Energy-Aware Computing

UG4/MSc

Lecture 1: Introduction & Overview

Why a new course?

- Power/energy consumption is a first-class problem for computer systems
  - Limits speed for high-perf computers
  - Limits battery life-time for mobile devices
  - Bad for the environment
  - Heat causes reliability issues
- Opens up challenges and opportunities

Learning outcomes

- Describe and discuss the factors which contribute to the consumption of power/ energy in computing systems and how they affect the system performance
- Explain in detail mechanisms found in modern computing systems for conserving energy
- Discuss, assess and compare the behaviour and performance of energy-saving techniques on computing micro-architectures

Learning outcomes

- Gain familiarity with state-of-the-art tools such as processor simulators, memory models and use them to implement and evaluate techniques described in the technical literature
- Locate, summarise and discuss critically peer-reviewed literature on a specific sub-area of energy-aware computing
- Write and present clear and concise descriptions of complex systems/methods

Pre-requisites

- ugrad computer-architecture course
  - Superscalar processors, caches, …
- ugrad computer-design (or similar) is useful but not required
- C programming
  - Tools used in coursework are in C
  - A good Java programmer should be able to cope easily

Assessment

- Coursework – 50%
  - One “mini-project”, 2-part submission
    - part 1, 5% introduction to tools
    - part 2, 45/35% is the bulk of the work
  - Critical review of a research paper (MSc students only) 10%
- Exam – 50%
  - In April/May 2011

CW1-Project

- Group-work: 2 students
  - 1st part individual
- Select from a list of available projects
- Implement and evaluate a known energy/power saving technique using a widely-used, research simulator
- Demonstrate your work at the end
  - Not directly assessed, but compulsory
- 6+ week duration
  - Impossible to do in just the last week!
  - Understanding the simulator code will take some time; start early!

Reading and resources

- Research papers will be made available during the course
- S. Kaxiras, M. Martonosi, Computer Architecture Techniques for Power-Efficiency, Synthesis Lectures on Computer Architecture. Morgan&Claypool publishers.
  - Free to download from University machines
- Hot Leakage/Wattch/Simplescalar, Cacti
  - Commonly used simulator(s)/tools by researchers in this field.
  - SPEC benchmarks/traces

Practicalities

- Lectures
  - Tuesdays, Fridays 2-3pm @ FH 1.B09
  - “Surgery” sessions at comp. lab if needed
- Web page
  - www.inf.ed.ac.uk/teaching/courses/eac
- Help
  - Use email for now. There will be a newsgroup/ web-forum soon.

