Introduction to Green Computing

Slides by: Jordi Cucurull Juan

Department of Computer and Information Science (IDA)
Linköping University
Sweden

Outline

- Sustainability
- Power aware computing
- Data centres

Sustainability

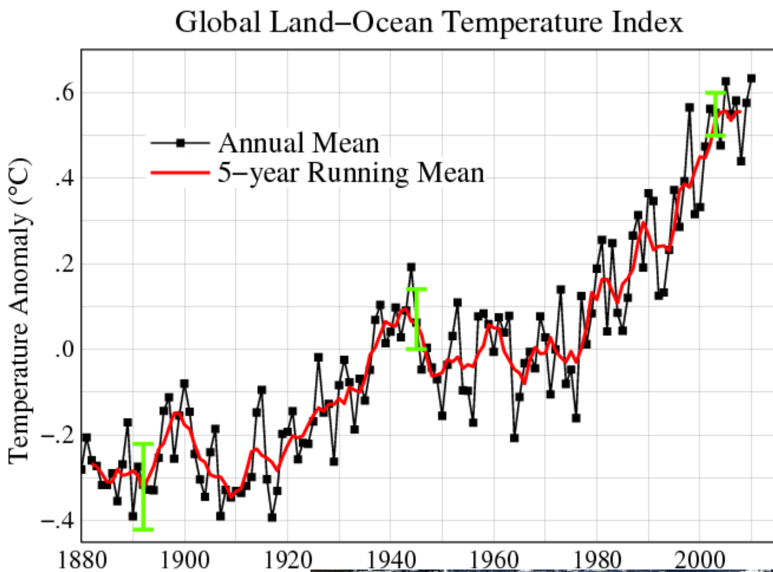
Sustainability

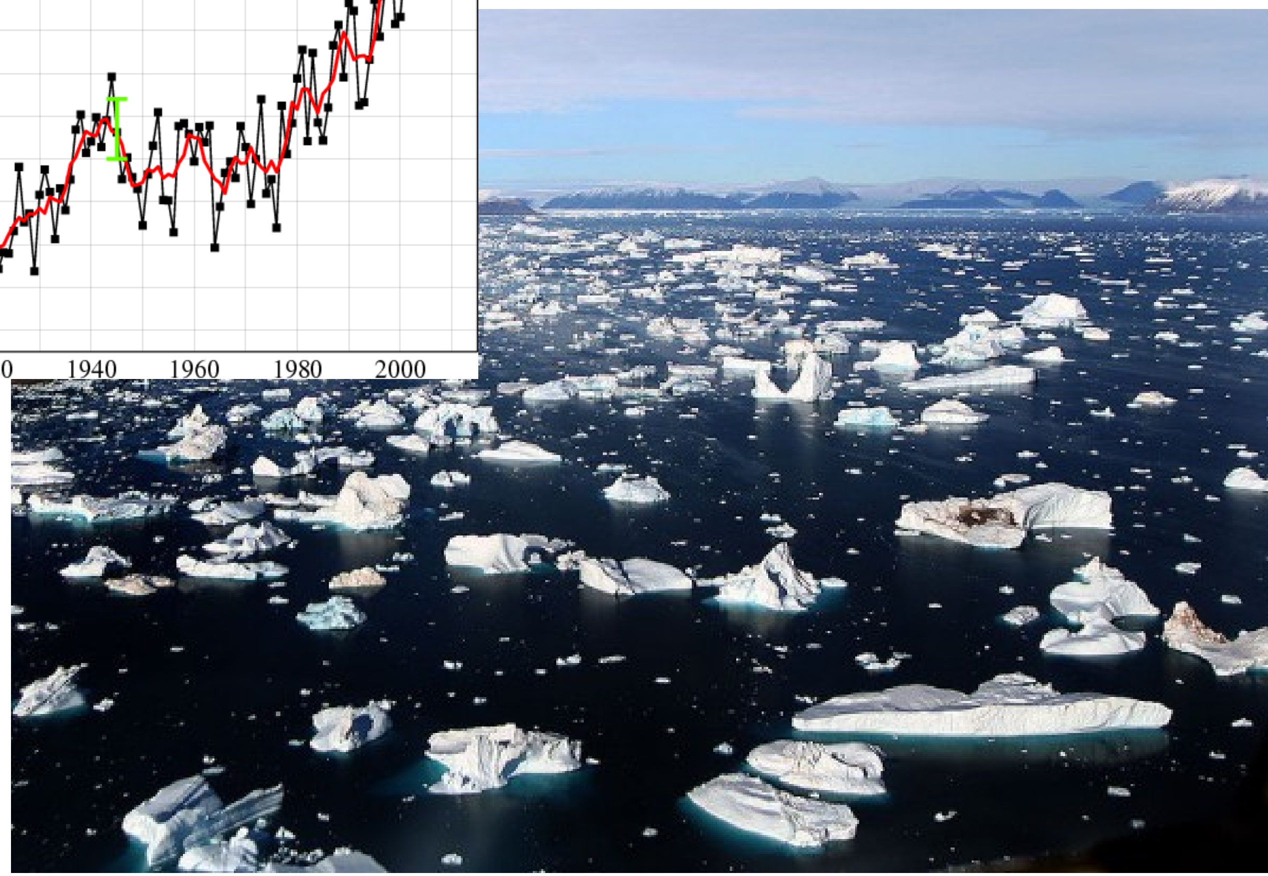
Introduction

Introduction

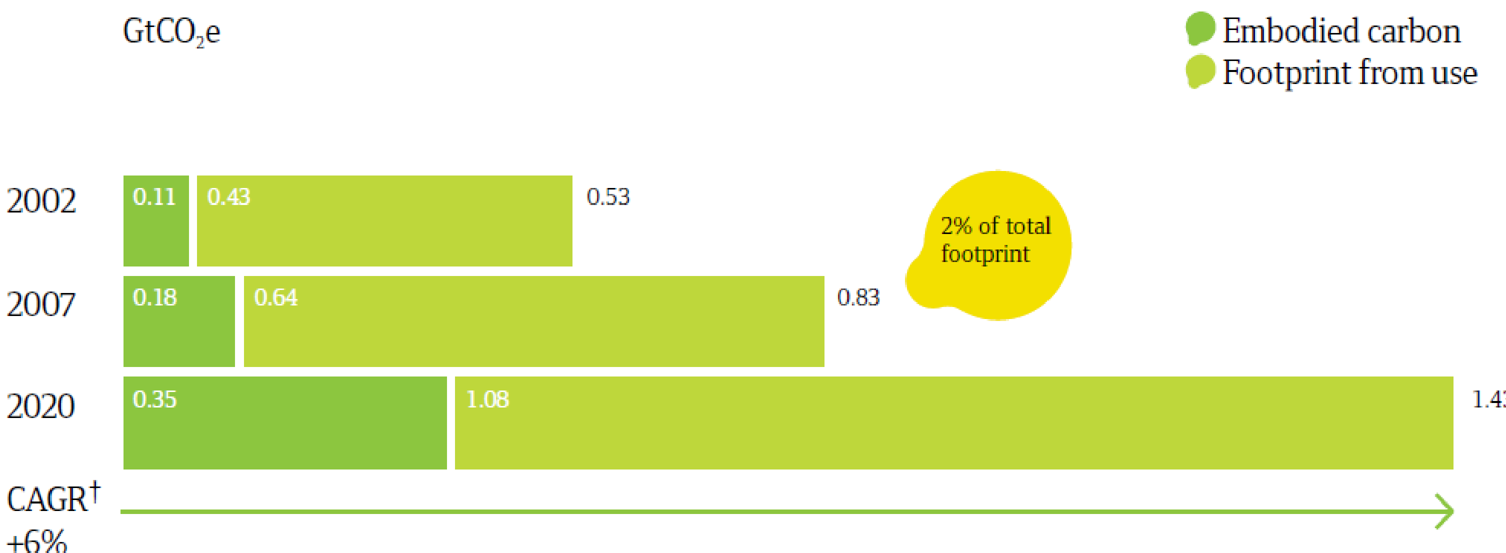
- Human activities are affecting the natural environment
- An example is the massive production of CO_2

Global Warming





Global ICT footprint (CO_2)



*ICT includes PCs, telecoms networks and devices, printers and data centres.
†Compounded annual growth rate.

http://www.smart2020.org/publications/

Green IT

Green IT is "...the study and practise of designing, manufacturing, using, and disposing computers, servers, and associated subsystems efficiently and effectively with minimal or no impact to the environment."

Green IT

Sustainable ICT

Sustainable IT

Green Computing

Murugesan 2008.

