
Spring 2017

John Wawrzynek
with
James Martin (GSI)

Thanks to John Lazaro for the slides

The Watt: Unit of power. A rate of energy (J/s). A gas pump hose delivers 6 MW.

120 KW: The power delivered by a Tesla Supercharger. Tesla Model S has a 306 MJ battery (good for 265 miles).

1 J = 1 W * s 1 W = 1 J/s.
Sad fact: Computers turn electrical energy into heat. Computation is a byproduct.

Energy and Performance

Air or water carries heat away, or chip melts.
The Joule: Unit of energy. Can also be expressed as Watt-Seconds. Burning 1 Watt for 100 seconds uses 100 Watt-Seconds of energy.

1 Joule heats 1 gram of water 0.24 degree C

This is how electric tea pots work...

1 Joule of Heat Energy per Second

The Watt: Unit of power. The amount of energy burned in the resistor in 1 second.

1 Watt

1 Ohm Resistor

20 W rating: Maximum power the package is able to transfer to the air. Exceed rating and resistor burns.

The Joule: Unit of energy.

Can also be expressed as Watt-Seconds. Burning 1 Watt for 100 seconds uses 100 Watt-Seconds of energy.
Cooling an iPod nano ...

Like resistor on last slide, iPod relies on passive transfer of heat from case to the air.

Why? Users don’t want fans in their pocket ...

To stay “cool to the touch” via passive cooling, power budget of 5 W.

If iPod nano used 5W all the time, its battery would last 15 minutes ...
Powering an iPod nano (2005 edition)

1.2 W-hour battery: Can supply 1.2 watts of power for 1 hour.

1.2 W-hr / 5 W ≈ 15 minutes.

More W-hours require bigger battery and thus bigger “form factor” — it wouldn’t be “nano” anymore :-).

Real specs for iPod nano:

- 14 hours for music,
- 4 hours for slide shows.

- 85 mW for music.
- 300 mW for slides.
Finding the (2005) iPod nano CPU ...

Two 80 MHz CPUs. One CPU used for audio, one for slides.

Low-power ARM roughly 1 mW per MHz ... variable clock, sleep modes.

85 mW system power realistic ...
What's happened since 2005?

2010 nano
0.74 ounces
(50% of 2005 Nano)

"Up to" 24 hours audio playback.
70% improvement from 2005 nano.

0.39 W Hr
(33% of 2005 Nano)
3.8 V, 0.78 Wh lithium-ion battery on 38mm model. Apple claims the 205 mAh battery should provide up to 18 hours of use (which translates to 6.5 hours of audio playback, 3 hours of talk time, or 72 hours of Power Reserve mode.)
A clever prism projects a layer over reality light.
2.1 Wh battery – 5.3x as much energy as 2010 Nano.

Battery life very usage dependent.

1.76 ounces – 4X the weight of iPod Shuffle

640 x 360 Liquid Crystal on Silicon (LCoS) prism projector.
4.7 inch iPhone 6: 1,810mAh battery

iPhone 5s: 1570mAh
The A8 is manufactured on a 20 nm process by TSMC. It contains 2 billion transistors. Its physical size is 89 mm$^2$. It has 1 GB of LPDDR3 RAM included in the package. It is dual core, and has a frequency of 1.38 GHz.
Notebooks ... as designed in 2006 ...

Performance: Must be “close enough” to desktop performance ... most people no longer used a desktop (even in 2006).

Size and Weight. Ideal: paper notebook.

Heat: No longer “laptops” -- top may get “warm”, bottom “hot”. Quiet fans OK.
Battery: Set by size and weight limits ...

Almost full 1 inch depth. Width and height set by available space, weight.

Battery rating: 55 W-hour.

At 2.3 GHz, Intel Core Duo CPU consumes 31 W running a heavy load - under 2 hours battery life! And, just for CPU!

46x more energy than iPod nano battery. And iPod lets you listen to music for 14 hours!

At 1 GHz, CPU consumes 13 Watts. "Energy saver" option uses this mode ...

CS 250 L6: Power and Energy

UC Regents Spring 2017 © UCB
MacBook Air ... design the laptop like an iPod
Non-removable, “form-fit” battery... 

