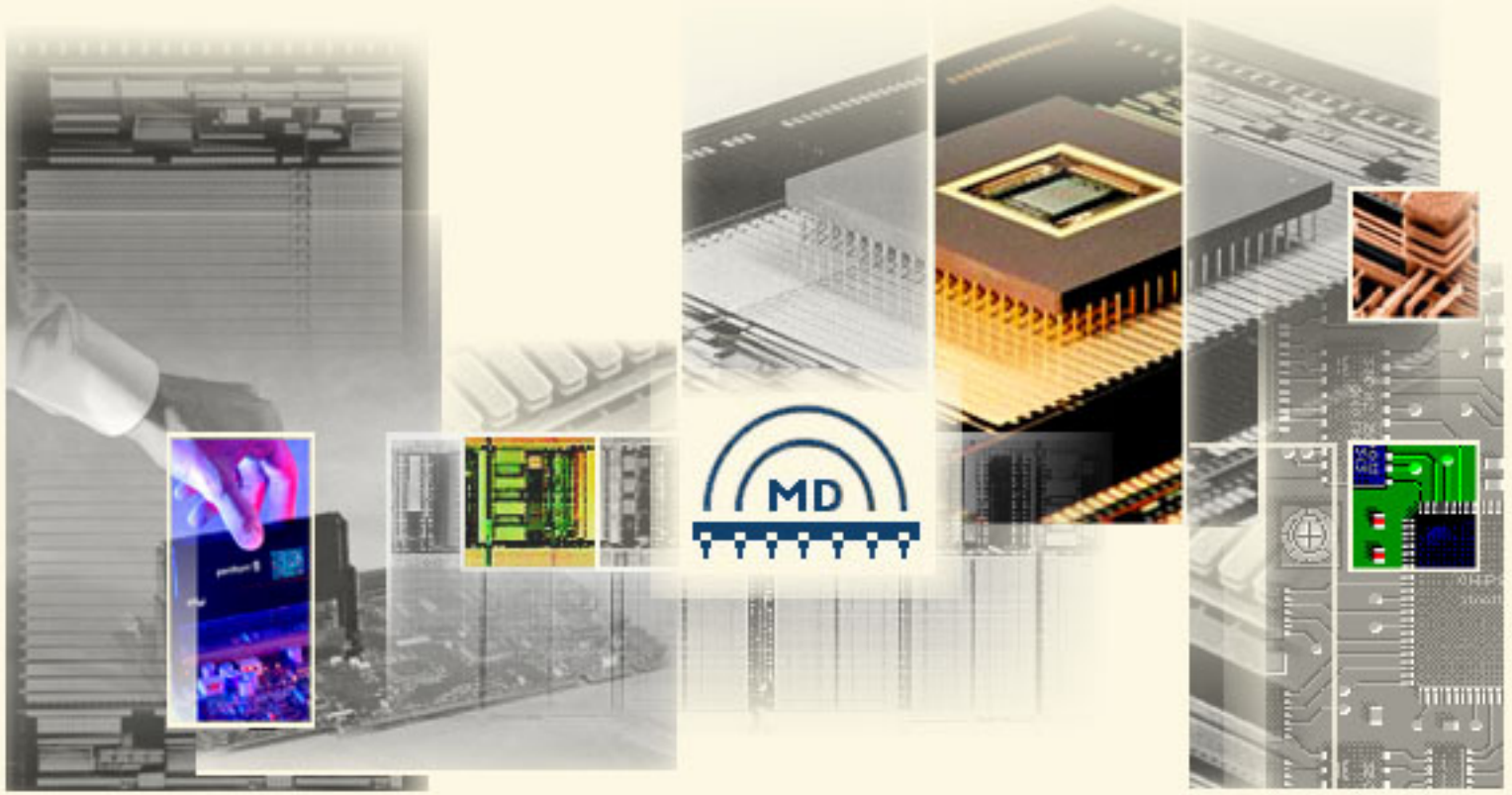
Low Power Techniques



Institut für Angewandte Mikroelektronik und Datentechnik

Low Power Techniques

Frank Sill

Institute of Applied Microelectronics and Computer Sience
Faculty for Computer Science and Electrical Engineering
University of Rostock
Tuesday, 27. January 2003

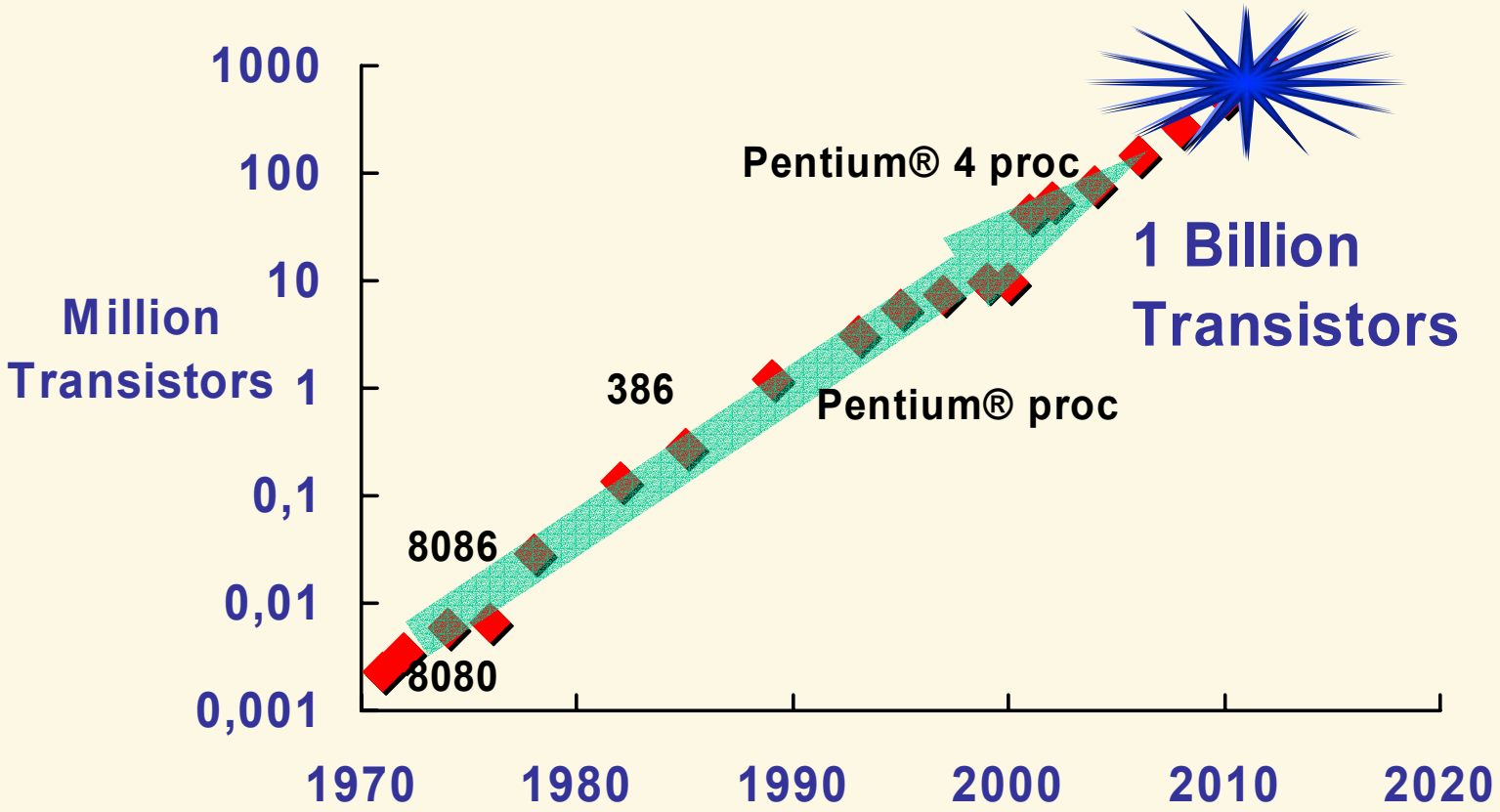
Outline

1) Power dissipation in today’s applications
2) Abstraction levels
3) Leakage currents
4) Summery

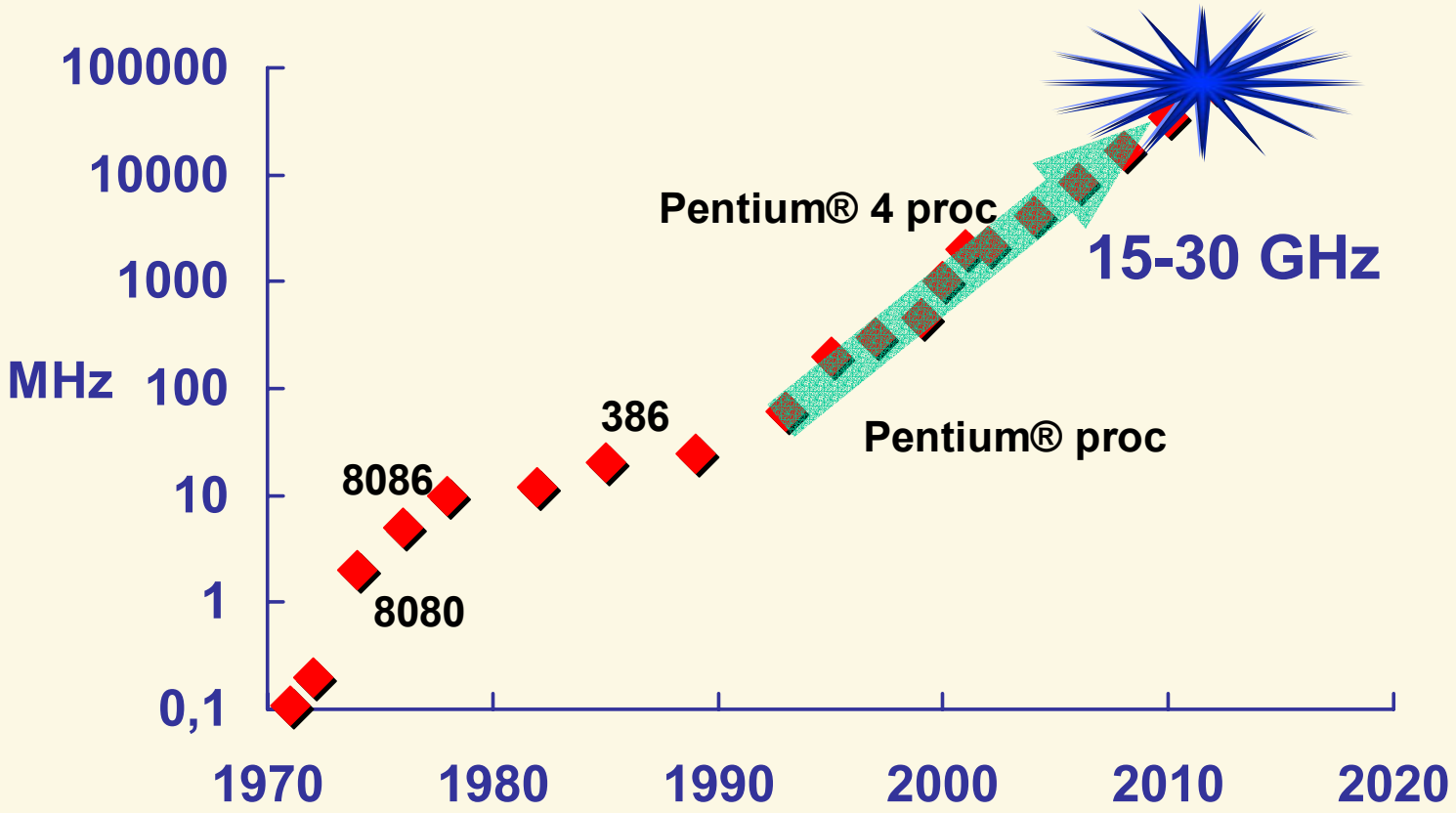
Outline

1) Power dissipation in today's applications
   - Performance trends
   - Power trends
2) Abstraction levels
3) Leakage currents
4) Summery