S. Murugesan, "Harnessing Green IT: Principles and Practices," IT Professional, vol.10, no.1, pp.24-33, Jan.-Feb. 2008

Focus areas

- Design for environmental sustainability
- Responsible disposal and recycling
- Use of renewable energy sources
- Regulatory compliance
- Eco-labeling of IT products
- Energy-efficient computing
- Power management
- Data center design, layout, and location
- Server virtualisation

Benefits

- Environmental
  - ICT respectful towards the environment
  - Less production of CO_2 and other contaminants
- Economical
  - Reduction of electricity bill
  - Less infrastructure same service (power supplies, cooling systems...)
  - Government financial incentives
- Public relations
  - Marketing
  - Competitiveness

Sustainability affects companies image

- Carbon Disclosure Project
  - Database with corporate climate change information
  - Companies disclose
    - Gas emissions
    - Strategies to prevent climate change

https://www.cdproject.net

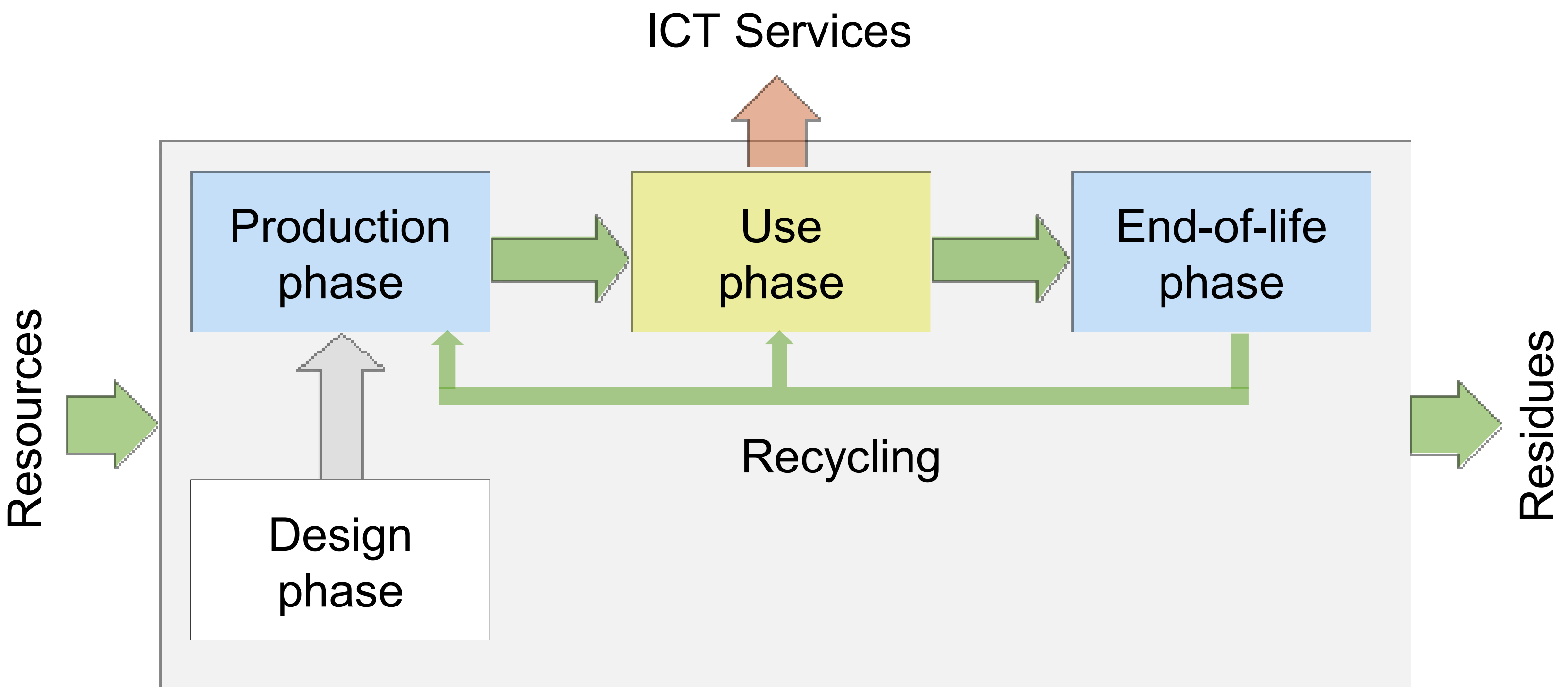
- Cool IT (Green Peace)
  - Initiative to track corporate climate leadership
  - Most influential IT brands are analysed according to:
    - Efforts to provide solutions to reduce greenhouse emissions
    - Initiatives to reduce their own emissions
    - Political advocacy to support climate and energy policies