Mainboard: fills about 25% of the laptop

35 W-h battery: 63% of 2006 MacBook’s 55 W-h
Servers: Total Cost of Ownership (TCO)

Machine rooms are expensive. Removing heat dictates how many servers to put in a machine room.

Electric bill adds up! Powering the servers + powering the air conditioners is a big part of TCO.

Reliability: running computers hot makes them fail more often.
Computations per W-h doubles every 1.6 years, going back to the first computer.

(Jonathan Koomey, Stanford).
Processors and Energy
Moore’s Law

curve shows transistor count doubling every two years
Switching Energy: Fundamental Physics

Every logic transition dissipates energy.

How can we limit switching energy?

1. Reduce # of clock transitions. But we have work to do ...
2. Reduce Vdd. But lowering Vdd limits the clock speed ...
3. Fewer circuits. But more transistors can do more work.
4. Reduce C per node. One reason why we scale processes.

Gate Switching Behavior

- Inverter:
  - For a gate: $E_{0\to1} = \frac{1}{2} C V_{dd}^2$
  - Strong result: Independent of technology.
Second Factor: Leakage Currents

Even when a logic gate isn’t switching, it burns power.

\[ \text{0V} = V_{\text{IN}} \]

\[ V_{\text{OUT}} \]

\[ I_{\text{Gate}} \]

\[ I_{\text{Sub}} \]

\[ C_L \]

\[ I_{\text{Sub}}: \text{Even when this nFet is off, it passes an } I_{\text{off}} \text{ leakage current.} \]

We can engineer any \( I_{\text{off}} \) we like, but a lower \( I_{\text{off}} \) also results in a lower \( I_{\text{on}} \), and thus a lower maximum clock speed.

\[ \text{IGate: Ideal capacitors have zero DC current. But modern transistor gates are a few atoms thick, and are not ideal.} \]

Intel’s 2006 processor designs, leakage vs switching power

A lot of work was done to get a ratio this good ... 50/50 is common.

Bill Holt, Intel, Hot Chips 17
Engineering “On” Current at 25 nm ...

We can increase $I_{on}$ by raising $V_{dd}$ and/or lowering $V_t$.

$0.25 = V_t$

$0.7 = V_{dd}$

$I_{ds} = 1.2 \text{ mA} = I_{on}$

$I_{off} = 0$ ???
Plot on a “Log” Scale to See “Off” Current

We can decrease $I_{off}$ by raising $V_t$ - but that lowers $I_{on}$.

$I_{off} \approx 10 \text{nA}$

$0.25 \approx V_t$

$0.7 = V_{dd}$

$1.2 \text{mA} = I_{on}$
Device engineers trade speed and power

We can reduce $CV^2(P_{\text{active}})$ by lowering $V_{dd}$.

We can increase speed by raising $V_{dd}$ and lowering $V_t$.

We can reduce leakage ($P_{\text{standby}}$) by raising $V_t$.

From: Silicon Device Scaling to the Sub–10–nm Regime
Meikei Leong,¹ Bruce Doris,² Jakub Kedzierski,¹ Ken Rim,¹ Min Yang¹
Customize processes for product types ...

Optimized transistors can provide ~1000x lower leakage.

Transistors Require Optimization to the Application

Performance vs. Leakage

Transistor physics revisited ...

The drain junction is also a capacitor, and puts - charges in the substrate.

Away from the surface, the drain-induced charges remain even when the gate is off!

As we make $L$ smaller, source and drain come closer, and $I_{off}$ gets larger!
Solution concept: Fully-depleted channel

We limit the depth of the channel so that the gate voltage “wins” over the drain voltage.

Done as shown, 5 to 7 nm depth for a 20 nm transistor.

Requires expensive wafers

“FD-SOI” -- Fully-Depleted Silicon-On-Insulator
Transistor channel is a raised fin. Gate controls channel from sides and top. Channel depth is fin width. 12-15nm for L=22nm.
A series of more elaborate manufacturing steps such as double patterning. The 20-nm node provides phenomenal benefits in terms of power as well as performance, but the cost is increasing marginally because of elaborate manufacturing to ensure silicon integrity.