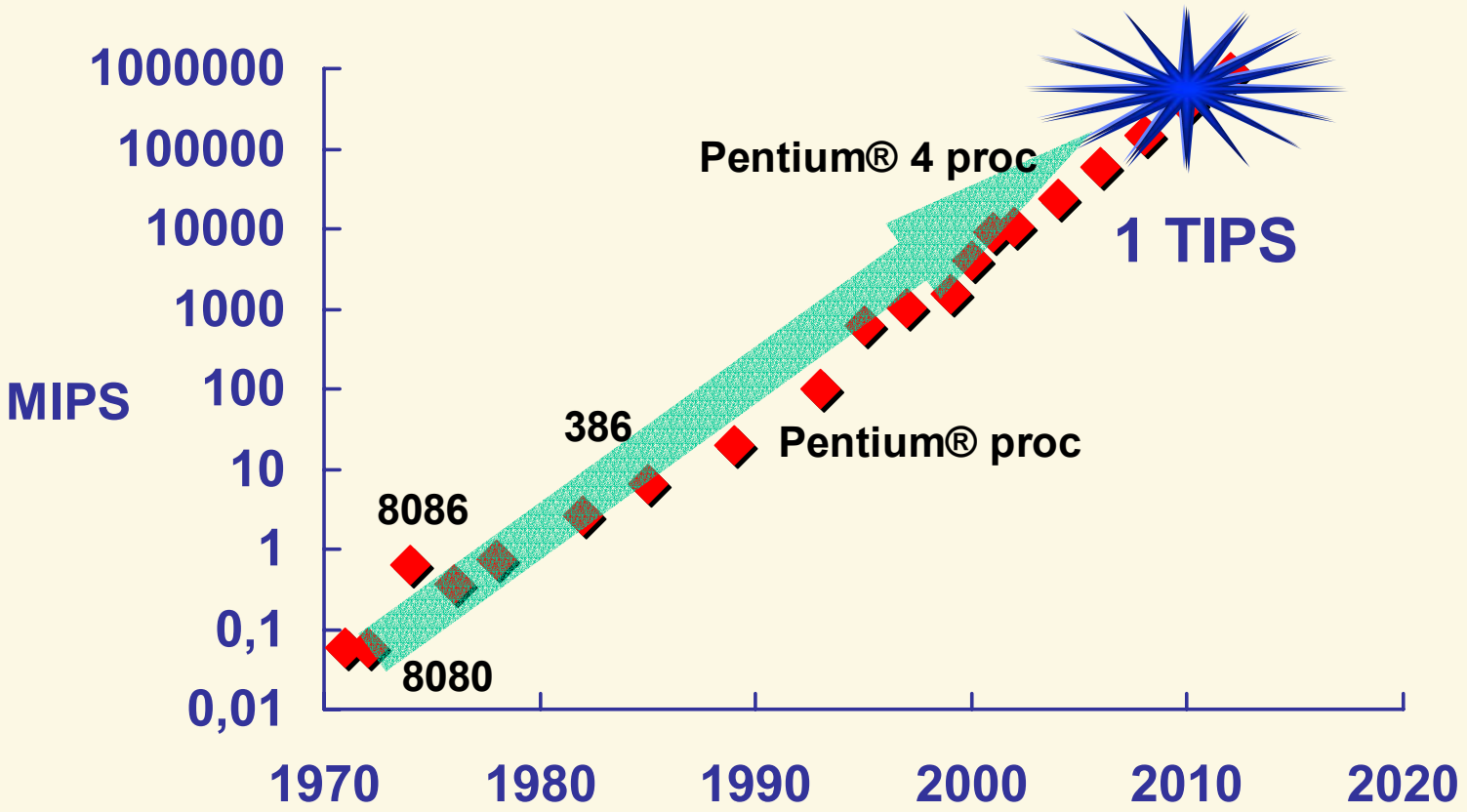
Trend: Transistor count



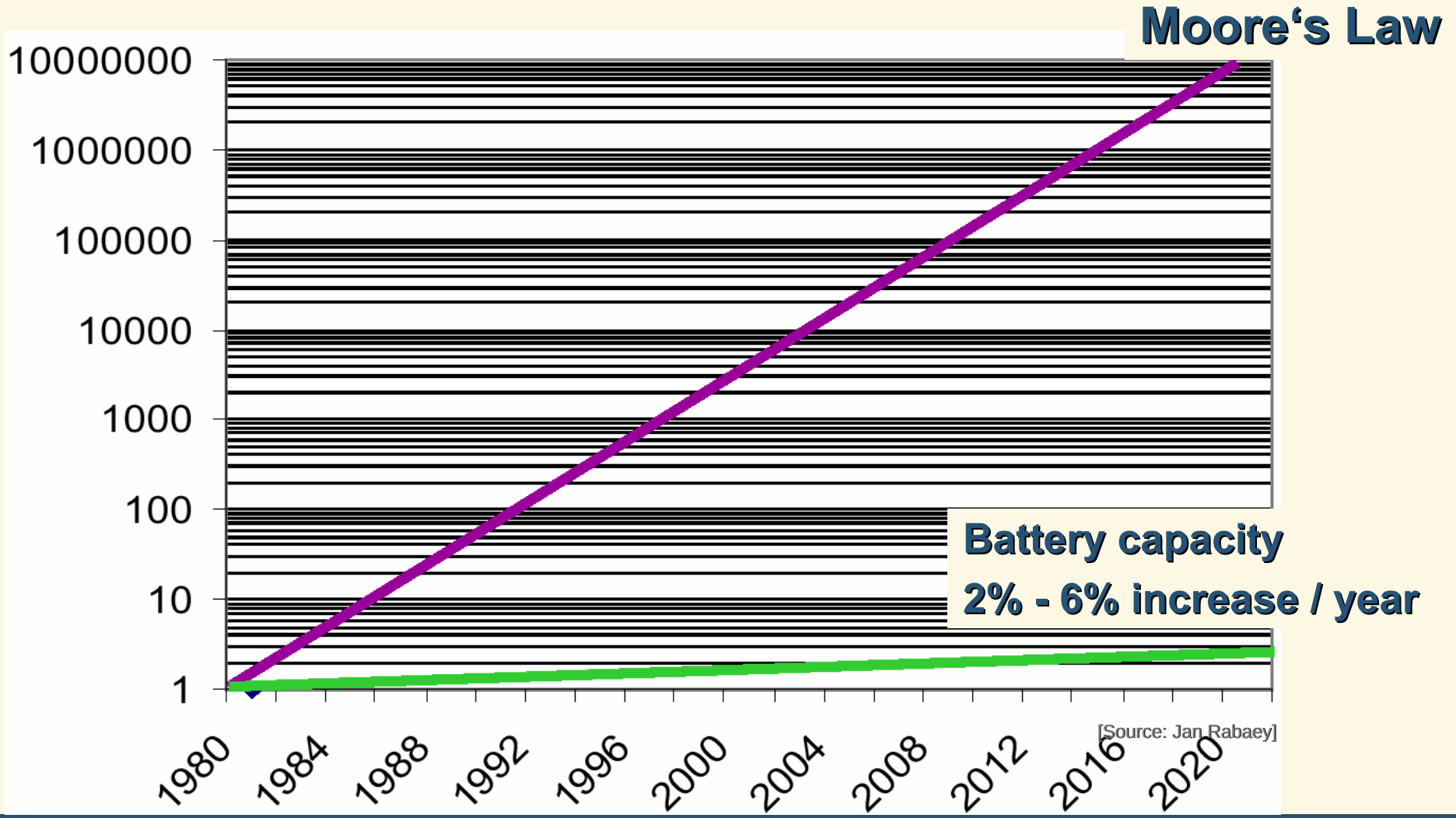
Trend: Frequency



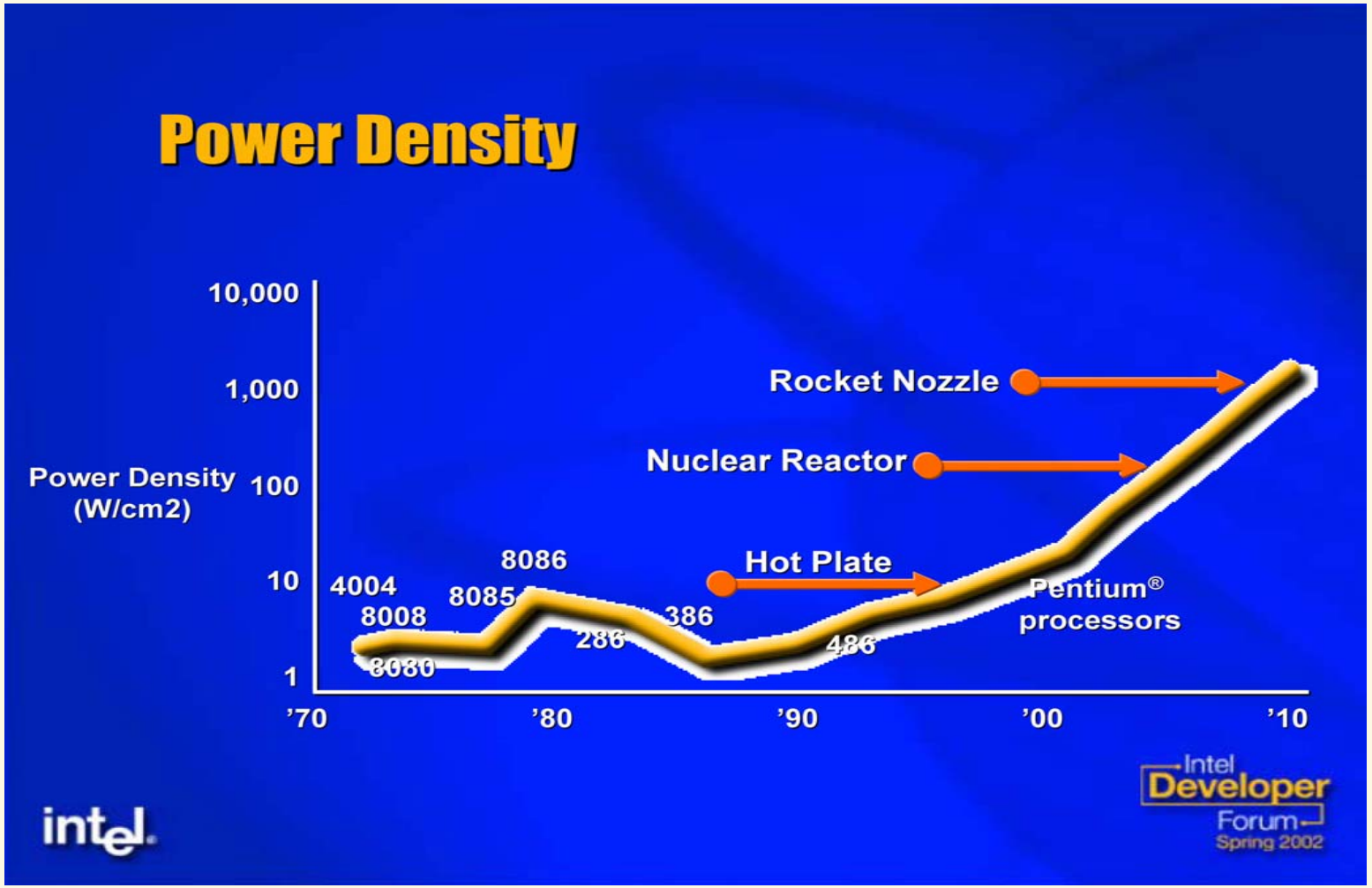
Trend: Performance



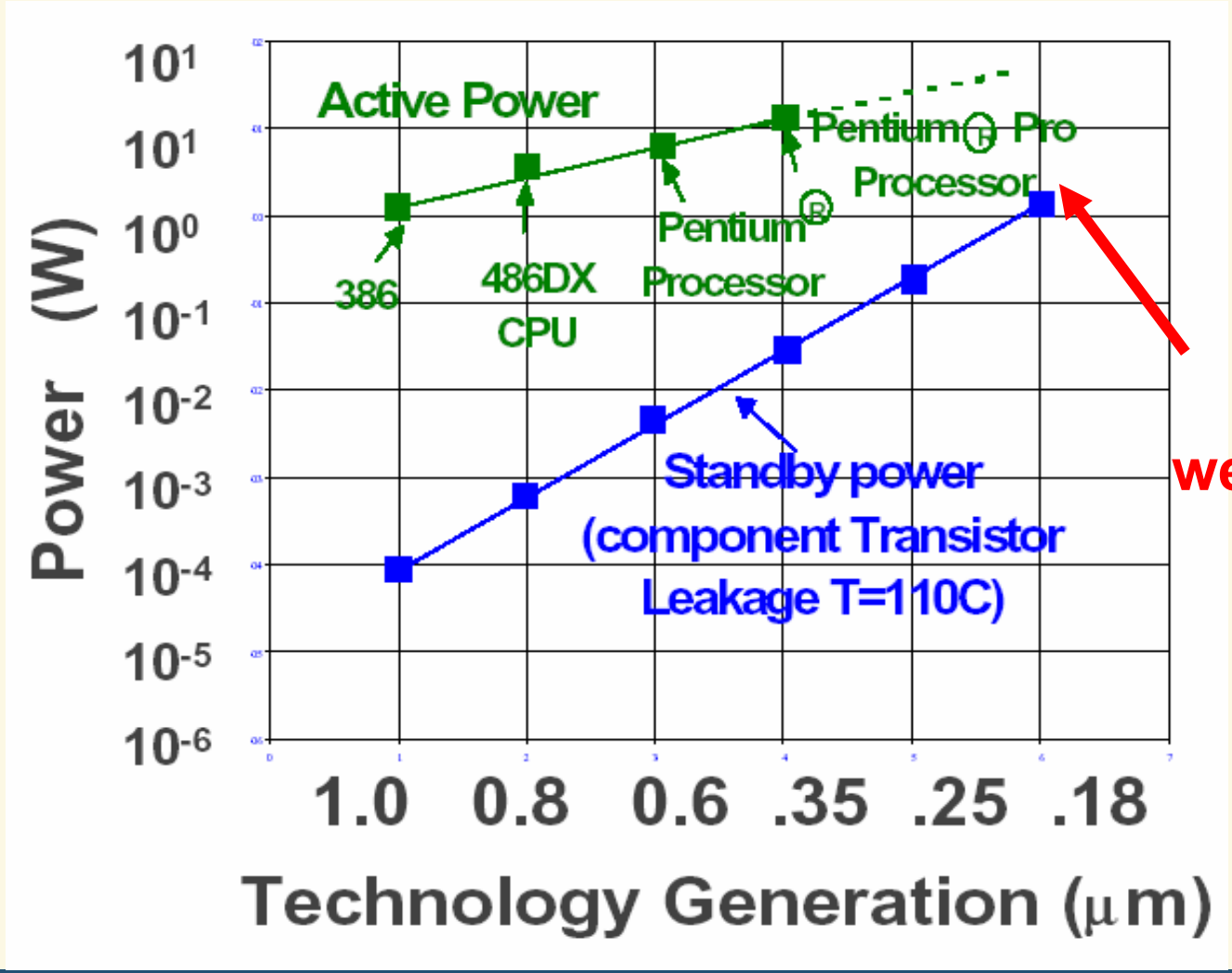
Battery capacity vs. Moore‘s Law



Trend power density



Dynamic Power vs. Leakage

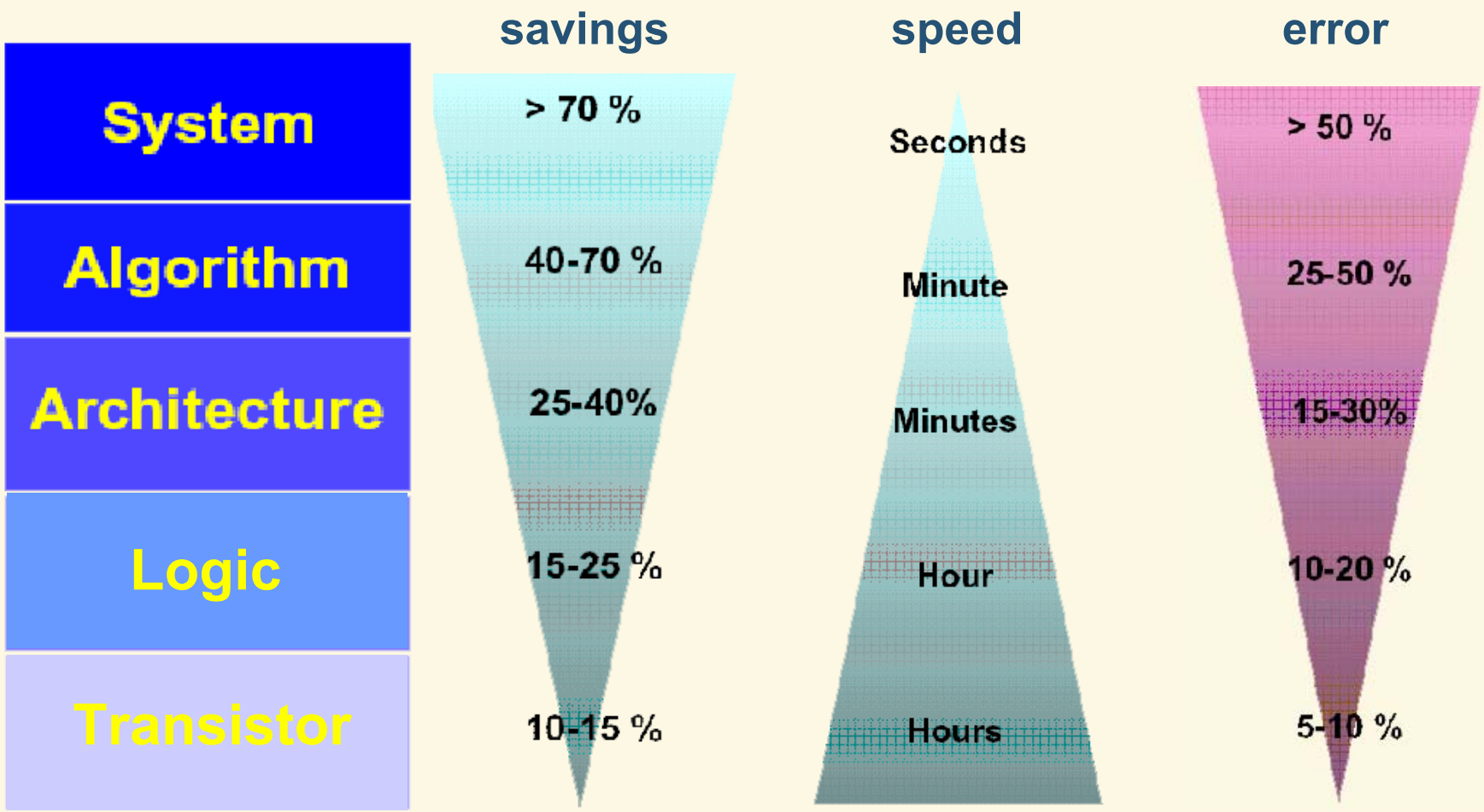


Outline

✓ Power dissipation in today‘s applications

2) Abstraction levels
   - Overview
   - technology level
   - circuit/gate level
   - architecture level
   - algorithm level
   - system level
3) Leakage currents
4) Summery