http://www.greenpeace.org/international/en/campaigns/climate-change/cool-it/

Sustainability

Life-cycle of ICT

Holistic approach: Life-Cycle of ICT



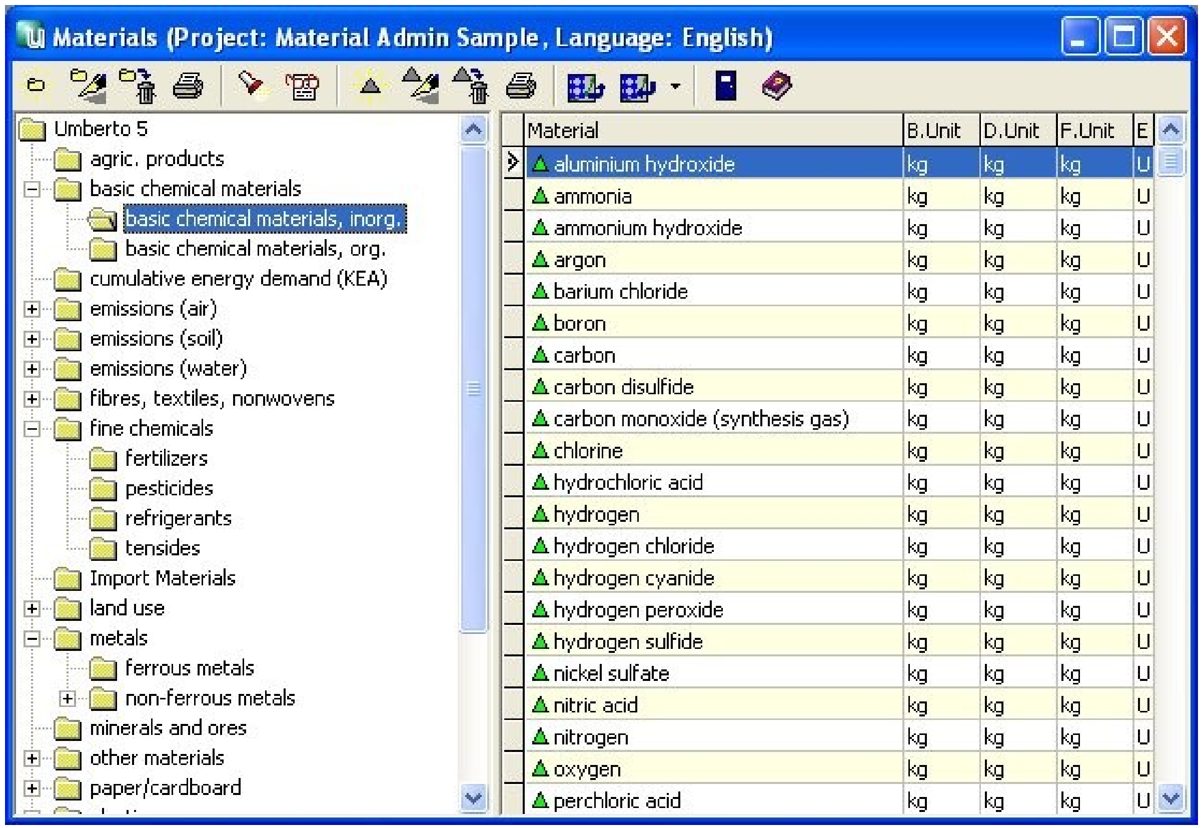
Lorenz M. Hilty. Information Technology and Sustainability: Essays on the Relationship between Information Technology and Sustainable Development, Books on Demand, 2008 ISBN: 978-3837019704

Life-Cycle Assessment (LCA)

- All phases of the chain must be considered
  - Consume material and energy
  - Use some infrastructure
- Life-Cycle Assessment (LCA)
  - Environmental impact of a product in entire life-cycle
  - Materials, energy and infrastructure have their own life-cycles
  - Recursive study of them
- LCA analysis performed with
  - LCA tools: *SimaPro*, *Umberto*...
  - Life-cycle inventory database: *Ecoinvent*...