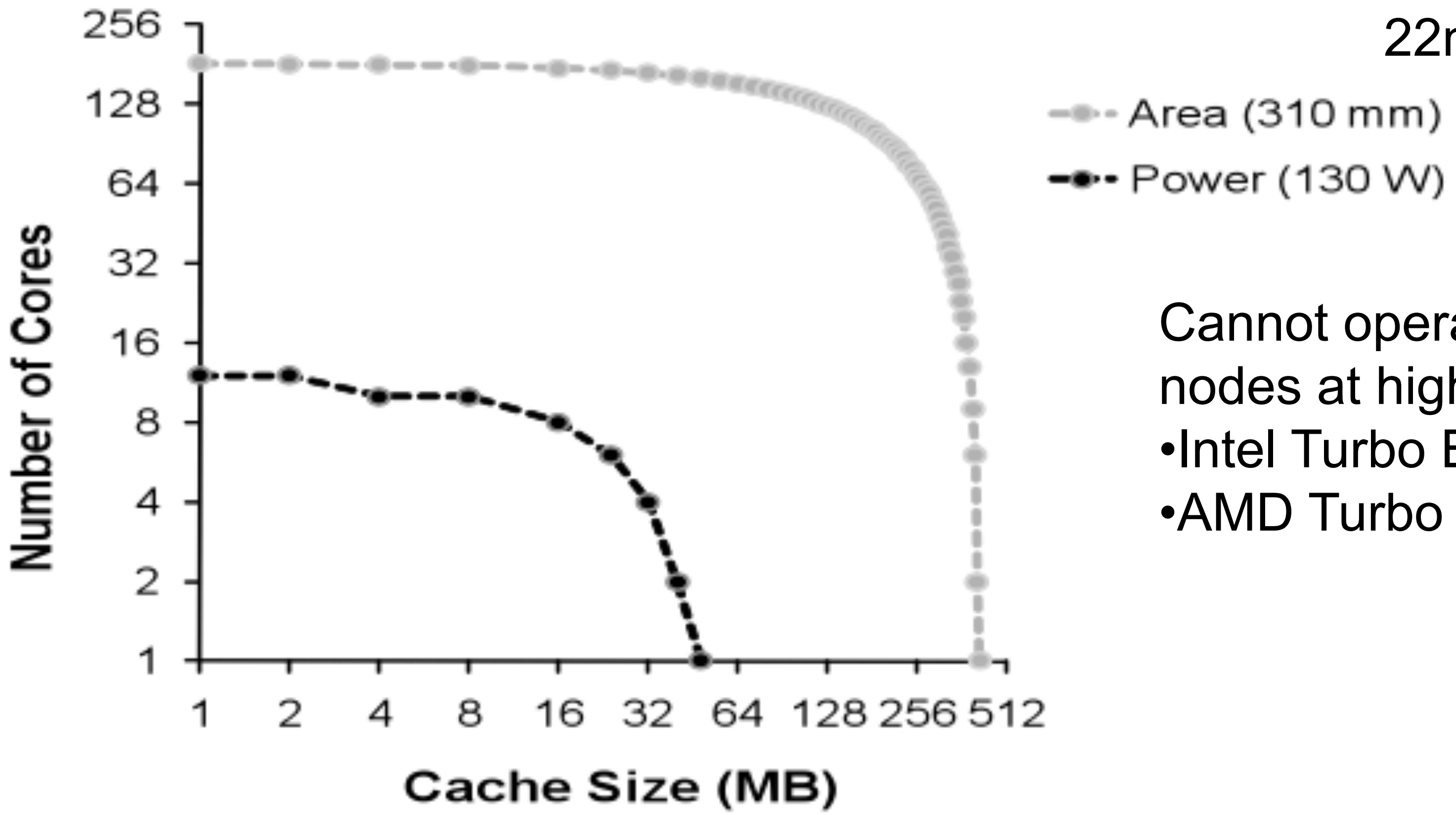
Topics

- CMOS technology basics and sources of power consumption
- Modelling and simulation
- Gate-level techniques
- Micro-architecture techniques
- Memory/cache
- Leakage reduction techniques
- Power management
- Software techniques

Why power matters?

- Limits scaling/integration
- Cooling
  - Chip packaging
  - Data centre room design
- Power delivery cost
- Battery lifetime and size
- System reliability
- Environmental concerns