Levels of Optimization



Sources of Power consumption

- Dynamic Power Consumption
  - charging and discharging capacitors
- Short Circuit Currents
  - short circuit path between supply rails during switching
- Leakage Current
  - leaking diodes and transistors
- Static Currents
  - design styles such as pseudo NMOS

CMOS Gate Energy & Power Equations

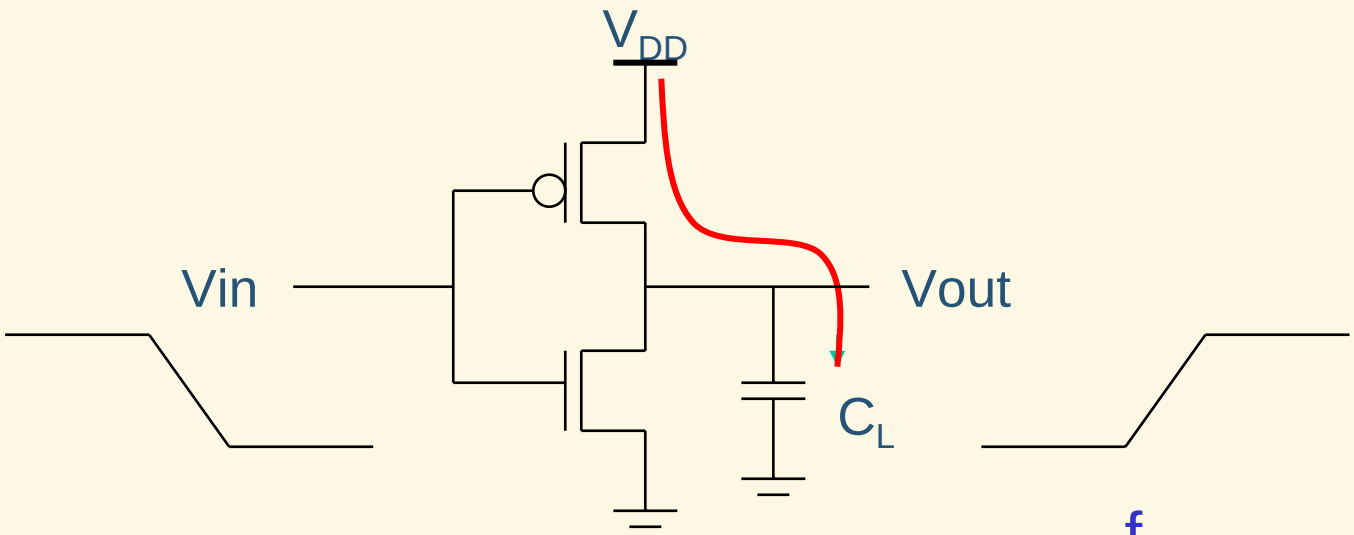
$$P = C_L V_{DD}^2 f_{0\rightarrow1} + t_{sc} V_{DD} I_{peak} f_{0\rightarrow1} + V_{DD} I_{leakage}$$

Dynamic term (~82% today and decreasing relatively)

Short-circuit term (~8% today and decreasing absolutely)

Leakage term (~10% today and increasing)

Dynamic Power Consumption



$$\text{Power} = C_L * V_{DD}^2 * W_{0\to1} * f$$

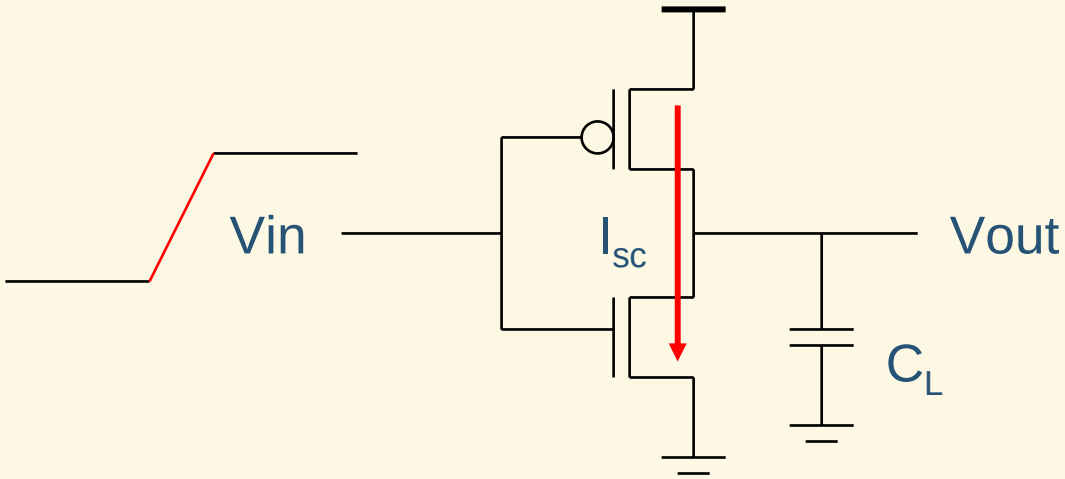
$f_{0\to1}$

Data dependent - a function of switching activity!

Lowering Dynamic Power

- Reducing V_{DD} has a quadratic effect!
  - Has a negative effect on performance especially as V_{DD} approaches $2V_T$
- Lowering C_L
  - Improves performance as well
  - Keep transistors minimum size
- Reducing the switching activity, $f_{0\rightarrow1} = W_{0\rightarrow1} * f$
  - A function of signal statistics and clock rate
  - Impacted by logic and architecture design decisions

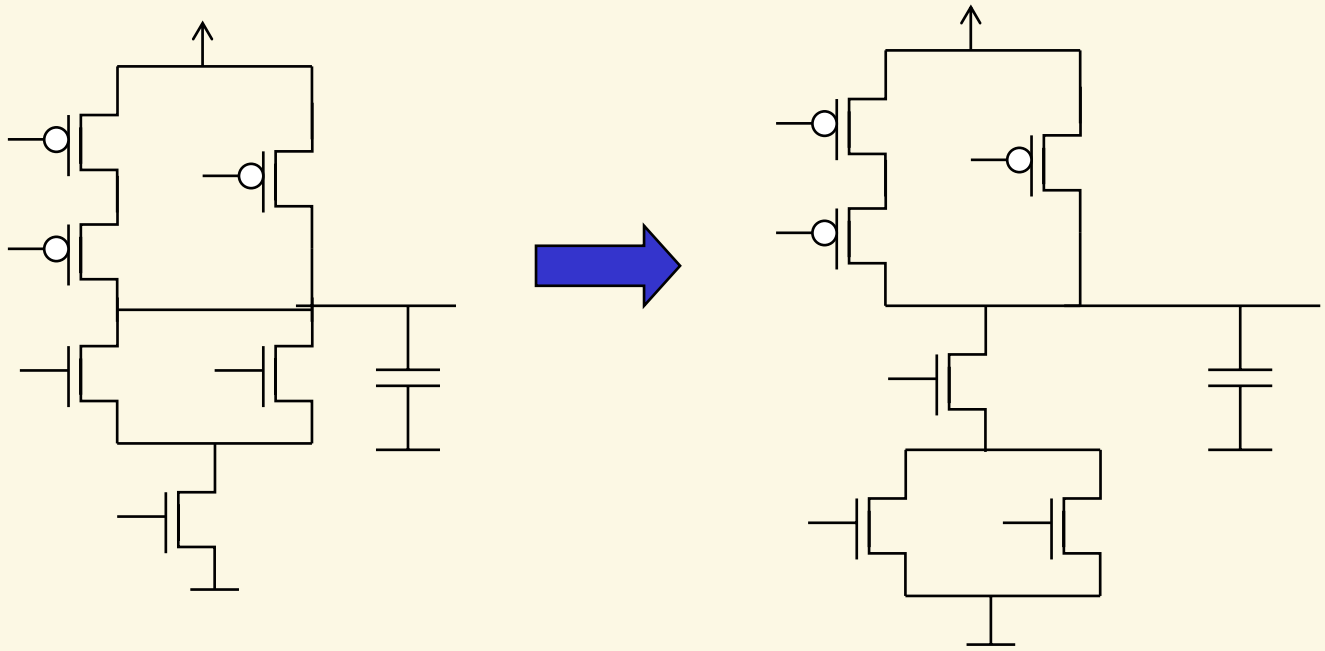
Short Circuit Power Consumption



Finite slope of the input signal causes a direct current path between V_{DD} and GND for a short period of time during switching when both the NMOS and PMOS transistors are conducting.

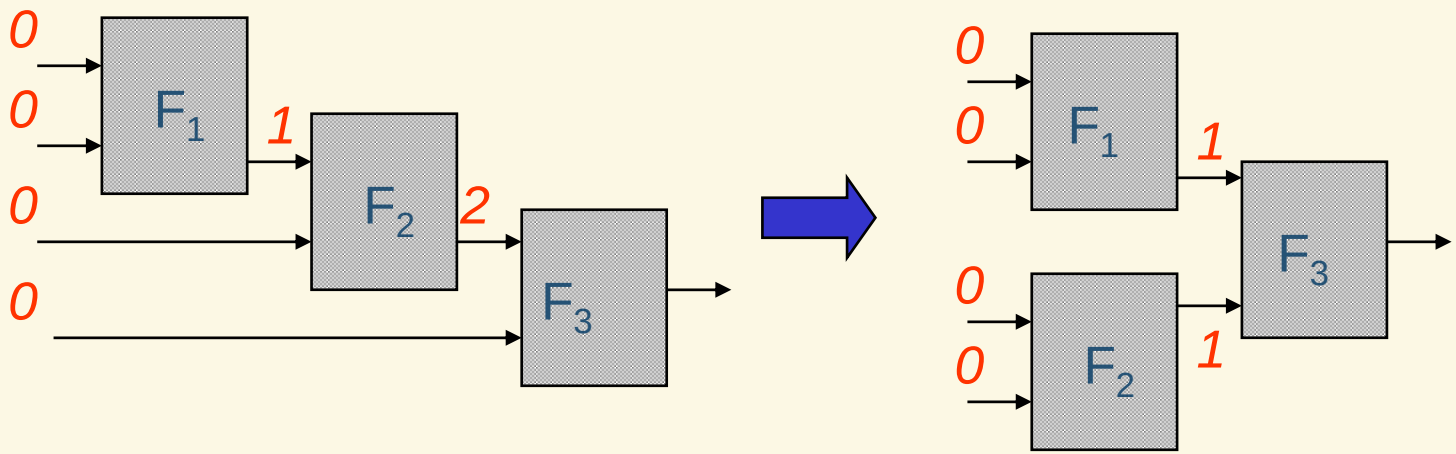
Logic Level: Gate Restructuring

Logically equivalent CMOS gates may not have identical energy/delay characteristics



Balanced Delay Paths

Reduce glitching by balancing the delay path



Equalize lengths of timing paths through logic

Dual Supply Voltages

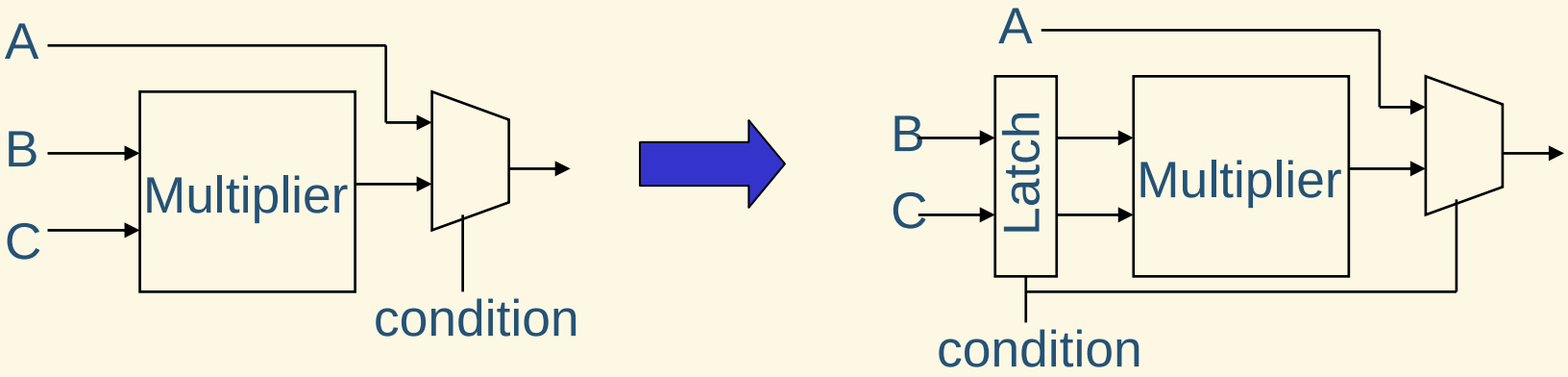
- Use two V_{DD}'s (e.g., 1.8V and 1.1V)
  - use the higher supply for gates on the critical path
  - use the lower supply for gates off the critical path
- Reduces energy without a performance loss
- Cons
  - slight area penalty
  - increased design time
  - need level converters to interconnect gates on different supplies (to avoid static currents)

Architecture level

- Architecture level techniques
  - processor sleep modes
  - pipelining, clock and signal gating
  - guarded evaluation, precomputation, power aware state encoding, etc.