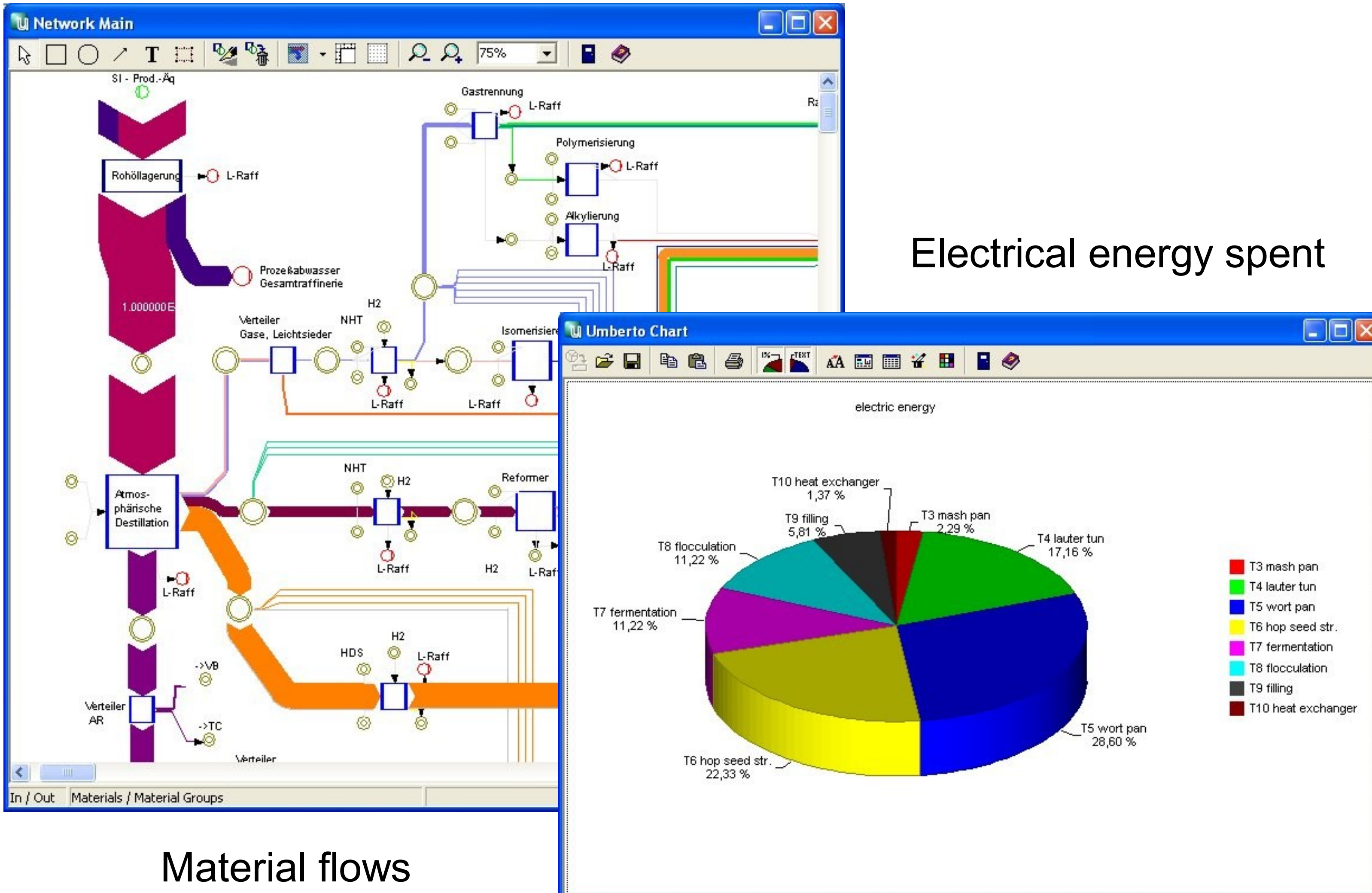
Umberto

Management of materials



http://www.umberto.de

Umberto



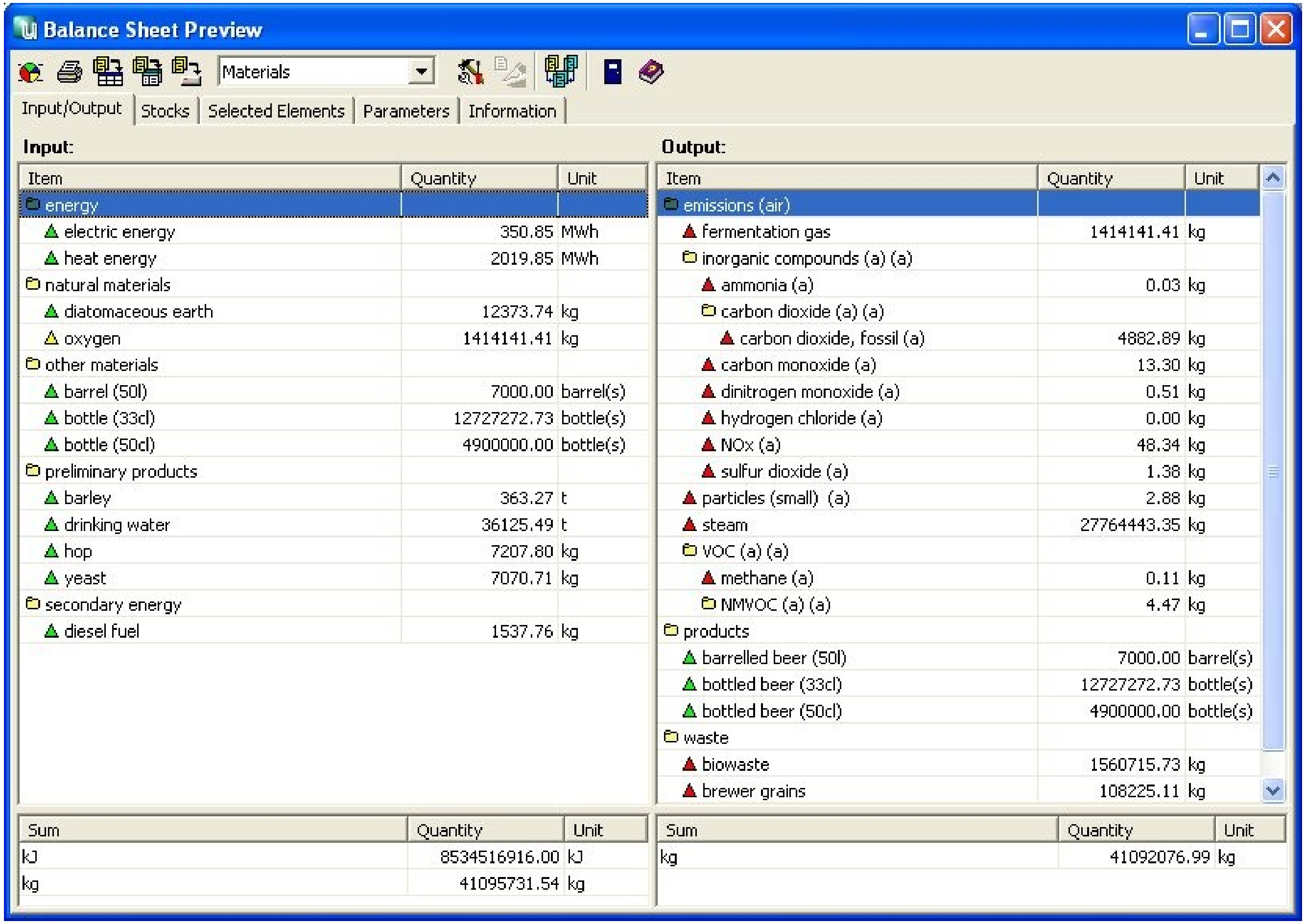
Electrical energy spent

Material flows

http://www.umberto.de

Umberto

Materials from flow network calculation



Input:

Item	Quantity	Unit
energy		
electric energy	350.85	MWh
heat energy	2019.85	MWh
natural materials		
diatomaceous earth	12373.74	kg
oxygen	1414141.41	kg
other materials		
barrel (50l)	7000.00	barrel(s)
bottle (33cl)	12727272.73	bottle(s)
bottle (50cl)	4900000.00	bottle(s)
preliminary products		
barley	363.27	t
drinking water	36125.49	t
hop	7207.80	kg
yeast	7070.71	kg
secondary energy		
diesel fuel	1537.76	kg

