number of Metal Layers

Technology Generation (um)
90nm Generation Interconnects

Copper Interconnects

Low-k CDO Dielectric

Source: M. Bohr
On-chip Interconnect Trend

- Local interconnects scale with gate delay
- Intermediate interconnects benefit from low-k material
- Global interconnects do not scale
Skin Effect

- Edge frequency is 5-9x the clock frequency
Capacitive vs. Inductive Coupling

- **Capacitive Coupling**
  - Due to electric field
  - “Near” field effect
  - Measures resistance to a voltage change

- **Inductive Coupling**
  - Due to magnetic field
  - “Far” field effect
  - Measures resistance to a current change
  - Frequency dependent
Inductive Noise

- Inductance of VLSI metal lines is becoming important at operating frequencies above 1GHz
- Need accurate R,L,C extraction tools
Inductance Effect

Xanthopoulos, ISSCC-2001
Noise Analysis

• Each circuit broken into equivalent models for various noise sources
  – Models charge-sharing, coupling, leakage, supply noise, contention
  – Calculated noise propagated to next stage to model amplification

• Macro-block results rolled up to fullchip analysis
  – Abstracted noise from block level are collected for the full chip and combined with coupling analysis
Agenda

- Microprocessor Design Trends
- Process Technology Directions
- Active Power Management
- Leakage Reduction Techniques
- Thermal Modeling
- Call to Action
- Summary
Itanium® 2 Processor Power Charts

- Maintain the same 130W power envelope
  - 50% frequency increase
  - 2X larger L3 cache
  - Leakage increased 3.5X

- Aggressive management of dynamic power
  - Reduced clock loading
  - Reduced contention power
  - L3 cache power management

130nm Itanium® 2 Processor 6M 1.5GHz
- Dynamic power: 74%
- I/O power: 14%
- Core leakage/static: 5%
- Cache leakage: 5%

180nm Itanium® 2 Processor 3M 1.0GHz
- Dynamic power: 90%
- I/O power: 5%
- All leakage/static: 5%
Active Power Reduction

Reduce switched capacitance:
• Minimize loading from diffusion, wire, gate
• Use more efficient layout techniques

Technology scaling:
• Dynamic voltage scaling
• Supply voltage scaling is slowing down
• Thresholds don’t scale

\[ P = \alpha C_L V^2 f_{CLK} \]

Reduce switching activity:
• Conditional execution
• Conditional clocking
• Conditional precharge
• Turn off inactive blocks

Reduce clock frequency:
• Use parallelism
• Less pipeline stages
• Use double-edge flip-flops
Clock Gating

- Save power by gating the clock when data activity is low
- Requires detailed logic validation
Active Power Management

- Voltage-frequency scaling with active thermal feedback
- Multi-operating states from high performance to deep sleep
- Power management reduces average and peak power
Exploit Memory Power Efficiency

- Static memory has 10X lower active power density
- Lower leakage than logic
- On-die cache provides higher bandwidth and lower latency
SRAM cell size continues to scale ~0.5x per generation

• Larger caches can be incorporated on die
Server Processors On-Die Cache Size Trends

- Increasing cache size is a power efficient way to improve server performance
Agenda

- Microprocessor Design Trends
- Process Technology Directions
- Active Power Management
- Leakage Reduction Techniques
- Thermal Modeling
- Call to Action
- Summary
Leakage Continues to Increase

- Design issues:
  - Dynamic circuits may fail
  - Need to guarantee burn-in functionality
Subthreshold Leakage Trend

- Intel 30nm transistor
- Intel 20nm transistor
- Intel 15nm transistor

Research data in literature (◆)
Production data in literature (●)
High-K gate dielectric will reduce gate leakage by up to 100x in the 45nm technology node.
Leakage Reduction Techniques

Body Bias

Stack Effect

Sleep Transistor

Equal Loading

Vdd

Vbp

+Ve

-Ve

Vbn

2-10X reduction

2-1000X reduction

©2004 Intel Corp.
Leakage is a Strong Function of Voltage

Normalized Leakage

Voltage (V)

Subthreshold Leakage

Gate Leakage

- Subthreshold and gate leakage reduce with lower supply voltage

130nm process
Standby Leakage Reduction: Sleep Transistor Design

- Motivation: Cut off power supply in sleep-mode
- Insert sleep transistor between main supply and functional unit’s supply rails
- Latches tied to main supply rails to retain state
- EDA tools needed to:
  - Size sleep transistor and distribute in layout
  - Model the timing impact
Burn-in Tolerant Dynamic Circuits

- Normal Mode Keeper
- Effective Burn-in Keeper

- Burn-in Enable
- Clock
- PK₁
- Min. sized
- PKᵦ
- Pull Down N-tree

- Leakage sensitive circuits not functional at burn-in
- Larger keepers increase delay at nominal condition
- Conditional keeper enables functional burn-in

A. Alvandpour et al, 2002 CICC
Agenda

• Microprocessor Design Trends
• Process Technology Directions
• Active Power Management
• Leakage Reduction Techniques
• Thermal Modeling
• Call to Action
• Summary
Microprocessor Package Evolution

- **1971 – 4004 Processor**
  - 16-pin ceramic package
  - Wire bond attach
  - 750kHz I/O

- **2003 – Pentium® 4 Processor**
  - 478-pin organic package
  - Flip-chip attach
  - 200MHz, quad-pumped I/O
Thermal Resistance

Package Thermal Resistance

Technology Generation

180nm 130nm 90nm

©2004 Intel Corp.
• With increasing power density and large on-die caches, detailed, non-uniform power models are required
Thermal Modeling

Simulated power density

Infrared Emission
Microscope measurement

Metal Reliability Verification

- Metal routing validated for self-heating (SH) and electromigration (EM)
  - Macro-blocks verified through every geometry for SH/EM
  - Fullchip EM correct-by-construction
  - Fullchip SH verified through all geometries

- Thermal map used at the macro-block level to tighten constraints
  - Hottest areas of die need to meet higher standards
Agenda

• Microprocessor Design Trends
• Process Technology Directions
• Active Power Management
• Leakage Reduction Techniques
• Thermal Modeling
• Call to Action
• Summary
Call to Action

• CAD tools must enable power and leakage reduction techniques with high productivity
  – All design flows must be power and leakage aware
  – Need support for multiple transistors flavors (Le and Vt) and sleep devices for leakage reduction

• CAD tools must comprehend process, temperature and voltage variations - worst casing is not practical
  – Major shift from deterministic to probabilistic design
  – Design optimization must consider parameter variations

• Simultaneous optimization of power, timing and noise
  – Need accurate R,L,C extraction tools
  – Explore multiple solutions for noise problems
Summary

• Moore’s Law will continue for at least another decade
  – 2X transistors growth per technology generation
  – 30nm and smaller transistors realized

• Power and leakage are a significant challenge
  – Exploit memory power efficiency → larger caches
  – Dynamic voltage and frequency adjustment
  – Circuit techniques (clock gating, sleep transistors)

• EDA industry’s job is to enable designers to keep pace with Moore’s law
  – Deliver tools and methodologies for increasingly complex designs
  – Focus on leakage and active power reduction