Thanks in large part to remarkable research started by Cal Berkeley professor Chenming Hu under a DARPA contract, the 20-nm process will likely be the last hurrah for the planar transistor (at least as we know it today), as the industry moves to FETs built with fins.

INS AND OUTS OF FINS

In a planar transistor of today, electrical current flows from source to drain through a flat, 2D horizontal channel underneath the gate. The gate voltage controls current flow through the channel. As transistor size shrinks with the introduction of each new silicon process, the planar transistor cannot adequately stop the flow of current when it is in an “off” state, which results in leakage and heat. In a FinFET MOSFET transistor, the gate wraps around the channel on three sides, giving the gate much better electrostatic control to stop the current when the transistor is in the “off” state. Superior gate control in turn allows designers to increase the current and switching speed and, thus, the performance of the IC.

Because the gate wraps around three sides of the fin-shaped channel, the FinFET is often called a 3D transistor (not to be confused with 3D ICs, like the Virtex-7 2000T, which Xilinx pioneered with its stacked-silicon technology).

In a three-dimensional transistor (see Figure 1b), gate control of the channel is on three sides rather than just one, as in conventional two-dimensional planar transistors (see Figure 1a). Even better channel control can be achieved with a thinner fin, or in the future with a gate-all-around structure where the channel will be enclosed by a gate on all sides.

The industry believes the 16-nm/14-nm FinFET process will enable a 50 percent performance increase at the same power as a device built at 28 nm. Alternatively, the FinFET device will consume 50 percent less power at the same performance. The performance-per-watt benefits added to the continued increases in capacity make FinFET processes extremely promising for devices at 16 or 14 nm and beyond.

That said, the cost and complexity of designing and manufacturing 3D transistors is going to be higher at least for the short term, as EDA companies figure out ways to adequately model the device characteristics of these new processes and augment their tools and flows to account for signal integrity, electromigration, width quantization, resistance and capacitance. This complexity makes designing ASICs and ASSPs even riskier and more expensive than before.

Xilinx, however, shields users from the manufacturing details. Customers can reap the benefits of increased performance per watt and Xilinx’s Generation Ahead design flows to bring innovations based on the new UltraScale architecture to market faster.

Figure 1 – The position of the gate differs in the two-dimensional traditional planar transistor (a) vs. the three-dimensional FinFET transistor (b).
Sandy Bridge
32nm planar
1.16B transistors

“Less than half the power @ same performance”

Ivy Bridge
22nm FinFet
1.4B transistors
Leakage reduction in “Tock” 22nm Intel CPUs

Haswell (2013, 22nm FinFET)

Ivy Bridge (2012, 22nm FinFET)

—Cary Chin is director of marketing for low-power solutions at Synopsys.
Break
Six low-power design techniques

- Parallelism and pipelining
- Power-down idle transistors
- Slow down non-critical paths
- Clock gating
- Data-dependent processing
- Thermal management
Trading Hardware for Power

via Parallelism and Pipelining ...
Gate delay roughly linear with Vdd

And so, we can transform this:

Logic Block

Freq = 1
Vdd = 1
Throughput = 1
Power = 1
Area = 1
Pwr Den = 1

Power Reduction

Slow

Fast

Low Supply Voltage

High Supply Voltage

Multiple Supply Voltages

Logic Block

Freq = 0.5
Vdd = 0.5
Throughput = 1
Power = 0.25
Area = 2
Pwr Den = 0.125

Replicated Designs

And so, we can transform this:

Block processes stereo audio. 1/2 of clocks for “left”, 1/2 for “right”.

Into this:

Top block processes “left”, bottom “right”.

Active Power Reduction

CV^2 power only

P ~ #blks x F x Vdd^2
P ~ 2 x 1/2 x 1/4 = 1/4

THIS MAGIC TRICK BROUGHT TO YOU BY CORY HALL...
### Architecture vs. Power (normalized)

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Power (normalized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>1</td>
</tr>
<tr>
<td>Parallel</td>
<td>0.36</td>
</tr>
<tr>
<td>Pipelined</td>
<td>0.39</td>
</tr>
<tr>
<td>Pipelined-Parallel</td>
<td>0.2</td>
</tr>
</tbody>
</table>