Sum	Quantity	Unit
kJ	8534516916.00	kJ
kg	41095731.54	kg

Output:

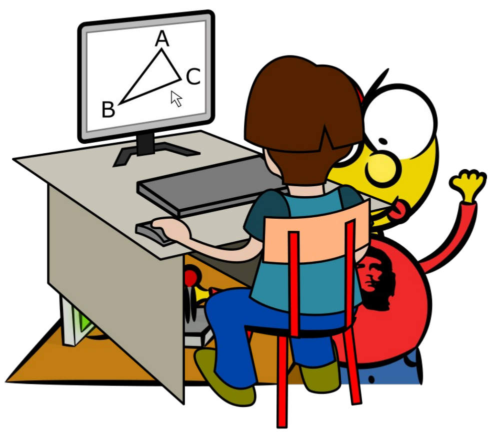
Item	Quantity	Unit
emissions (air)		
fermentation gas	1414141.41	kg
inorganic compounds (a) (a)		
ammonia (a)	0.03	kg
carbon dioxide (a) (a)		
carbon dioxide, fossil (a)	4882.89	kg
carbon monoxide (a)	13.30	kg
dinitrogen monoxide (a)	0.51	kg
hydrogen chloride (a)	0.00	kg
NOx (a)	48.34	kg
sulfur dioxide (a)	1.38	kg
particles (small) (a)	2.88	kg
steam	27764443.35	kg
VOC (a) (a)		
methane (a)	0.11	kg
NMVOC (a) (a)	4.47	kg
products		
barrelled beer (50l)	7000.00	barrel(s)
bottled beer (33cl)	12727272.73	bottle(s)
bottled beer (50cl)	4900000.00	bottle(s)
waste		
biowaste	1560715.73	kg
brewer grains	108225.11	kg

Sum	Quantity	Unit
kg	41092076.99	kg

ICT production *vs.* usage

- Comparison of phases in Desktop PCs

Production in China
consumes 2.4 GJ



One year usage
consumes 0.8 GJ

ICT End-of-life

- Waste Electrical and Electronic Equipment (e-Waste)
- e-Waste has become a serious problem
  - Total annual global volume around 40 million tonnes
  - Treatment is a challenge, recycling is the key
    - Recycling metals can save up 20-25% production costs

ICT End-of-life

- Informal recycling
  - Informal industry in emerging economies
  - Health and environmental impacts not considered
    - None or poor safety measures used for manipulation
    - High levels of contaminants in the activity areas
      - Air, bottom ash, dust, soil, waters...

Pictures by courtesy of Technology and Society Lab, Empa Materials Science and Technology, Switzerland

Ecolabelling and standards

Ecolabelling, standards and specifications

- Sustainability of products depends on several factors
  - Energy consumption
  - Materials used for construction
  - Disposal of the products

- Ecolabelling facilitates product selection
  - Easy identification of sustainable products
  - Easy comparison of sustainable products

- Standards, directives and product databases
  - ENERGY STAR
  - RoHS
  - EPEAT and IEEE P1680

ENERGY STAR

- Set of **energy** performance specifications that qualified products must fulfill
  - Several categories: Computers, Servers, Battery chargers...
- ENERGY STAR Program Requirements for Computers 5.0
  - Defines categories of computers
    - Desktop computers, integrated desktop computers, notebooks, workstations, game consoles, small-scale servers, thin clients
  - Defines operational modes
    - Off mode, sleep mode, idle mode, active mode
  - Defines maximum annual consumption for each category
    - Tables with operational mode weighting (% time idle, sleep...)
    - Tables with maximum annual energy according to operational mode weighting defined (Typical Energy Consumption - TEC)
  - Defines test procedures to qualify products

http://www.energystar.gov

ENERGY STAR

- Example with desktop computer category B (2 cores, 2 GB)

Operational mode	Percentage of time
Toff	55%
Tsleep	5%
Tidle	40%

$$E_{TEC} = (8760/1000) \cdot (P_{off} \cdot T_{off} + P_{sleep} \cdot T_{sleep} + P_{idle} \cdot T_{idle})$$

E_{TEC} for category B computer ≤ 175.0 (kWh)