Power limits tech scaling

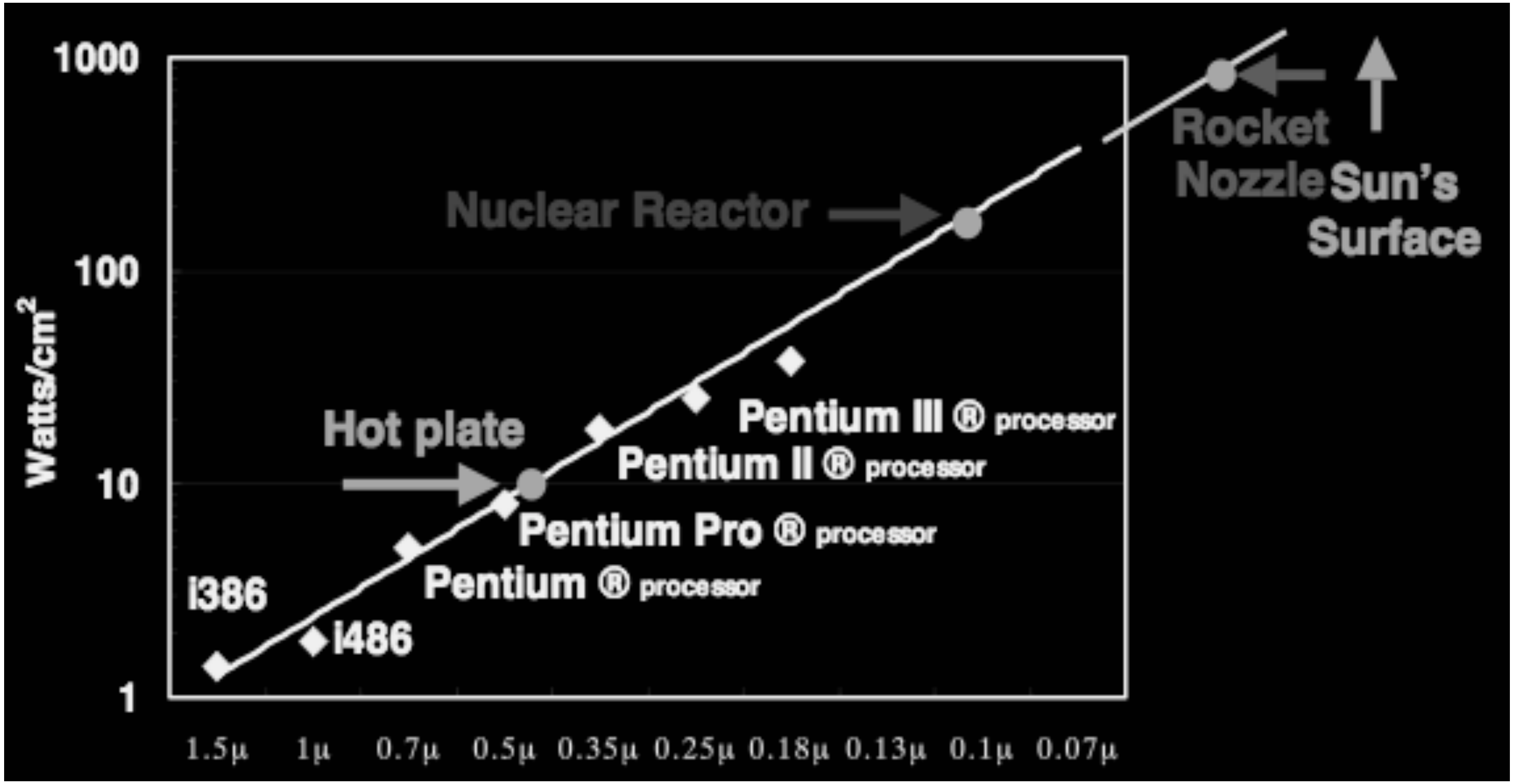


22nm CMOS

Cannot operate all nodes at high speed
•Intel Turbo Boost,
•AMD Turbo CORE

Source: Babak Falsafi: Milliwatt Chips: The Viable Scalability Path for Datacenters