### Architecture vs. Area (normalized)

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Area (normalized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>1</td>
</tr>
<tr>
<td>Parallel</td>
<td>3.4</td>
</tr>
<tr>
<td>Pipelined</td>
<td>1.3</td>
</tr>
<tr>
<td>Pipelined-Parallel</td>
<td>3.7</td>
</tr>
</tbody>
</table>

### Architecture vs. Voltage

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Voltage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>5V</td>
</tr>
<tr>
<td>Parallel</td>
<td>2.9V</td>
</tr>
<tr>
<td>Pipelined</td>
<td>2.9V</td>
</tr>
<tr>
<td>Pipelined-Parallel</td>
<td>2.0V</td>
</tr>
</tbody>
</table>

From: Minimizing Power Consumption in CMOS Circuits

Anantha P. Chandrakasan
Robert W. Brodersen
Regents Spring 2017 © UCB

---

Pipelined-PoPpThroughput

Throughput

Figure 1.15

Area = 636 x 833 µ²

Parallel-PoP

Parallel

Pipelined-PoP

Pipelined

From: Minimizing Power Consumption in CMOS Circuits

Anantha P. Chandrakasan
Robert W. Brodersen
Regents Spring 2017 © UCB
Example: Intel Graphics Pipeline IP

A 2.05 GVertices/s 151 mW Lighting Accelerator for 3D Graphics Vertex and Pixel Shading in 32 nm CMOS

Farhana Sheikh, Member, IEEE, Sanu K. Mathew, Member, IEEE, Mark A. Anders, Member, IEEE, Himanshu Kaul, Member, IEEE, Steven K. Hsu, Member, IEEE, Amit Agarwal, Member, IEEE, Ram K. Krishnamurthy, Fellow, IEEE, and Shekhar Borkar, Fellow, IEEE
The authors are with the Circuit Research Lab, Intel Labs, Intel Corporation, Hillsboro, OR 97124, USA (e-mail: farhana.sheikh@intel.com; amit1.agarwal@intel.com; ram.krishnamurthy@intel.com). This paper was approved by Guest Editor Timothy Fischer.

IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 48, NO. 1, JANUARY 2013

Maximal clock frequency $F_{\text{max}}$ and throughput $\text{V} \text{Vertices/s}$ as a function of supply voltage $V_{\text{supply}}$ for 32nm CMOS, 25°C. Graph shows a 190MV vertices/s at 560mV and a 205G vertices/s at 1.05V. Farhana Sheikh, Member, IEEE, Sanu K. Mathew, Member, IEEE, Mark A. Anders, Member, IEEE, Himanshu Kaul, Member, IEEE, Steven K. Hsu, Member, IEEE, Amit Agarwal, Member, IEEE, Ram K. Krishnamurthy, Fellow, IEEE, and Shekhar Borkar, Fellow, IEEE.
Multiple Cores for Low Power

Trade hardware for power, on a large scale ...
Cell: The PS3 chip
Cell (PS3 Chip): 1 CPU + 8 "SPUs"

- L2 Cache: 512 KB
- 8 Synergistic Processing Units (SPUs)
- PowerPC
A “Schmoo” plot for a Cell SPU ...

The lower Vdd, the less dynamic energy consumption.

\[ E_{0\rightarrow1} = \frac{1}{2} CV_{dd}^2 \quad E_{1\rightarrow0} = \frac{1}{2} CV_{dd}^2 \]

The lower Vdd, the longer the maximum clock period, the slower the clock frequency.

Failed
Clock speed alone doesn’t help $E/$op ... 

But, lowering clock frequency while keeping voltage constant spreads the same amount of work over a longer time, so chip stays cooler ...

$$E_{0 \rightarrow 1} = \frac{1}{2} CV_{dd}^2 \quad E_{1 \rightarrow 0} = \frac{1}{2} CV_{dd}^2$$
Scaling V and f does lower energy/op

1 W to get 2.2 GHz performance. 26 C die temp.

7 W to reliably get 4.4 GHz performance. 47 C die temp.

If a program that needs a 4.4 Ghz CPU can be recoded to use two 2.2 Ghz CPUs ... big win.

![Diagram showing scaling of V and f for energy and performance]
How iPod nano 2005 puts its 2 cores to use ...