Guarded Evaluation

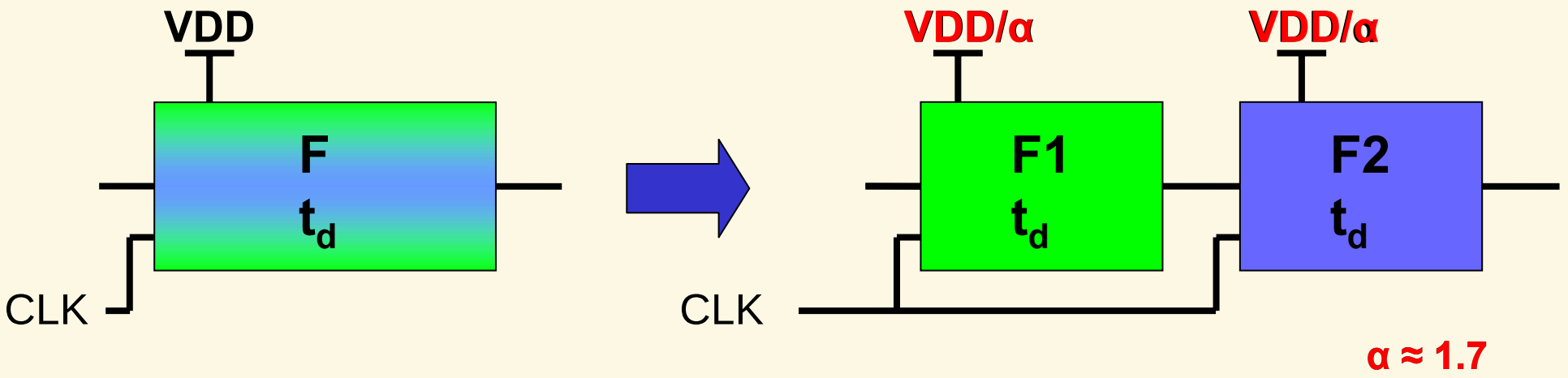
- Reduce switching activity by adding latches at the inputs if outputs are not used



- Latch preserves previous value of inputs to suppress activity
- could also use AND gates to mask inputs to zero → forced zero

Pipelining

- Reduces the propagation time of a block by half
- Voltage can be reduced, without changing operating frequency
- The throughput does not change



Algorithm level

Sources of Software Energy Consumption

- Datapaths in integer ALU and FP units
- Cache and memory systems
- System buses (address, instruction, and data)
- Control circuitry and clock logic and distribution

Levels of Optimization

- Algorithm and Application Design
- Compiler Optimizations
- Operating System Control

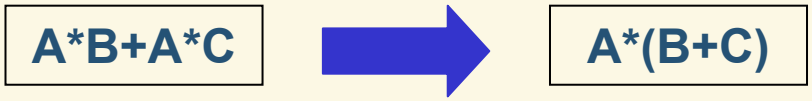
Coding styles

- Use processor-specific instruction style:
  - Variable types
  - Function calls style
  - Conditionalized instructions (for ARM)
- Follow general guidelines for software coding
  - Use table look-up instead of conditionals
  - Make local copies of global variables so that they can be assigned to registers
  - Avoid multiple memory look-ups with pointer chains

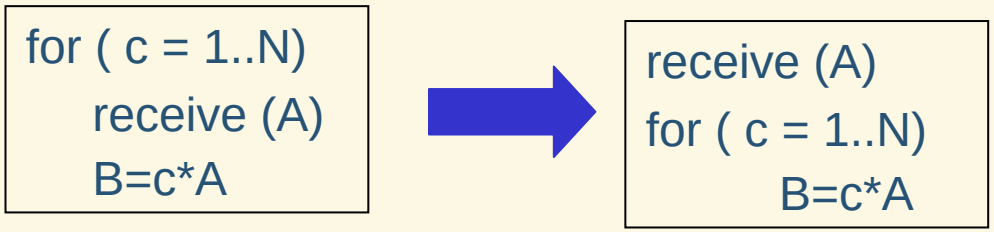
Source-code transformations

Minimize power-consuming activity

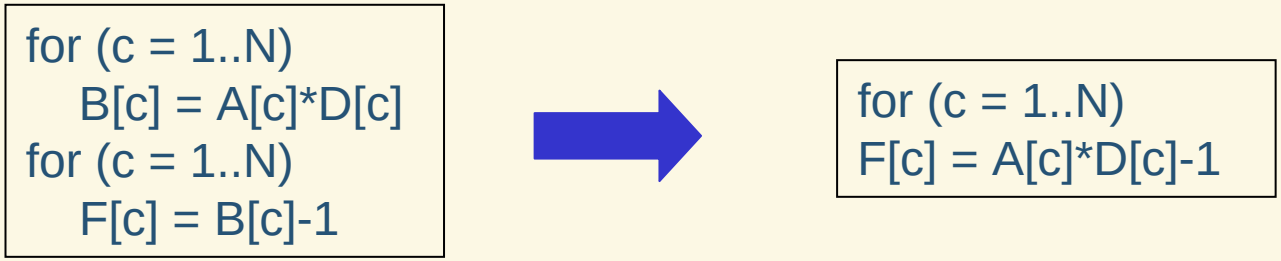
- Computation



- Communication



- Storage

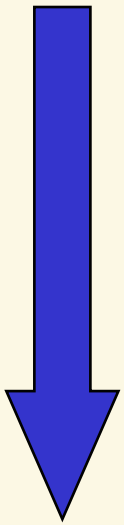


System level: Dynamic power management

- Systems are:
  - Designed to deliver peak performance, but …
  - Not needing peak performance most of the time
- Components are idle sometimes
- Dynamic power management (DPM):
  - Puts idle components in low-power non-operational states when idle
- Power manager:
  - Observes and controls the system
  - Power consumption of power manager is negligible

Processor Sleep Modes

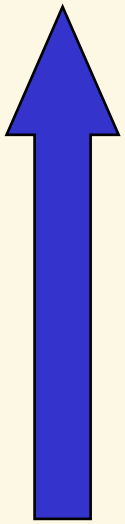
Software power control - power management



DOZE most units stopped except on-chip cache memory (cache coherency)

NAP cache also turned off, PLL still on, time out or external interrupt to resume

SLEEP PLL off, external interrupt to resume



Deeper sleep mode saves more power

Deeper sleep mode requires more latency to resume

Transmeta LongRun

- Hardware that scales supply voltage and clock frequency in response to software demands
  - 32 levels of V_{DD} (use 5 to 7 in practice) from 1.1V to 1.6V
  - clock frequency from 200MHz to 700MHz in increments of 33MHz
- Controlled through 5 pin interface
  - triggered when CPU load change detected by software
    - heavier load $\rightarrow$ ramp up supply voltage, when stable scale up clock frequency
    - lighter load $\rightarrow$ scale down clock frequency, when PLL locks onto new rate, ramp down supply voltage
  - always keeps clock frequency within limits required by supply voltage to avoid clock skew problems

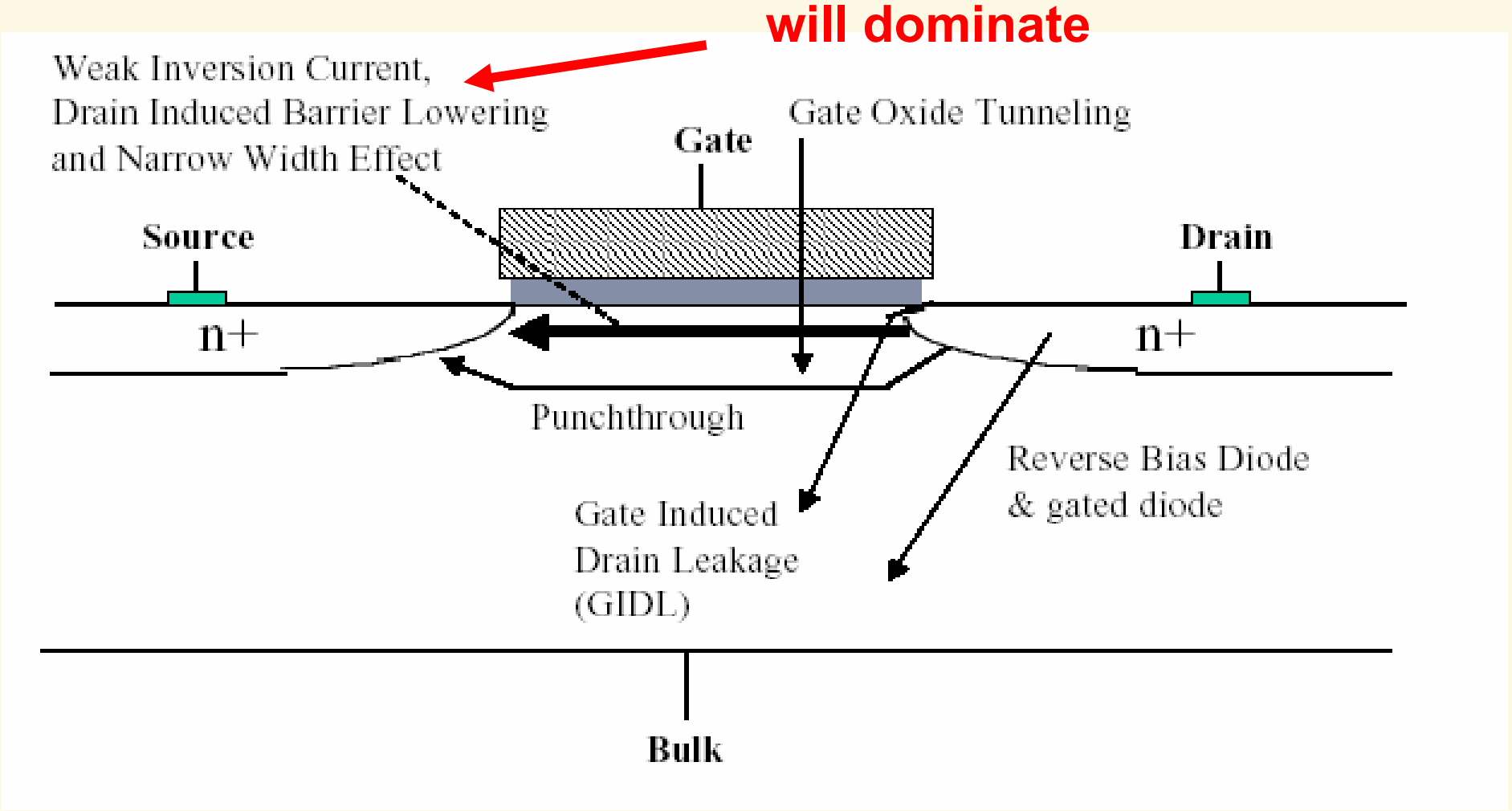
Outline

- ✓ power dissipation in today‘s applications
- ✓ Abstraction levels

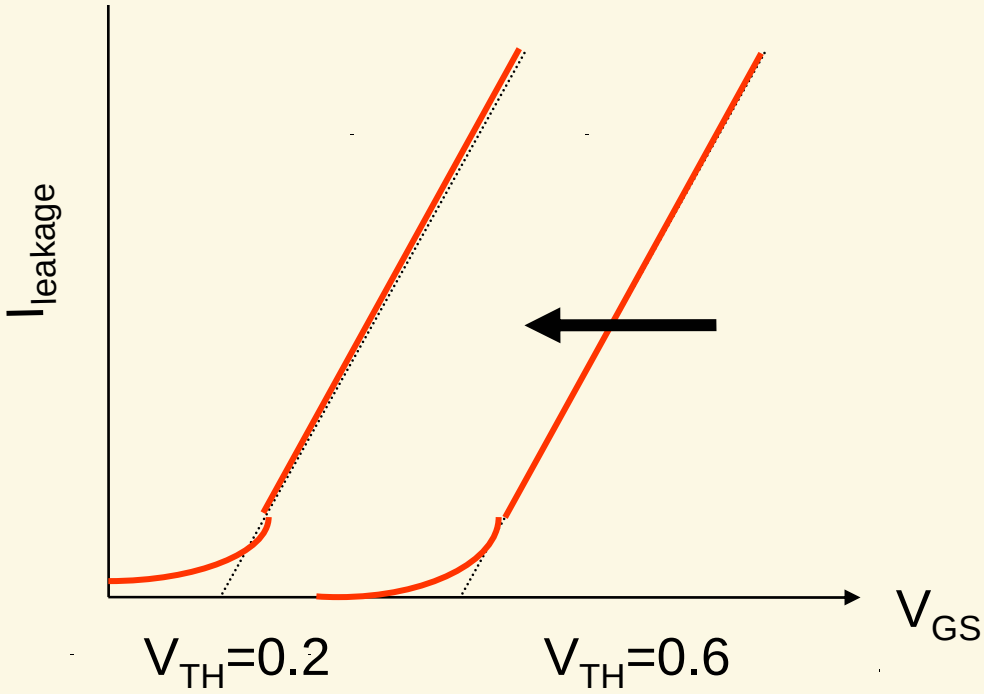
3) Leakage currents
   - overview
   - MTCMOS
   - VTCMOS
   - stacking
   - input vector control

4) Abstract

Contributions to Leakage



Leakage vs. V_{TH}

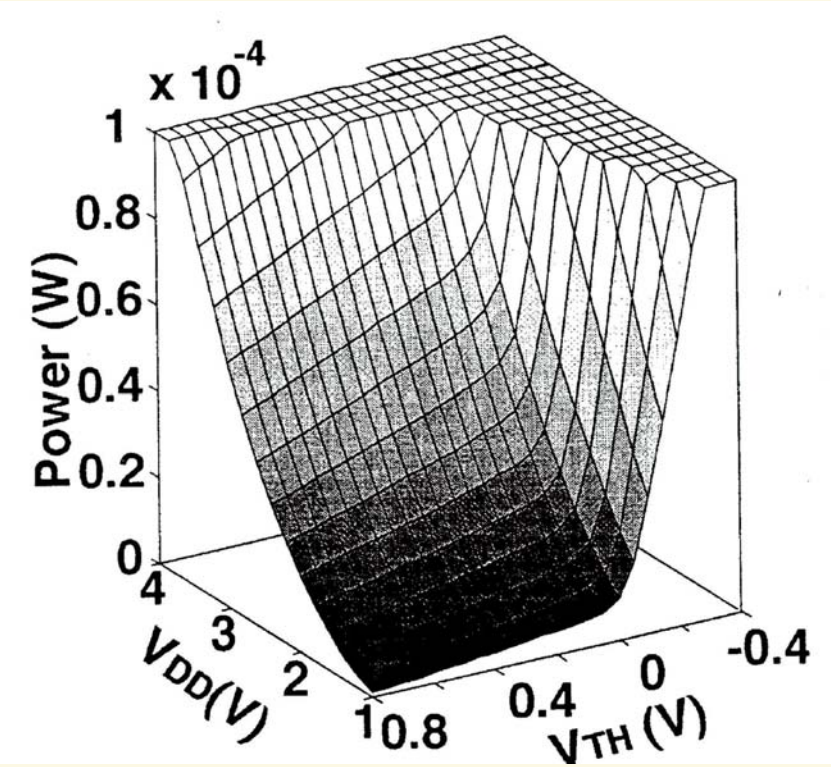


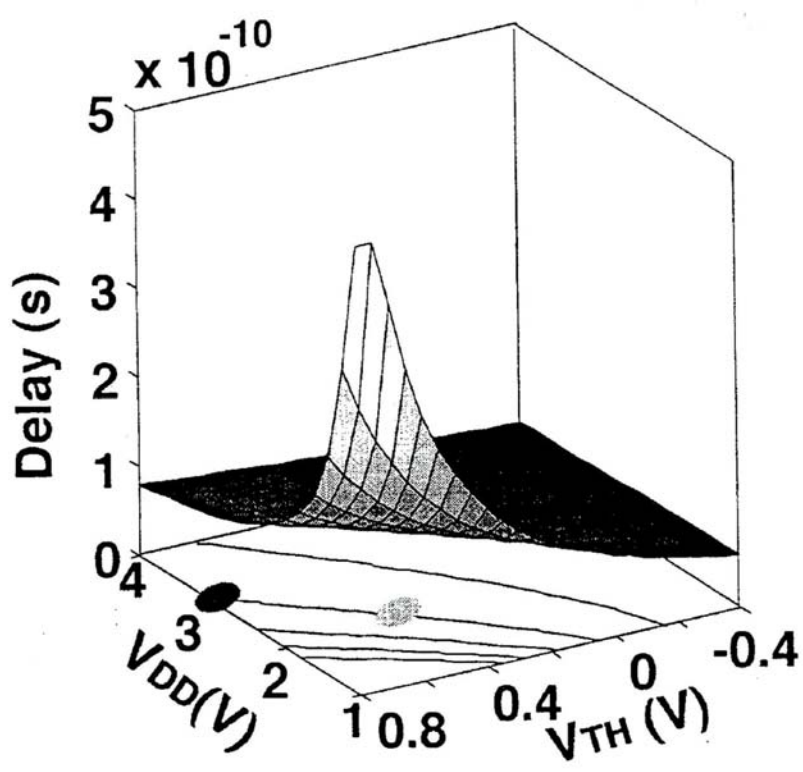
Continued scaling of supply voltage will make subthreshold conduction a dominate source of power dissipation.

