

# Power Wall

Power has become the biggest challenge facing computer designers across all classes of computers, creating three primary concerns:

## Maximum Power Requirements

The peak power a processor requires must be met to ensure correct operation. Exceeding power supply capabilities can cause voltage drops and device **malfunction**.

## Thermal Design Power (TDP)

The sustained power consumption determines **cooling** requirements. Inadequate cooling can cause junction temperatures to exceed maximum values, resulting in device failure or **permanent damage**.

## Energy Efficiency

Energy consumption ( $\text{power} \times \text{time}$ ) is the most important metric for comparing processors, as it directly relates to **electricity costs** for servers and **battery life** for mobile devices.

# Energy vs. Power

## Why Energy Matters More Than Power

Energy is always a better metric because it's tied to a specific task and the time required for that task:

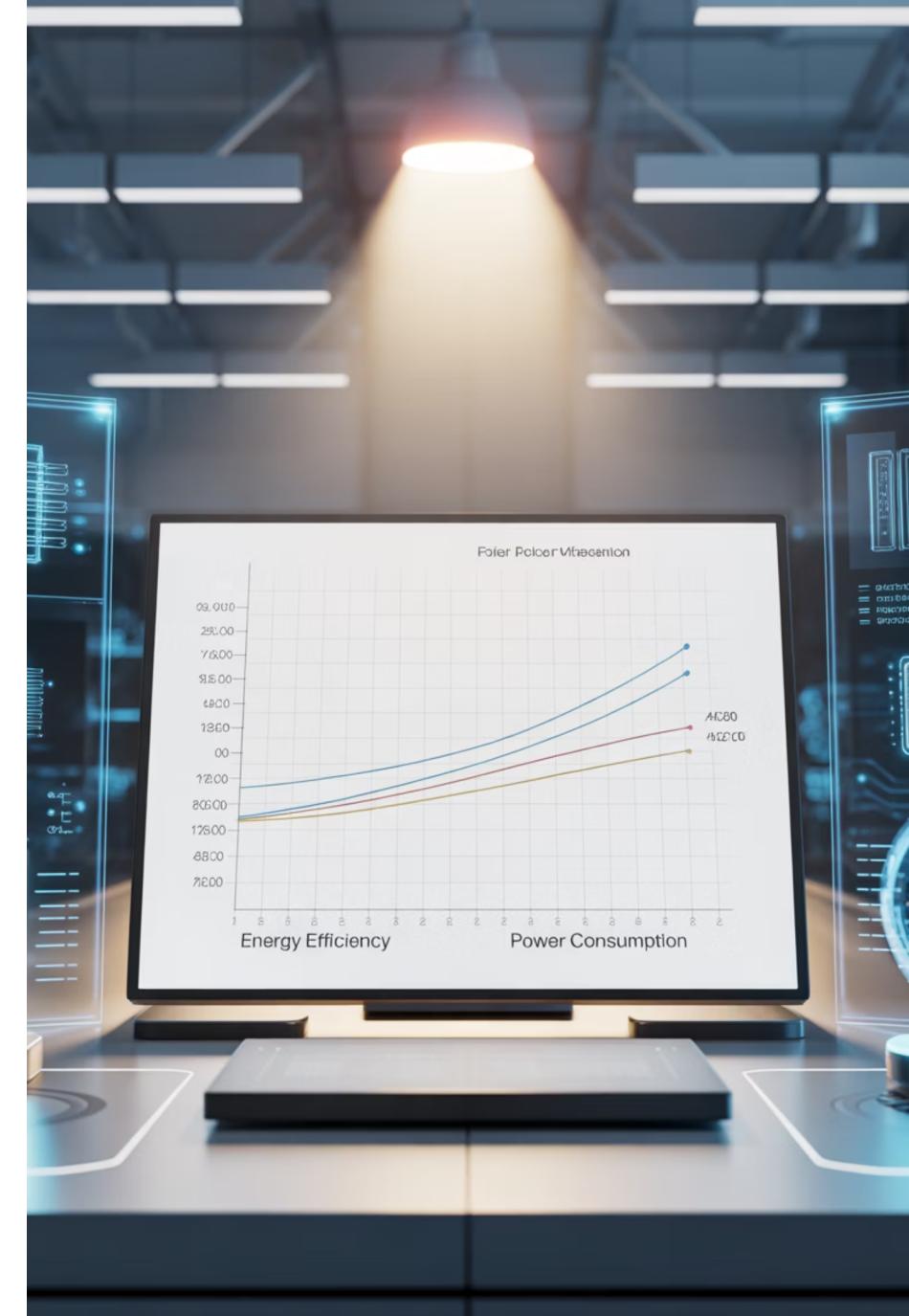
$$\text{Energy} = \text{Average Power} \times \text{Execution Time}$$

Example: Processor A has 20% higher power consumption than Processor B, but executes a task in 70% of the time. Its energy consumption is  $1.2 \times 0.7 = 0.84$ , making it more efficient.

## When to Consider Power

Power consumption is useful primarily as a constraint (e.g., a chip limited to 100 watts).