Two 80 MHz CPUs. Was used in several nano generations, with one CPU doing audio decoding, the other doing photos, etc.
Powering down idle circuits
Add “sleep” transistors to logic ...

Example: Floating point unit logic.

When running fixed-point instructions, put logic “to sleep”.

+++ When “asleep”, leakage power is dramatically reduced.

--- Presence of sleep transistors slows down the clock rate when the logic block is in use.
Intel example: Sleeping cache blocks

>3x SRAM leakage reduction on inactive blocks

A tiny current supplied in “sleep” maintains SRAM state.

The 32nm CPU in the Intel® Atom™ Processor Z2480 is a process-shrink of the original 45nm Atom™ micro-architecture [5]. The 32nm CPU doubled the size of the Gshare branch predictor to 8K entries and optimized memory copy performance. Additional low-power enhancements include operating the CPU at lower minimum voltage, reducing active power of the CPU PLL, separating voltage rails for CPU and the L2 cache, and enabling full power-gating of the CPU in the C6 standby mode. This section highlights the results of these low-power optimizations, and how they apply to smartphone use cases.

As shown in Figure 5, the Intel® Atom™ CPU provides a wide dynamic performance/power operating range. On Medfield, fine-grained active CPU power management is accomplished through dynamic frequency voltage scaling that is controlled by Enhanced Intel® SpeedStep® Technology [7], also known as CPU P-states. The dynamic range of the CPU ranges from 600MHz @ ~175mW to a sustained high frequency mode (HFM) of 1.3GHz @ ~500mW. For bursty workloads (e.g., interactive use of a web browser) the CPU can burst up to 2.0GHz @ ~1,200mW for short periods of time.

As long as thermal headroom allows, the CPU can run in burst mode until thermal monitors in the platform, the SoC, or the CPU indicate that thermal design power limits have been reached. When this occurs, a combination of firmware and software throttle the CPU back into a lower P-state. Additionally, CPU w/512KB L2$ 

LPDDR2 
eMMC 
SD/MMC 
Primary Camera: 8MP, 15fps, 1080p 
Secondary Camera: 1.3MP, 1080p 

Power Delivery IC: VRs Audio CODEC 
IMC 6260 HSPA+ Modem 
TI WiFi & BT 
CSR GPS 
SDIO UART 
HDMI 1.3a 
MIPI-HSI MIPI-CSI MIPI-DSI MIPI-HSI SDIO UART HDMI 1.3a CSR GPS SDIO UART HDMI 1.3a 

Digital Object Indentifier 10.1109/MM.2013.22             0272-1732/$26.00 2013 IEEE
Intel Medfield

Switches 45 power “islands.”

Fine-grained control of leakage power, to track user activity.

“Race to idle” strategy -- finish tasks quickly, to get to power down.
Playing a game ...

Active system looks like this.
Watching a video ...

Figure 6 – S0 System State (CPU and Graphics Active)

Figure 7 – S0 System State (CPU Off, Media blocks still on)

Active system looks like this

CPU w/512KB L2$

IO

IO

IO

Security Engine

Power Manager

Low Power Audio

Storage

CPU w/512KB L2$

IO

IO

IO

2D/3D Graphics

Video Encode/Decode (1080p30)

Image Signal Processor

Display Controller (3 pipes)

Low Power

Audio

Power Manager

Security Engine

CPU is now off!
Looking at phone screen, not doing anything ...

Figure 8 – S0i1 System State with power consumption in the mW range

Figure 9 – S0i3 System State with power consumption in the uW range

S0i1 – low activity

CPU w/512KB L2$

IO

Security Engine

Power Manager

Low Power Audio

Storage

CPU w/512KB L2$

IO

2D/3D Graphics

Video Encode/Decode (1080p30)

Image Signal Processor

Display Controller (3 pipes)

Low Power

Audio

Power Manager

Security Engine

2D/3D Graphics

Video Encode/Decode (1080p30)

Image Signal Processor

Display Controller (3 pipes)

Digital Object Indentifier 10.1109/MM.2013.22             0272-1732/$26.00 2013 IEEE

This article has been accepted for publication in IEEE Micro but has not yet been fully edited. Some content may change prior to final publication.
Phone in your pocket, waiting for a call ...