Power & Delay dependence

$$P = p_t \cdot f_{CLK} \cdot C_L \cdot V_{DD} + I_0 \cdot 10^{\frac{-V_{TH}}{S}} \cdot V_{DD}$$

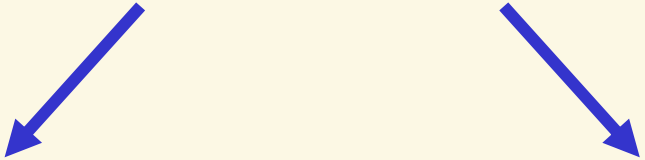
$$t_d = \frac{k \cdot Q}{I} = \frac{k \cdot C_L \cdot V_{DD}}{(V_{DD} - V_{TH})^{\alpha}}$$





Approaches to reduce leakage

approaches for different states



Idle states

- transistor stacking
- input vectors for standby mode (MLV)
- SCCMOS

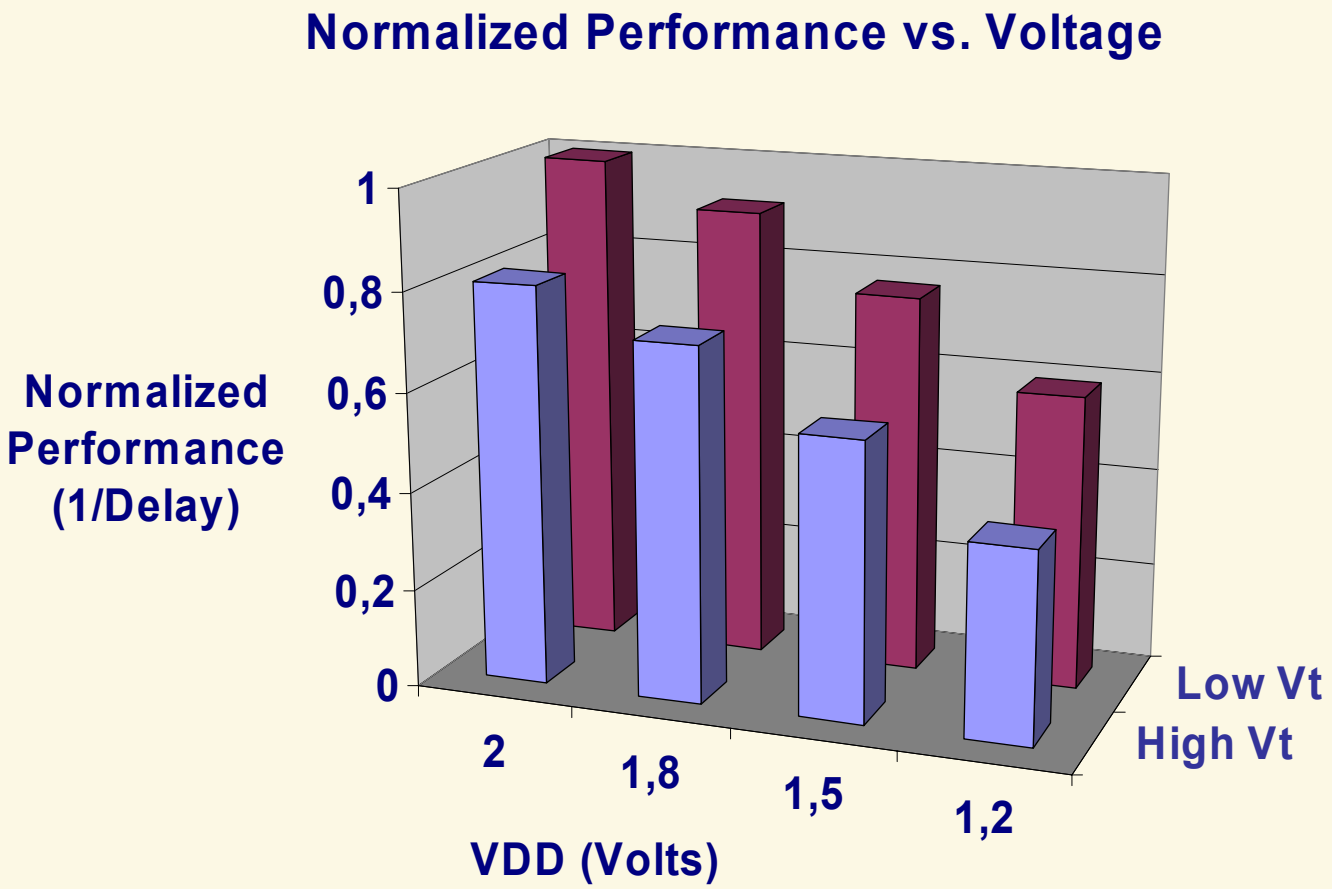
Active states

- multiple threshold transistors (MTCMOS, MHS-Domino)
- variable threshold transistors (VTCMOS, DTMOS)

Multiple threshold voltages (MTCMOS)

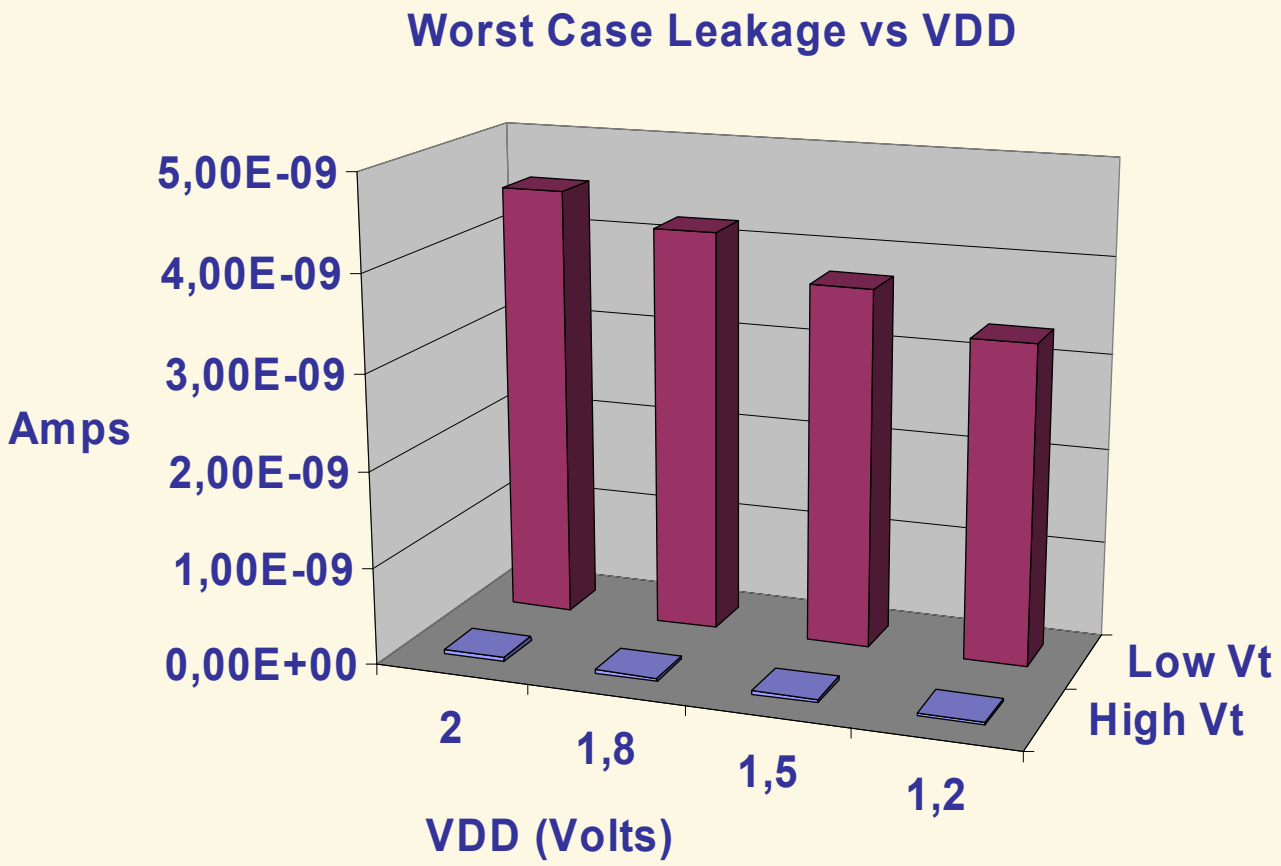
- Use different V_T's (e.g., 0.6V and 0.3V for V_{DD} = 2.5V)
  - use the lower threshold for gates on the critical path
  - use the higher threshold for gates off the critical path
- Improves performance without an increase in power
- Cons
  - increased fabrication complexity
  - increased design time

MTCMOS - Benchmark



Based on NAND3
TSMC 0.25um
Process

MTCMOS – Benchmark cont'd



MHS - Domino

- Multi-Threshold High Speed Domino
- MTCMOS - implementation
- modifies a clock-delayed keeper circuit to force internal dynamic nodes into a low leakage state

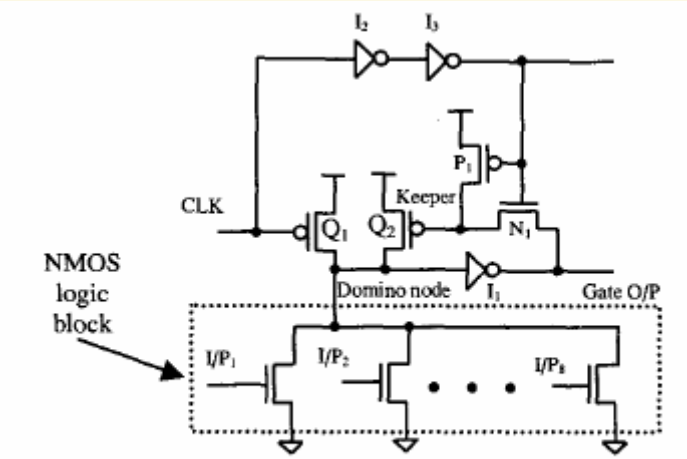


Figure 3: An 8-input HS-Domino OR gate

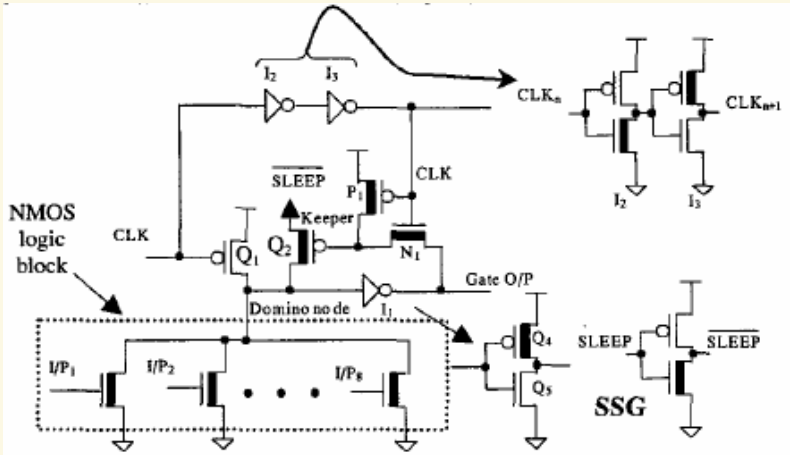
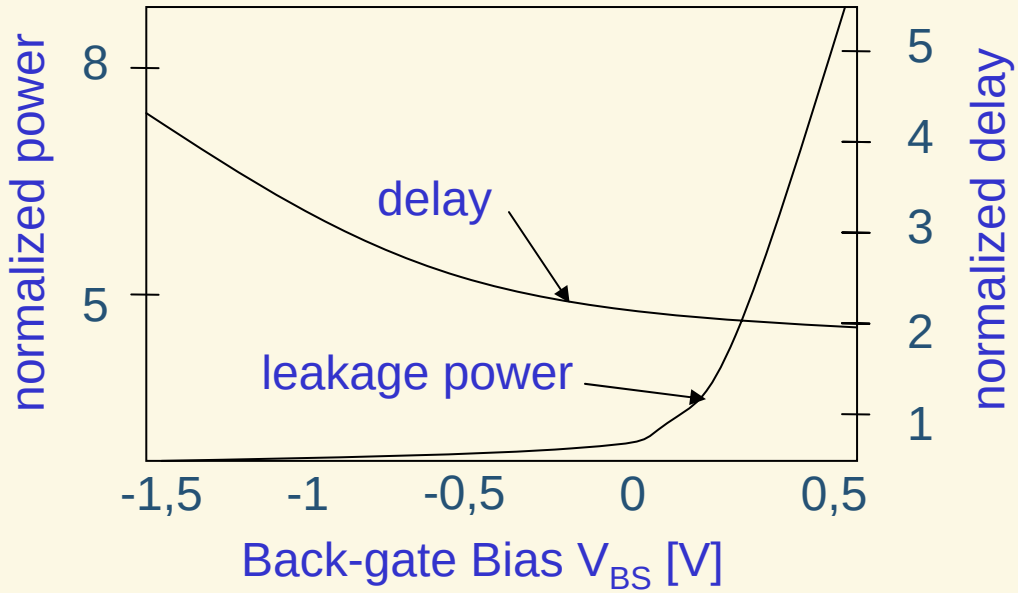