For fixed workloads—whether for warehouse-size clouds or smartphones—comparing energy is the right approach, as electricity bills and battery life are determined by energy consumed.



# Dynamic Energy and Power

**1**

## Dynamic Energy

$$Energy_{dynamic} \propto \text{Capacitive load} \times \text{Voltage}^2$$

For a single transition (0→1 or 1→0):

$$Energy_{dynamic} \propto \frac{1}{2} \times \text{Capacitive load} \times \text{Voltage}^2$$

**2**

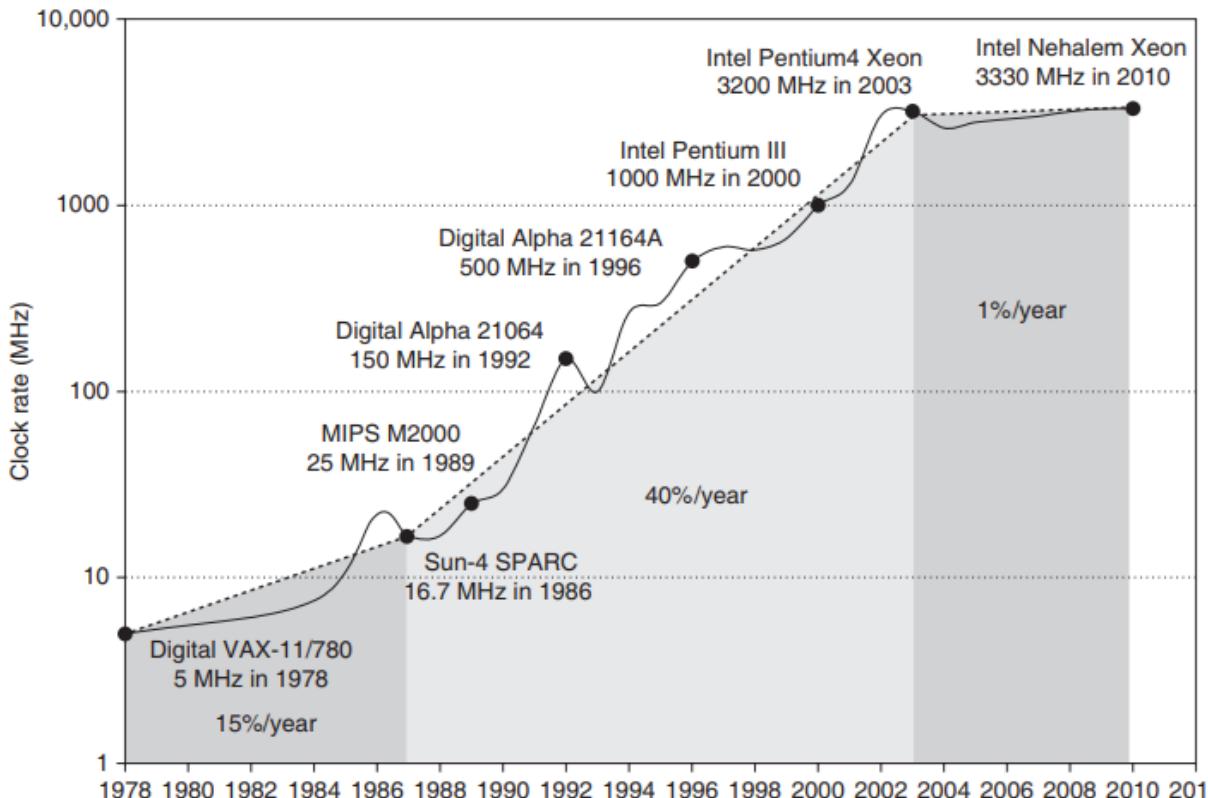
## Dynamic Power

$$Power_{dynamic} \propto \frac{1}{2} \times \text{Capacitive load} \times \text{Voltage}^2 \times \text{Frequency switched}$$

For a fixed task, slowing clock rate reduces power, but not energy.

Example: A 15% reduction in voltage may result in a 15% reduction in frequency. This reduces energy to about 72% of the original and power to about 61% of the original.

# Clock Rate Growth Has Stalled



- As the graph shows, clock frequency growth has slowed dramatically since 2003:
  - 1978-1986: Clock rate improved less than 15% per year
  - 1986-2003: "Renaissance period" with clock rates increasing almost 40% per year
  - Post-2003: Clock rate nearly flat, growing less than 1% per year
- This flattening corresponds directly to the period of slow performance improvement in modern processors, as **power constraints have become the limiting factor** rather than silicon area.

# Energy Efficiency Techniques



## Do Nothing Well

Turn off clocks of inactive modules to save energy and dynamic power. If floating-point units or cores are idle, their clocks are stopped.



## Design for Typical Case

Optimize for common usage patterns with low-power modes for memory and storage, relying on temperature sensors for emergency slowdowns.