Standby State – just waiting for wakes
Slow down “slack paths”
Fact: Most logic on a chip is “too fast”

The critical path

Most wires have hundreds of picoseconds to spare.

Use several supply voltages on a chip ...

Why use multi-Vdd? We can reduce dynamic power by using low-power Vdd for logic off the critical path.

What if we can’t do a multi-Vdd design?
In a multi-Vt process, we can reduce leakage power on the slow logic by using high-Vth transistors.

Logical partition into 0.8V and 1.0V nets done manually to meet 350 MHz spec (90nm).

Level-shifter insertion and placement done automatically.

Dynamic power in 0.8V section cut 50% below baseline.

Leakage power in 1.0V section cut 70% by using high $V_T$. 

From a chapter from new book on ASIC design by Chinnery and Keutzer (UCB).
Gating clocks to save power
On a CPU, where does the power go?

Half of the power goes to latches (Flip-Flops).

Most of the time, the latches don’t change state.

So (gasp) gated clocks are a big win.
But, done with CAD tools in a disciplined way.

From: Bose, Martonosi, Brooks: Sigmetrics-2001 Tutorial
Synopsis Design Compiler can do this ...

Up to 70% power savings at the block level, for applicable circuits

Synopsis Data Sheet
Design Technique #5 (of 6)

Data-Dependent Processing
Example: Video Decode Transform

Most of the time, the inputs flip between small positive and negative integers.

In 2’s complement, wastes power:

+1: 0b00001
-1: 0b11110
Solution: Add bias value to all inputs

30+% power reduction for a bias of 64. For this linear transform, correcting the output for the bias is trivial.
Thermal Management
Keep chip cool to minimize leakage power

Figure 3: $I_{CCINTQ}$ vs. Junction Temperature with Increase Relative to 25°C

<table>
<thead>
<tr>
<th>Junction Temperature ($T_J$ °C)</th>
<th>Normalized Static Power or $I_{CCINTQ}$ Typical</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1.00</td>
</tr>
<tr>
<td>50</td>
<td>1.46</td>
</tr>
<tr>
<td>85</td>
<td>2.50</td>
</tr>
<tr>
<td>100</td>
<td>3.14</td>
</tr>
</tbody>
</table>

A recipe for thermal runaway

Optimizing Designs for Power Consumption through Changes to the FPGA Environment

CS 250 L6: Power and Energy

WP285 (v1.0) February 14, 2008 UC Regents Spring 2017 © UCB
IBM Power 4: How does die heat up?

4 dies on a multi-chip module

2 CPUs per die
115 Watts: Concentrated in “hot spots”

66.8 °C == 152 °F
82 °C == 179.6 °F
Idea: Monitor temperature, servo clock speed

Power consumption varies greatly by workload

Time (as we run a benchmark suite)

* Based on internal AMD modeling using benchmark simulations

TDP = Thermal Design Point
# Intel realtime temp monitoring

<table>
<thead>
<tr>
<th>CPUID</th>
<th>306C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threads</td>
<td>8</td>
</tr>
<tr>
<td>GPU</td>
<td>36°C</td>
</tr>
<tr>
<td>Clock Speed</td>
<td>3591.683 MHz</td>
</tr>
<tr>
<td>Load</td>
<td>12.9%</td>
</tr>
<tr>
<td>VID</td>
<td>1.0463</td>
</tr>
<tr>
<td>Power</td>
<td>15.2 W</td>
</tr>
</tbody>
</table>

- **Temperature (°C)**:
  - 45
  - 45
  - 49
  - 62
  - **Package**: 62

- **Distance to TJ Max**:
  - 55
  - 55
  - 51
  - 38
  - **Package**: 38

- **Minimum**:
  - 39°C
  - 39°C
  - 39°C
  - 39°C
  - **Sensor**: 41°C

- **Maximum**:
  - 54°C
  - 61°C
  - 62°C
  - 62°C

- **Thermal Status**:
  - OK
  - OK
  - OK
  - OK
  - OK
Six low-power design techniques

- Parallelism and pipelining
- Power-down idle transistors
- Slow down non-critical paths
- Clock gating
- Data-dependent processing
- Thermal management