Figure 6: An 8-input MHS-Domino OR Gate

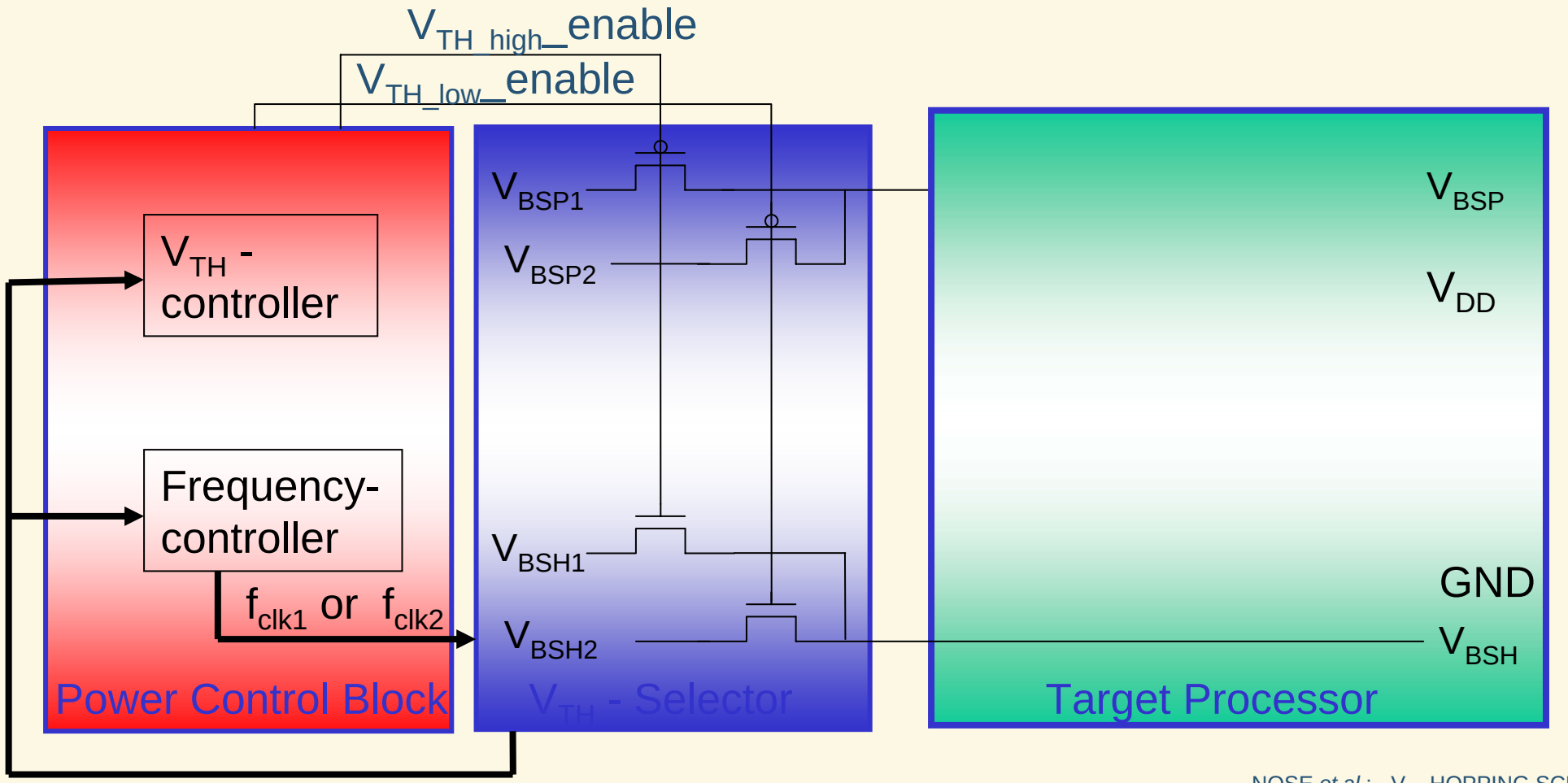
M. W. Allam *High-speed dynamic logic styles*

VTCMOS (Variable Threshold CMOS)

- threshold voltage dynamically controlled
- dynamic adjustment of frequency and V_{TH} through back-gate bias control

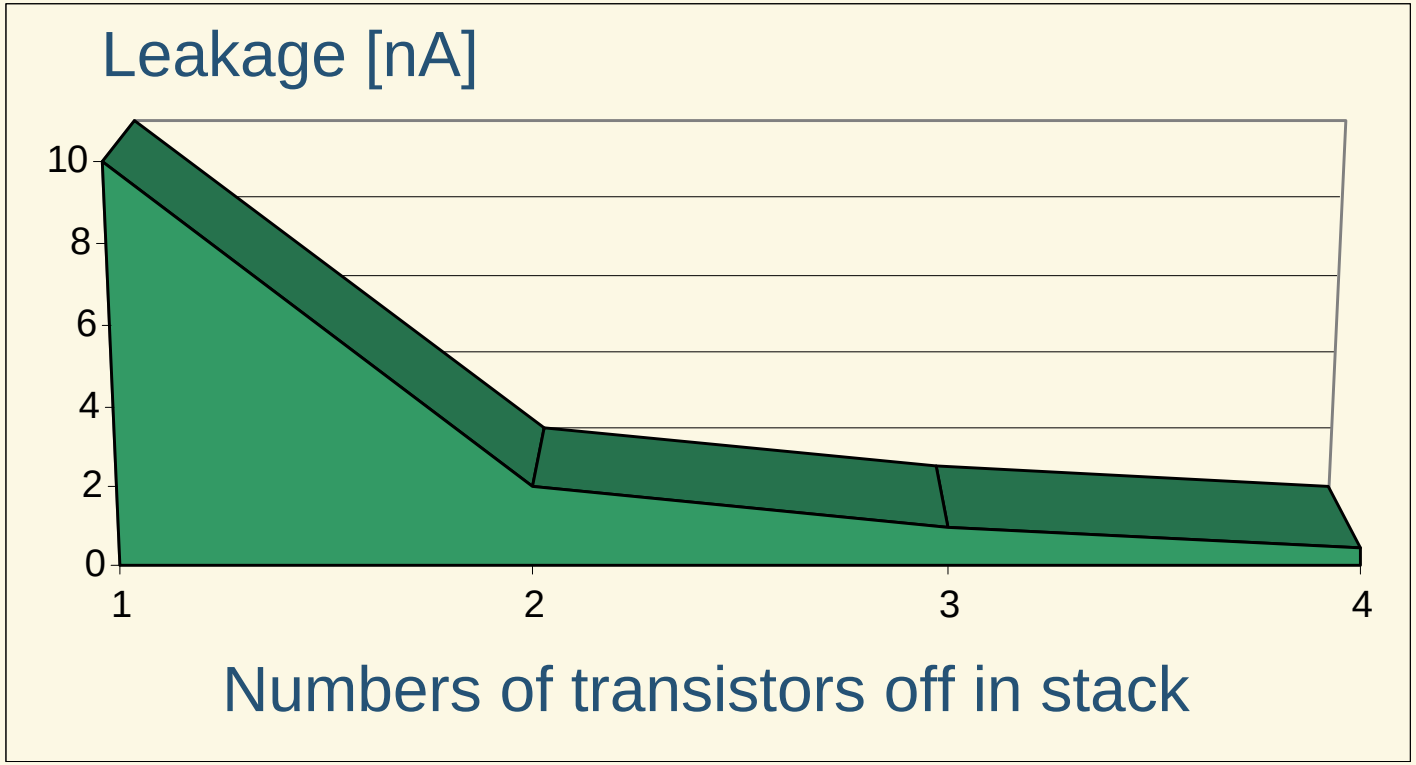


VTCMOS: V_{TH}-hopping scheme



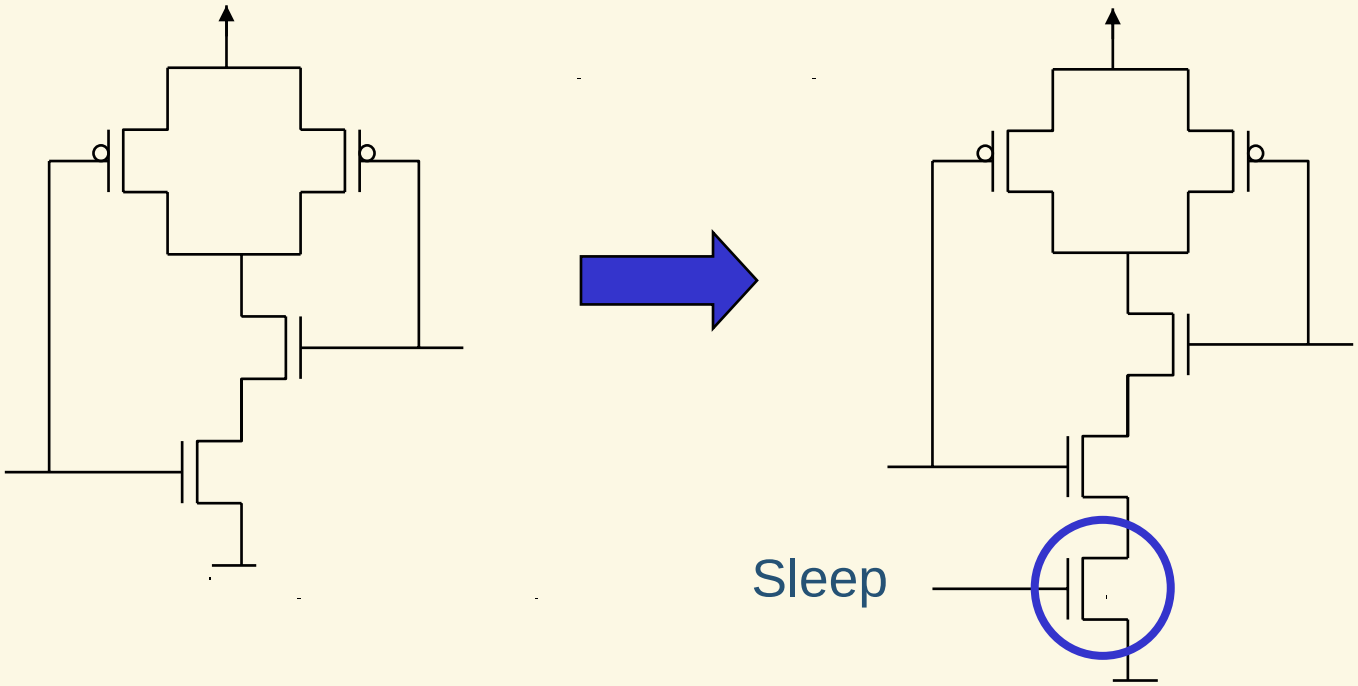
NOSE *et al.*: - V_{TH} HOPPING SCHEME

Leakage vs. Transistors Off



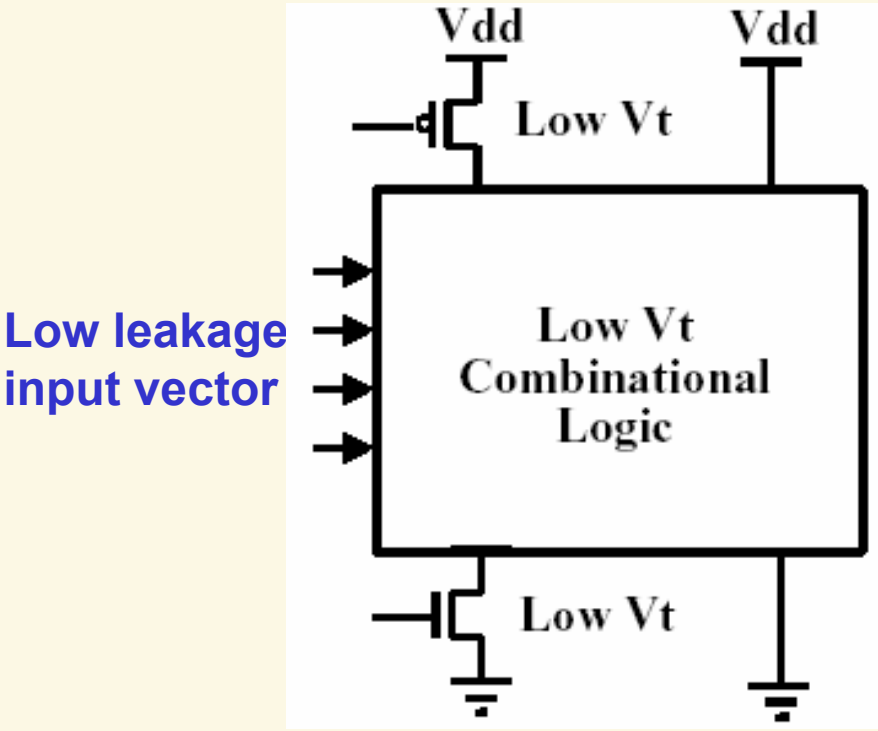
Roy *Lecture*

Transistor Stacking



Leakage Control

- Sleep mode:
  - low leakage input vector
  - leakage control transistors off
- Sub-circuits in low leakage state or in critical path
  - Normal power & ground
- Sub-circuits in high leakage state
  - Power & ground via leakage control transistors



Outline

- ✓ Power dissipation in today‘s applications
- ✓ Abstraction levels
- ✓ Leakage currents

4) Abstract

Summery

- Increasing Power dissipation in future devices
- Different levels for solutions
- Leakage Currents increase strong
- Approaches differ between solutions for idle and active states

Reserve

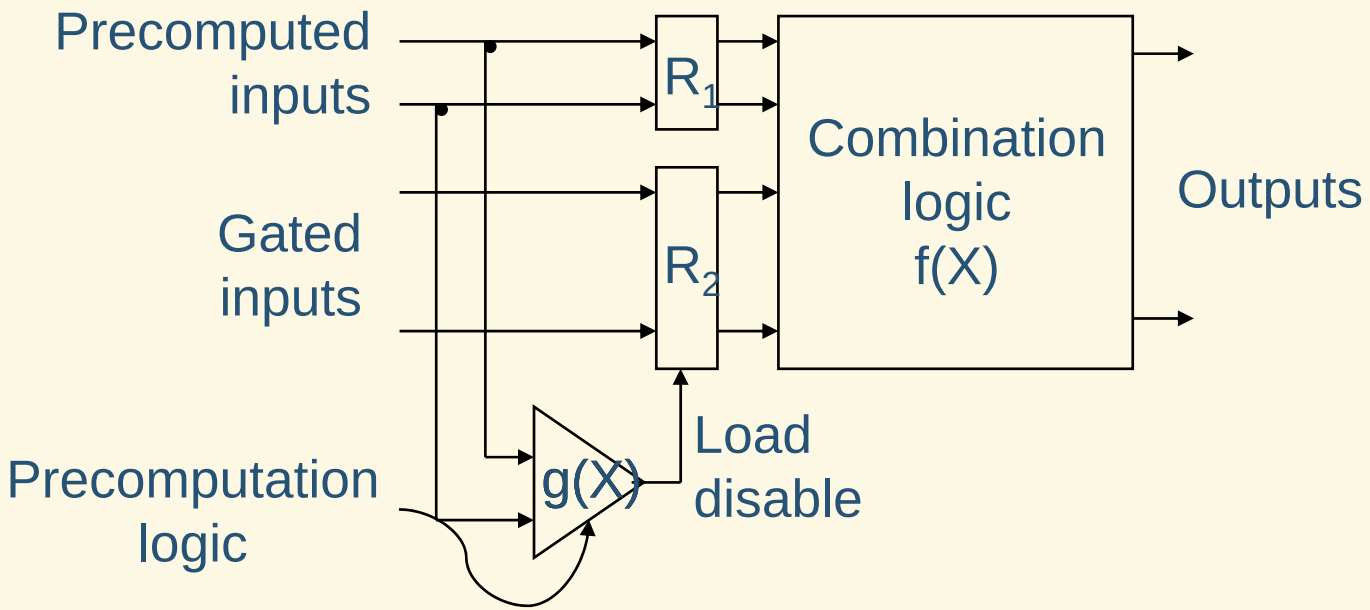
Algorithm - Power Tradeoffs

- Trading of Quality for Power
  - Black and White Video instead of Color Video
  - Fetch Map with less details
  - Scalable Encryption Algorithms
  - Network-driven partitioning of algorithms