Processors are getting hotter



Fred Pollack, Micro-32 keynote

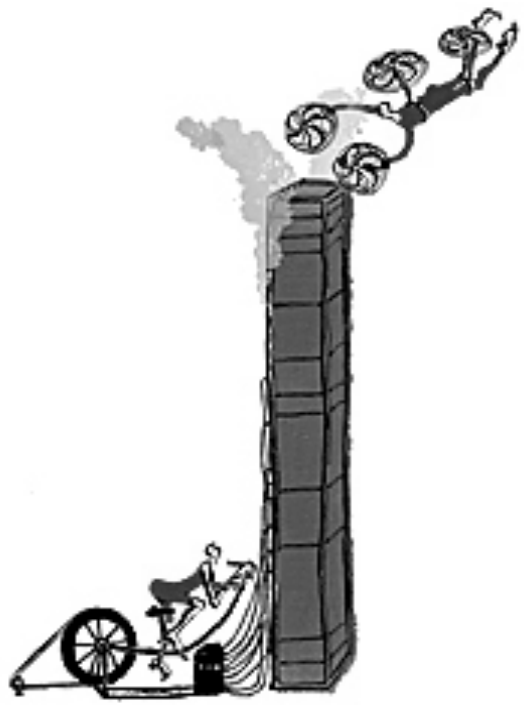
Chip packaging

- Heat needs to be transferred away, or the chip dies
  - For every 10 degree Celcius increase in temperature, the lifetime of a chip reduces by half!
  - Solutions exist (e.g. liquid cooling) but are expensive
- Fans consume power too!
- Handheld devices cannot use fans, not even hit-sinks.
  - Need to dissipate less than 3W

Data centres

- Struggle to keep up with the power requirements of new machines.

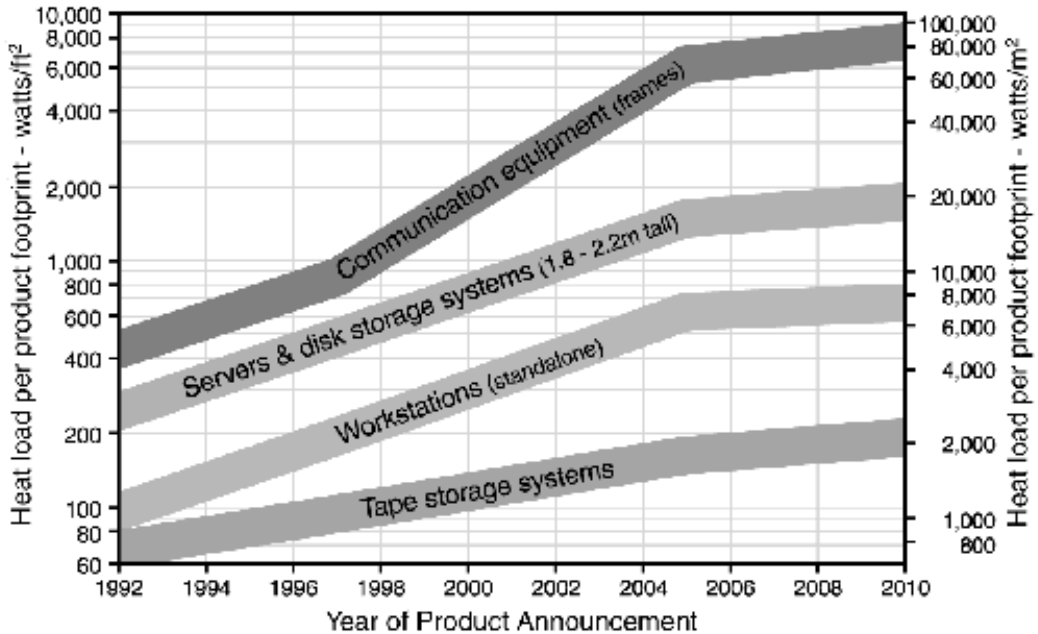
“What matters most to the computer designers at Google is not speed but power - low power, because data centers can consume as much electricity as a city” Eric Schmidt, Google CEO