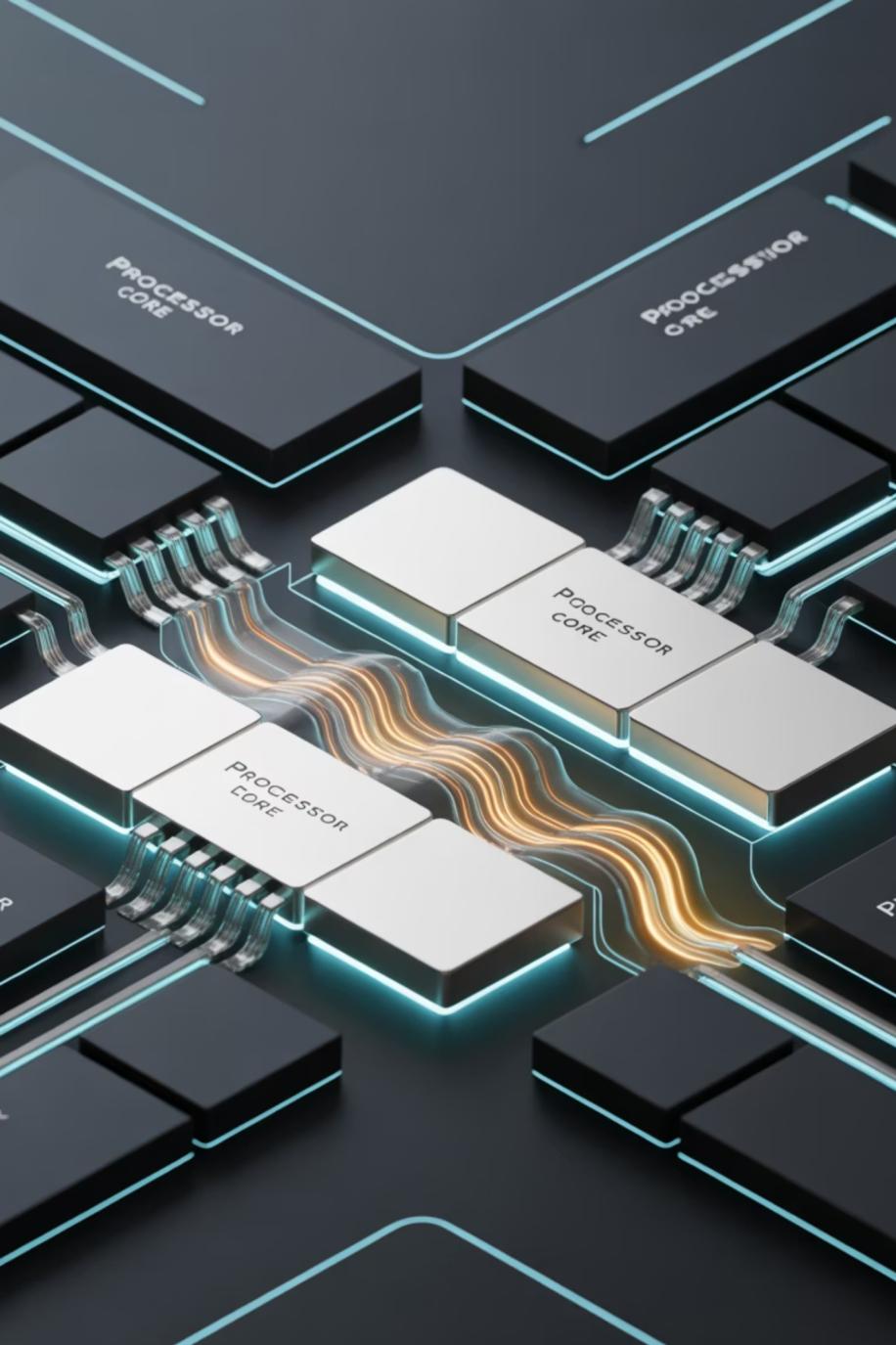
## Dynamic Voltage-Frequency Scaling (DVFS)

Offer multiple clock frequencies and voltages for periods of low activity, significantly reducing power consumption.

## Overclocking (Turbo Mode)

Run at higher clock rates for short periods until temperature rises, offering ~10% performance boost over nominal rates.

# Static Power: The Growing Challenge



## Static Power Consumption

$$\text{Power}_{\text{static}} \propto \\ \text{Current}_{\text{static}} \times \text{Voltage}$$

Static power is proportional to the number of devices, meaning more transistors increase power consumption even when idle.

Leakage current increases in processors with smaller transistor sizes, becoming a **major concern in modern designs**.

## Addressing Static Power

- Power gating: turning off power supply to inactive modules
- Goal for leakage in 2011: 25% of total power consumption
- High-performance designs often exceed 50% leakage
- Large SRAM caches contribute significantly to leakage

# Performance per Watt

## New Efficiency Metrics

The primary evaluation metric has shifted from performance per mm<sup>2</sup> of silicon to:

- Tasks per joule
- Performance per watt

## Race-to-Halt Strategy

Sometimes using a faster, less energy-efficient processor allows the rest of the system to enter sleep mode sooner, reducing overall energy consumption.

## Future Directions

Power and energy constraints are fundamentally reshaping approaches to parallelism and processor design, driving innovation in computer architecture.

The interplay between transistor scaling, performance, and power consumption creates complex design challenges that will continue to shape the future of computing.