Power vs. Energy

- Power consumption in Watts
  - determines battery life in hours
  - sets packaging limits
- Peak power
  - determines power ground wiring designs
  - impacts signal noise margin and reliability analysis

Precomputation

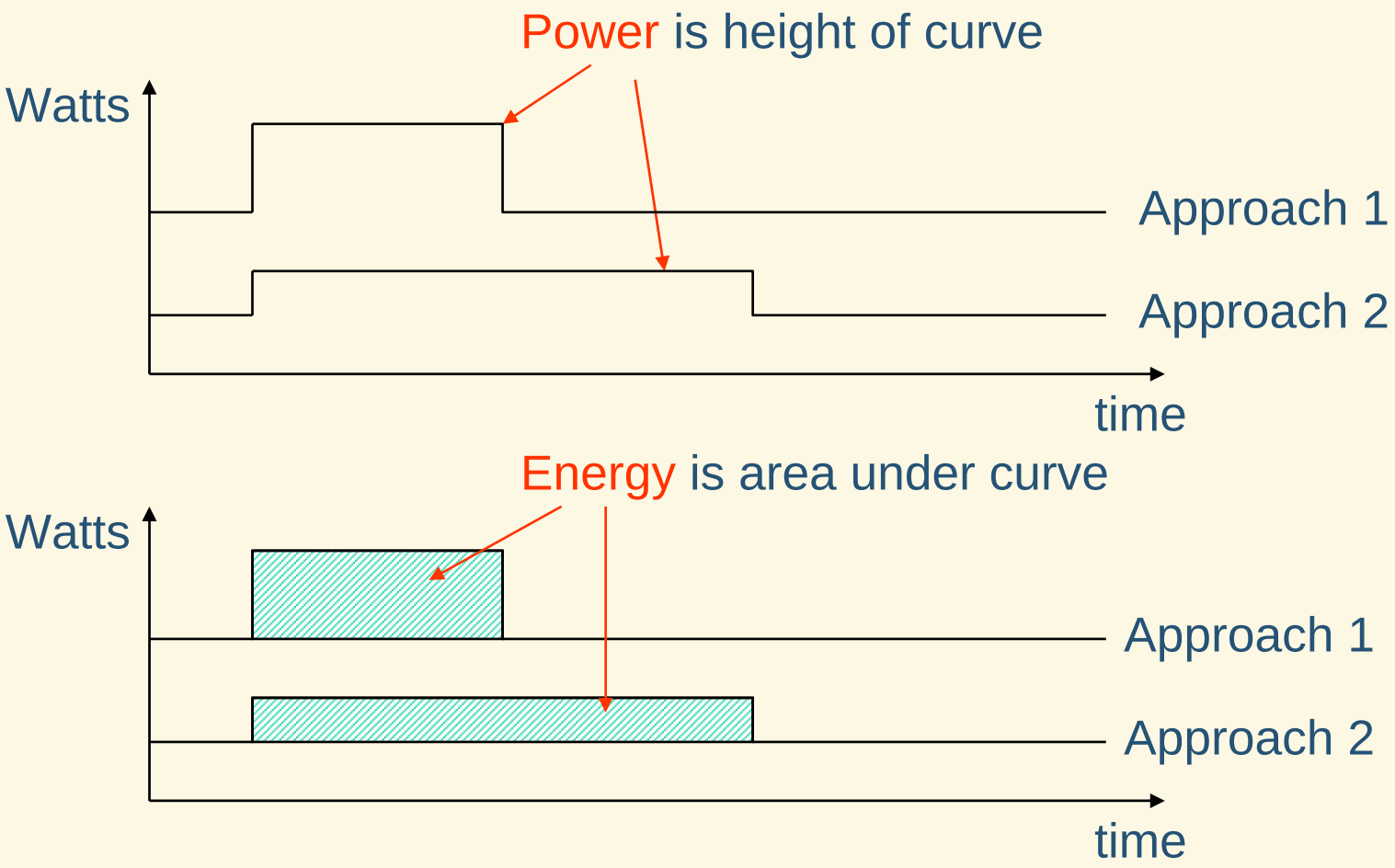


- Identify logical conditions at inputs that are invariant to the output
  - since those inputs don't affect output, disable input transitions
  - trade area for energy

Energy vs. Power cont'd

- Energy efficiency in Joules
  - rate at which energy is consumed over time
- Energy = power * delay
  - joules = watts * seconds
  - lower energy number means less power to perform a computation at the same frequency

Power vs. Energy cont'd



Intel's SpeedStep

- Hardware that steps down the clock and supply voltage when the user unplugs a mobile computer from AC power
  - step down the PLL from 650MHz → 500MHz

  NEUE WERTE aus I-NET

- CPU stalls during SpeedStep adjustment

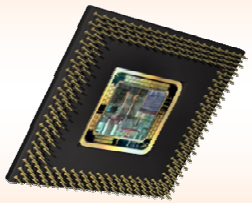
Dissipation in a Notebook

Peripherals

Disk

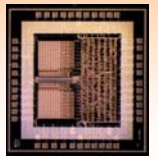
Display

Processing



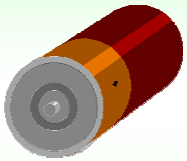
programmable µPs or DSPs

ASICs



Memory

Power supply



Battery

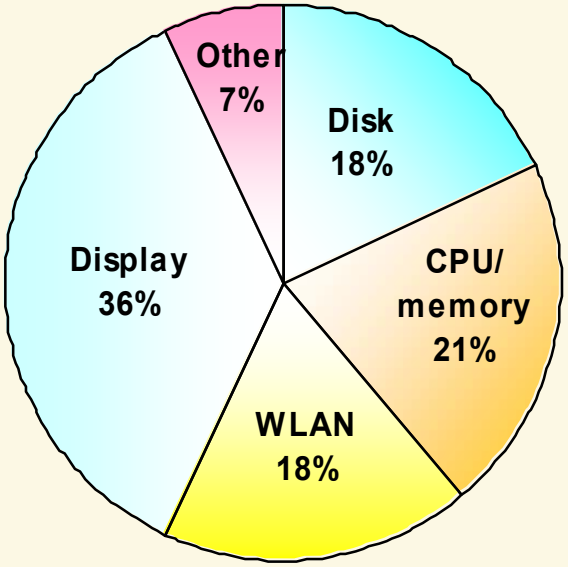
DC-DC converter

Communication

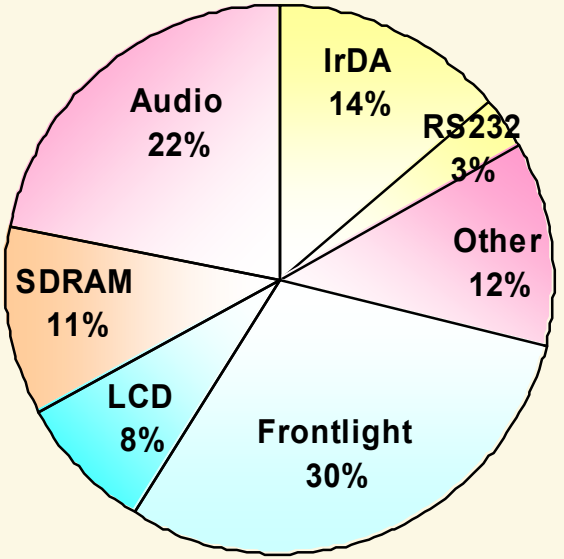
WLAN

Ethernet

Power dissipation



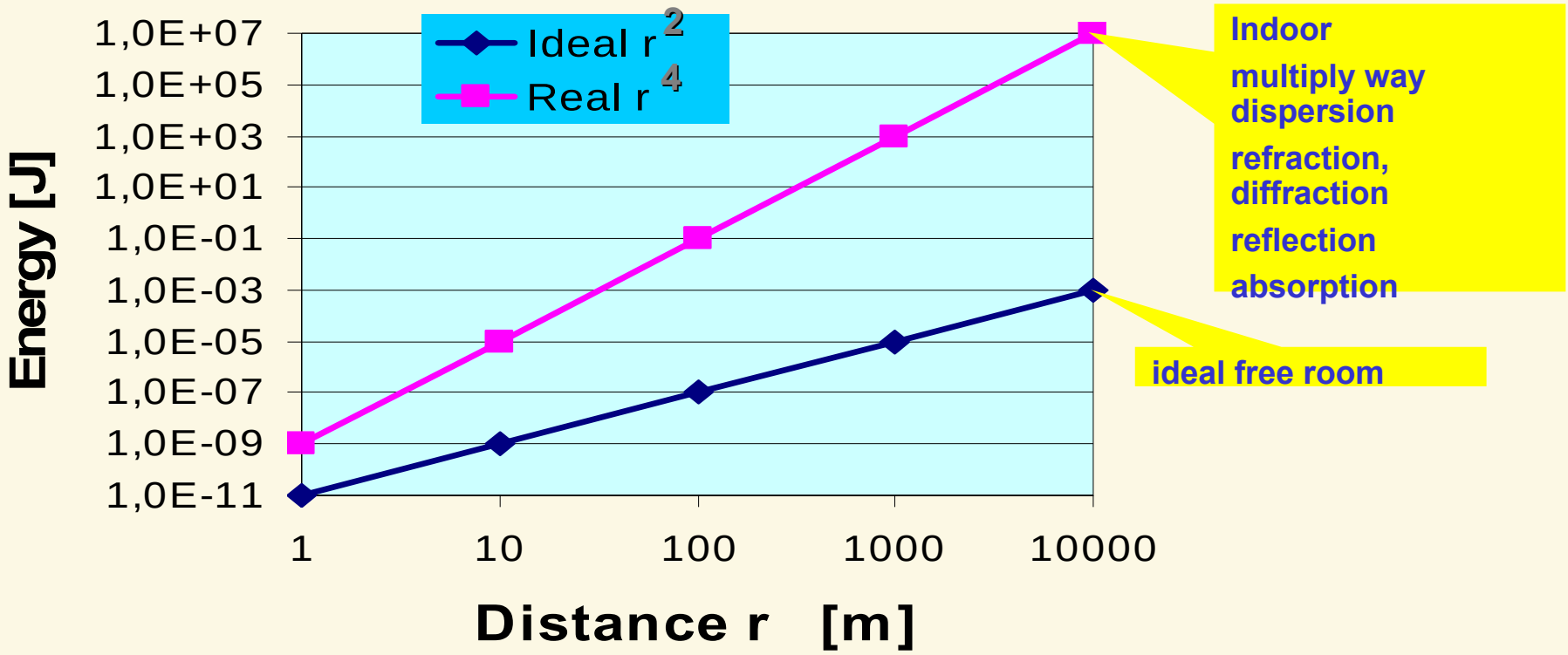
Power dissipation in a notebook



Power dissipation ar Compaq‘s iPAQ

Power for sending

- needed power to send over distance *r* with constant receive energy



Transmeta LongRun, con't

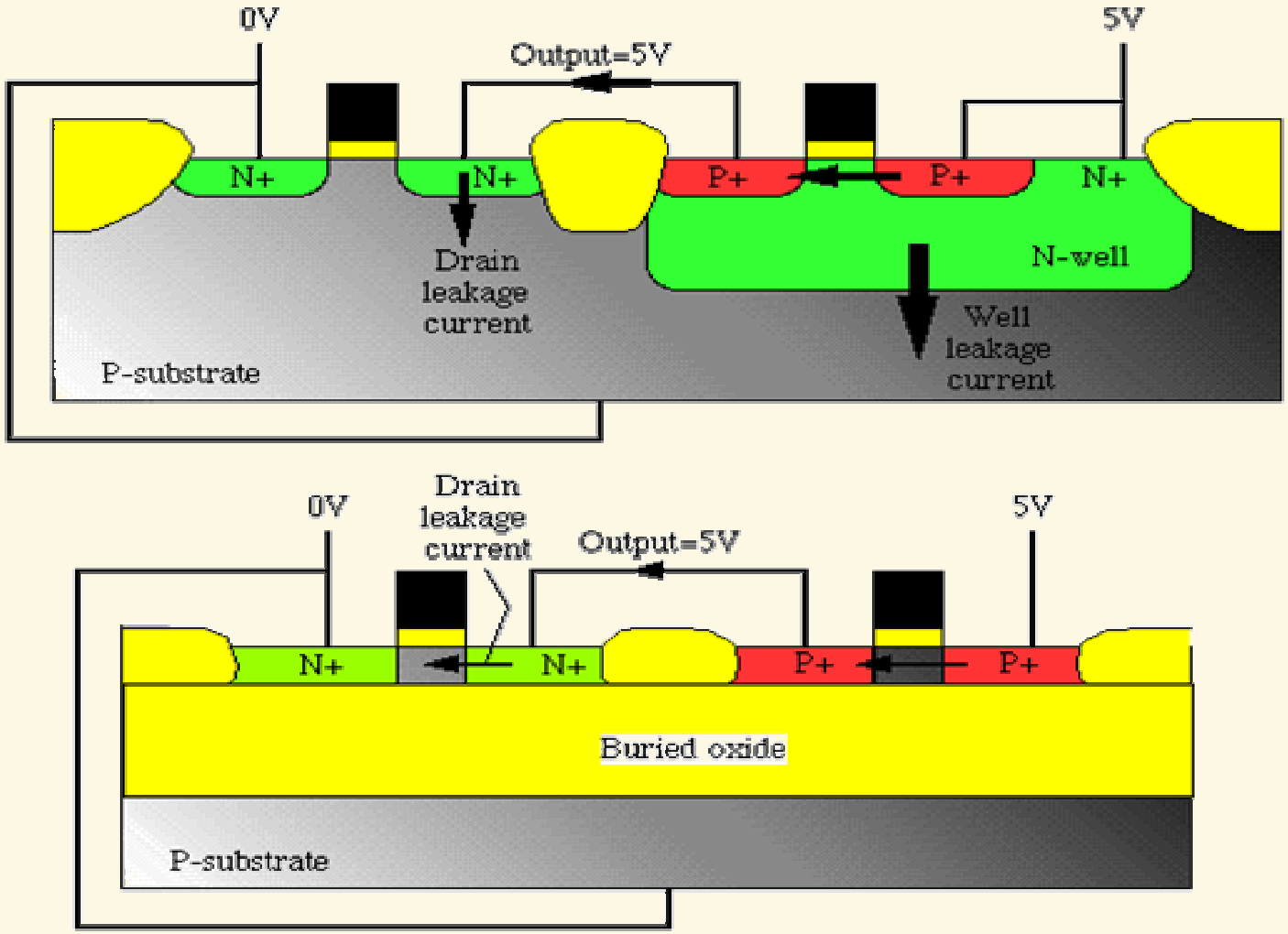
- Software can detect CPU load change in ~ 1/2 microsecond
- LongRun scales
  - the voltage up or down in less than 20 microseconds per step
  - the clock frequency in one step
- Worst-case scenario of a full swing from 1.1V to 1.6V and from 200MHz to 700MHz takes 280 microseconds
- CPU stalls only during PLL relock (< 20 microseconds in the worst case)