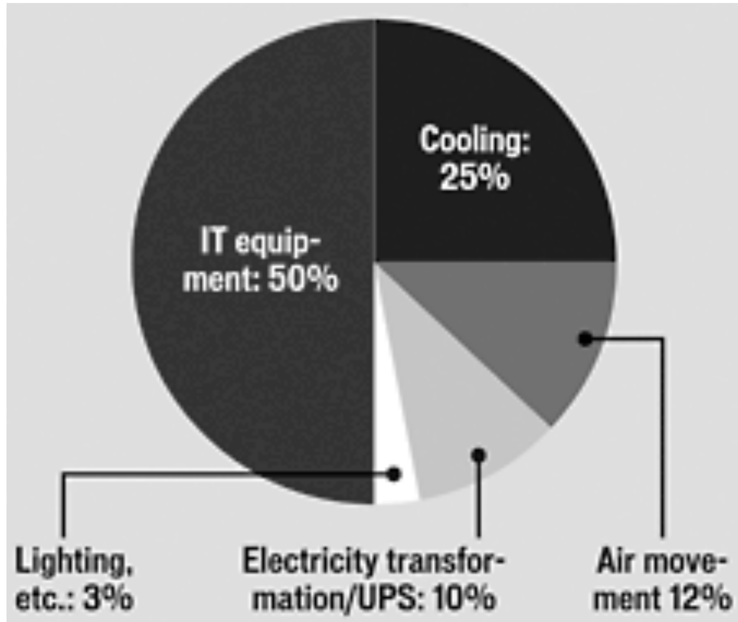
Credit: Belle Mellor

Power Struggle: How IT managers cope with the data center power demands,Robert Mitchell Computer World, April 2006

Data centres



The Uptime institute, 2000



Source: EYP Mission Critical Facilities Inc.

Power delivery system

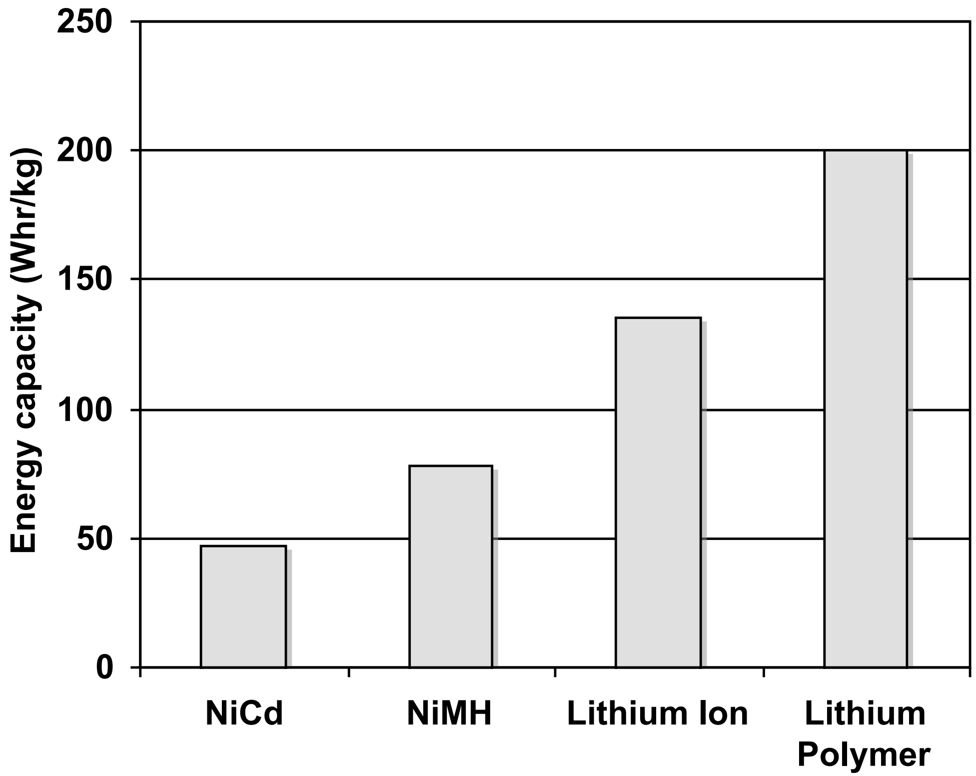
- The subsystem that delivers power to the chip but also the on-chip delivery system
- Increased current through PDS
  - Operating voltages decrease
  - More transistors on chip
- Problems
  - IR drop - variation in voltage at point of delivery
  - Electromigration - reliability issue
- More complex PDS
  - High cost
  - High design/verification effort