New Technologies - SOI

- Scaling of bulk CMOS below 0.13 micron is extremely difficult due to short-channel effects
  - As transistor channel length shrinks, parasitic factors become dominate
    - loss of gate control (and transistor gain)
    - high gate-overlap capacitance
    - subthreshold leakage
    - tunneling
- Silicon on insulator (SOI)

New Technologies - SOI

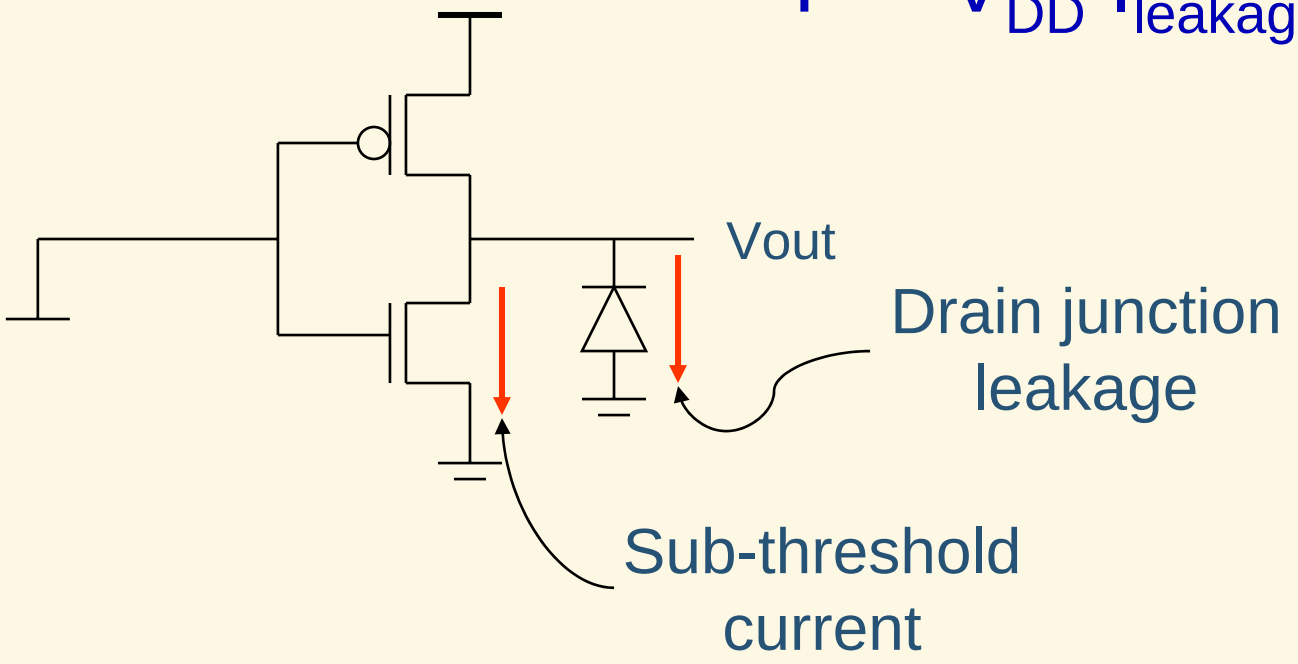
SOI completely isolates the transistor from its neighboring transistors and other circuit components with a layer of insulator material



Quelle: Nazmul Islam "SOI"

Leakage currents

$$P = V_{DD} I_{leakage}$$



Sub-threshold current is the dominant factor.
Increases exponentially with temperature!

Leakage currents

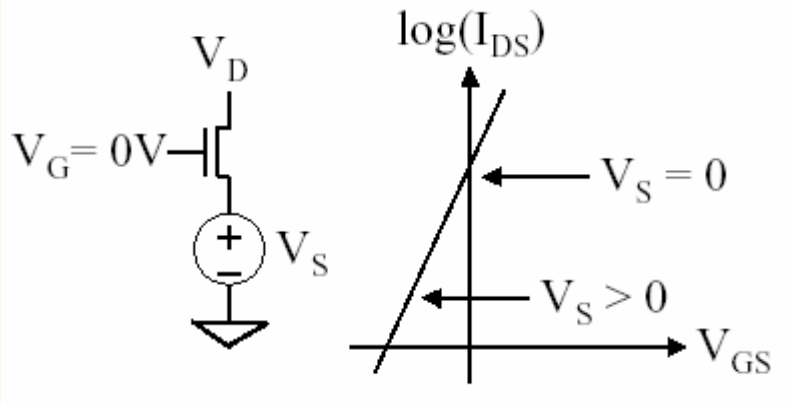
- Decreasing supply voltage means that threshold voltages will decrease to
  - help maintain signal integrity
  - buy back loss in speed

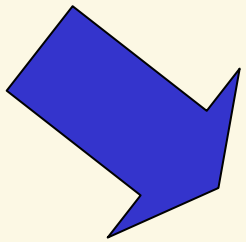
- But decreasing the threshold voltage means an increase in leakage current
  - increase in sub-threshold current
  - leakage currents increase exponentially with temperature!

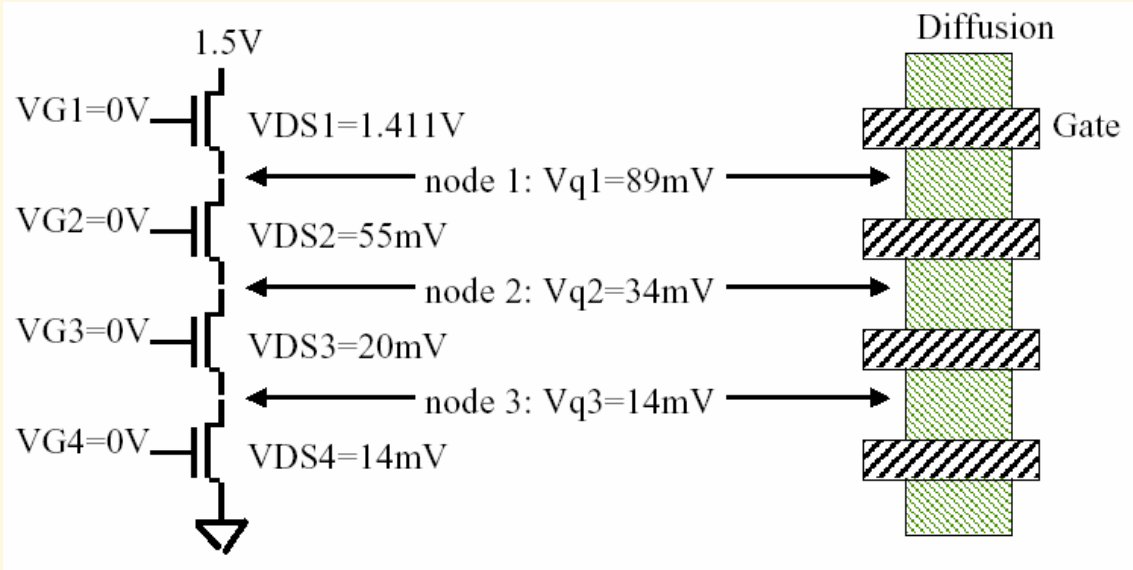
Leakage currents

- Transistor sizing
  - leakage current increases with decreasing VT and channel length while speed increases
- Transistor stacking
- Multiple threshold voltages (MTCMOS)
  - use lower threshold devices only where speed is important
- Limit leakage through dynamic control (threshold is lower when circuit is active, elevated when idle)
  - substrate biasing
- New technologies
  - SOI

Stacking Effect







increasing count of OFF transistors in stack, caused lower leakage

Energy Analysis Techniques

- **Simulation based techniques**
  - characterization - using lower level analysis results to construct higher level models
  - very computational intensive; cycle accurate
- **Probabilistic based techniques**
  - signals viewed as random zero-one processes with certain statistical characteristics
  - computationally efficient; average values for sequences of cycles
  - large errors are possible if statistical assumptions are incorrect

Leakage Estimation

- Motivation: Efficient and Accurate
- Idea and tools:
  - Gate leakage power characterization(Spice)
  - Leakage power Library (Library Compiler)
  - Logic Synthesis (Design Compiler)
  - Logic Simulation (Verilog – XL)
  - Power analysis (Power Compiler)
- Benchmark and Technology
  - ISCAS-89
  - 0.07um and 0.1um

Leakage Estimation cont'd

Equation: $P_{static} = V_{cc} N K_{design} I_{leak}$

V_{cc} : Power supply voltage

N : Number of transistor in design

K_{design} :

design dependent parameter empirically determined parameter representing the characteristics of an average device

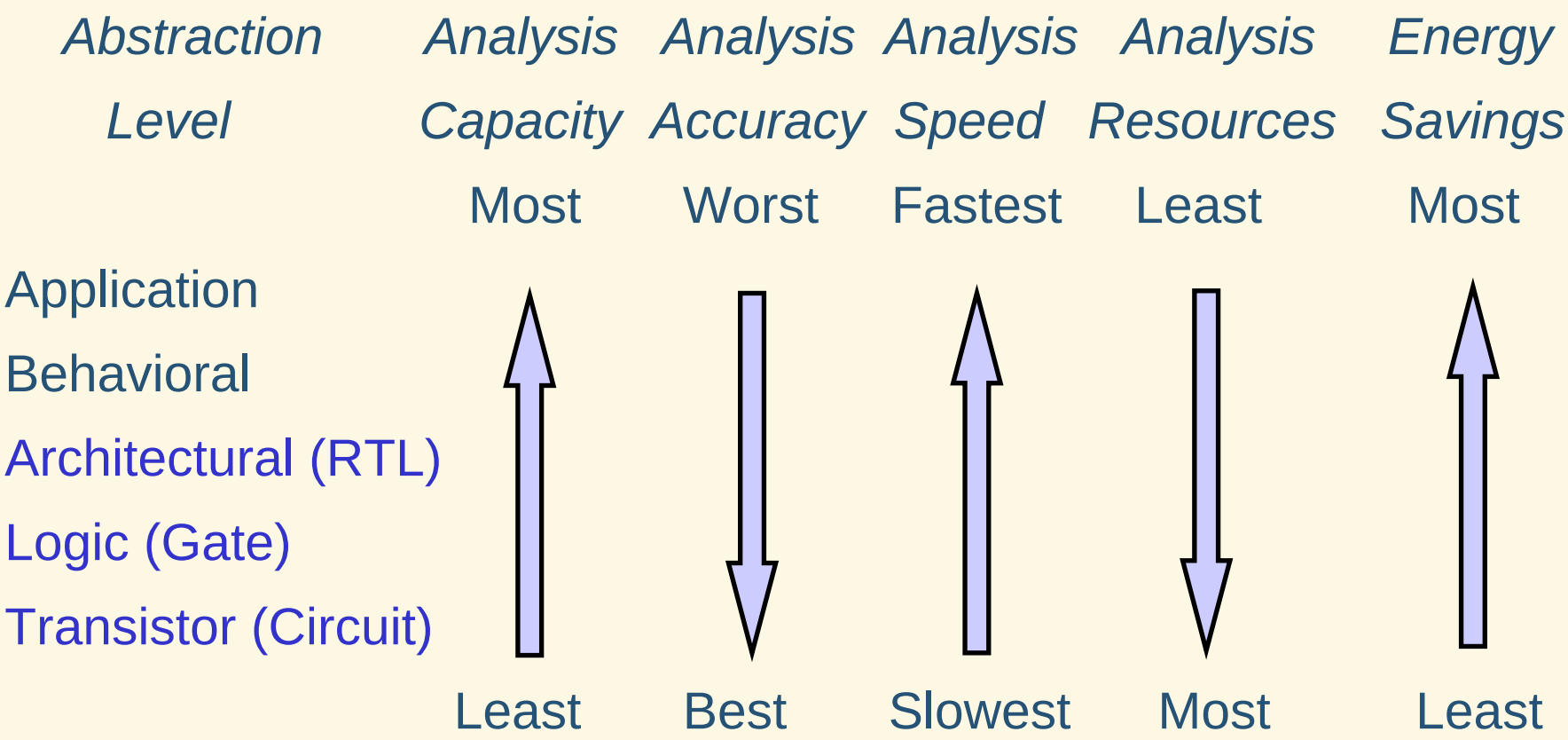
I_{leak} :

technology dependent parameter describing the per device sub-threshold leakage, highly dependent on aggressive of Vt scaling

Input Vector Determination Techniques

- The vector searching is one-time effort, and can be multiplexed onto the logic design.
- The searching for a low-leakage input vector depends on being able to estimate the leakage for a given vector
  - Spice and build up look-up table for each gate
  - Run a simple steady state logic simulation for the circuit for each vector.
- How many cycles needed:
  - $N>\ln(1-\alpha)/\ln(1-\varepsilon)$: α: confidence; ε: error tolerance.
  - For 95% confidence and 5% error tolerance, we need 59 vectors
- Comment:
  - this techniques ignore the circuit topology
  - the primary input combination is randomly generated with no knowledge of history information

Analysis Abstraction Levels



Architectural Level Simulation

- Perform simulation to obtain the input activity for each major functional unit
  - architectural description in behavioral VHDL or Verilog or C, C++
- Energy characterization of functional units
  - analytical energy models
    - caches, DRAMs
  - macro energy models
    - system buses
    - ALUs, register file, pipeline registers

Circuit Level Simulation

- Extract circuit netlist description from layout
  - captures internal (diffusion) and external (wiring and gate fanout) capacitances
- Run an analog simulation
  - characterization of device models
  - solution of large system of equations so very computationally intensive (< few thousand transistors)
  - can accurately estimate (within a few %) dynamic and leakage power dissipation

Transistor Level Simulation

- Perform logic simulation to obtain the switching events for each net
  - logic description in structural VHDL or Verilog
  - zero-delay or unit-delay timing models
- Determine frequency of each net
- Pre-layout must estimate capacity