Batteries

- Battery capacity is not improving fast

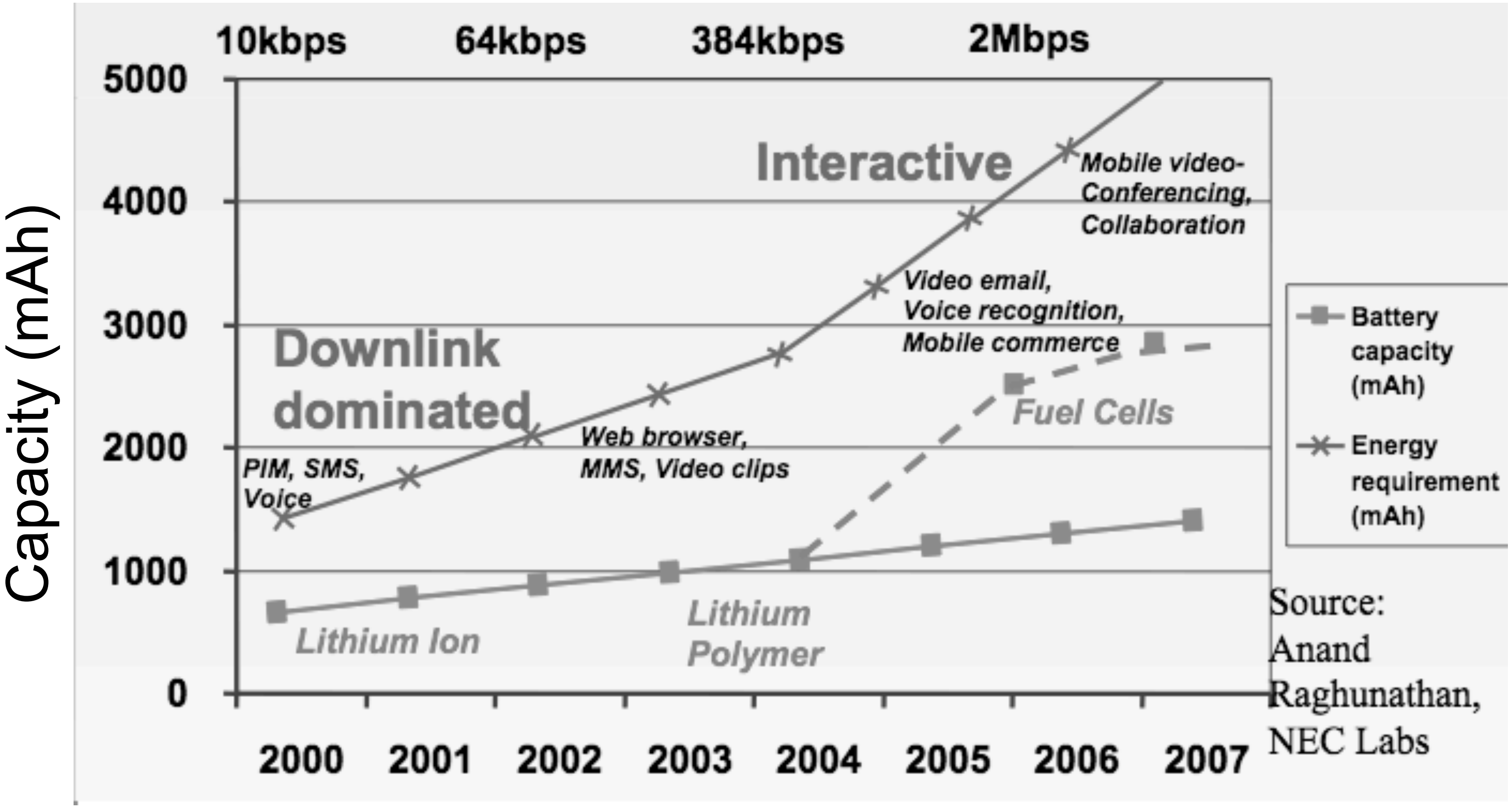
- Limits the functionality of portable devices

- Forces manufacturers to make feature vs attractive design trade-offs



Batteries

Gap between energy needs of applications and battery capacities



What can we do?

- Understand where/when power is dissipated
- Find ways of reducing it at all levels of design (circuits, architecture, OS, applications software)

Next time

- CMOS technology basics
- Power, energy in CMOS
- Metrics combining power and speed