P_{off}, P_{sleep}, and P_{idle} must be measured and fed into the formula that must give less or equal than 175 kWh to qualify the product

http://www.energystar.gov

ENERGY STAR

- ENERGY STAR Program Requirements for Displays 5.0
  - Defines criteria for qualifying products
    - Power source, television tuners, automatic brightness control, external power supply, power management requirements
  - Defines operational mode requirements
    - Maximum consumption in off and sleep modes
    - Maximum consumption in on mode
      - Depends on size, resolution and brightness settings
  - Defines test procedures to qualify products
- ENERGY STAR provides a database for qualified products
  - http://www.eu-energystar.org/en/database.htm

RoHS - EU Directive 2002/95/EC

- Restriction of use of certain Hazardous Substances (RoHS)
  - In electrical and electronic equipment
  - To protect human health and environment
  - For products put on the market since 1st July of 2006

- Restricted substances
  - Lead
  - Mercury
  - Cadmium
  - Hexavalent chromium
  - Polybrominated biphenyls (PBB)
  - Polybrominated diphenyl ethers (PBDE)

http://www.rohs.gov.uk/
http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:02002L0095-20100925:EN:NOT

EPEAT / IEEE P1680

- Global registry of electronic products
  - Covers design, production, use, and disposal of products
  - Operation and criteria based on IEEE 1680 standards
  - 23 required criteria and 28 optional
- Products registered and declared by manufacturers
  - Independent verification of their claims
  - Fast product presence in the register
- Environmental product ranking
  - Bronze: Meets all 23 required criteria
  - Silver: Bronze plus 50% of the optional criteria
  - Gold: Bronze plus 75% of the optional criteria

http://www.epeat.net

EPEAT / IEEE P1680

- Criteria categories
  - Reduction/elimination of environmentally sensitive materials
  - Materials selection
  - Design for end of life
  - Product longevity / life cycle extension
  - Energy conservation
  - End of life management
  - Corporate performance
  - Packaging

- EPEAT enforces products to meet ENERGY STAR requirements RoHS EU directive

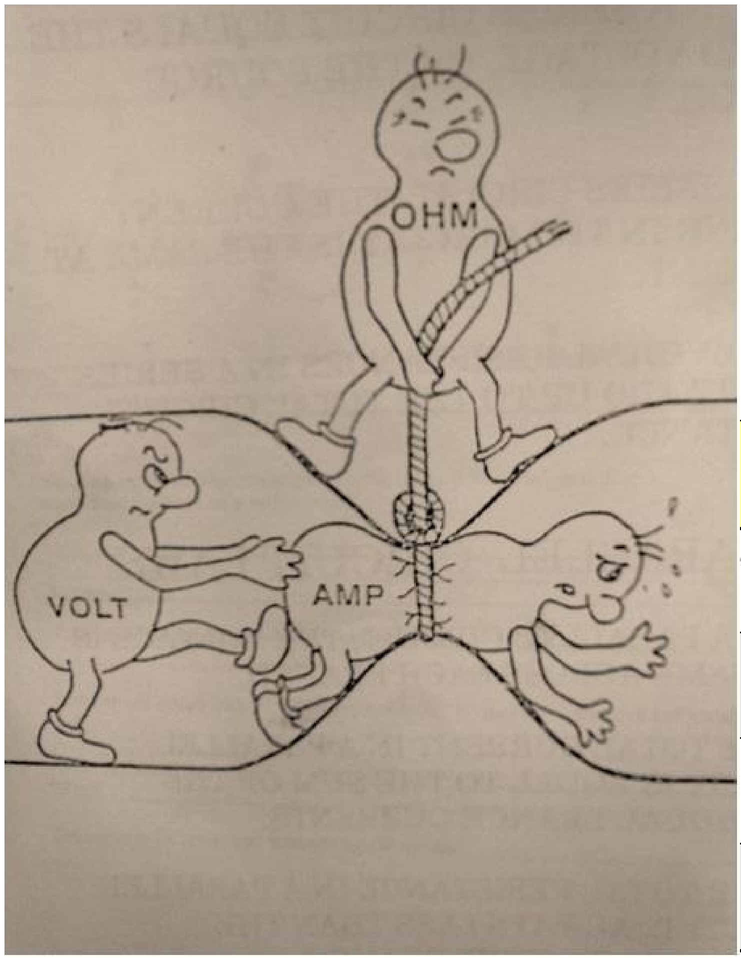
http://www.epeat.net/Docs/Summary%20of%20EPEAT%20Criteria.pdf

Power aware computing

Power aware computing

Electrical backgrounds

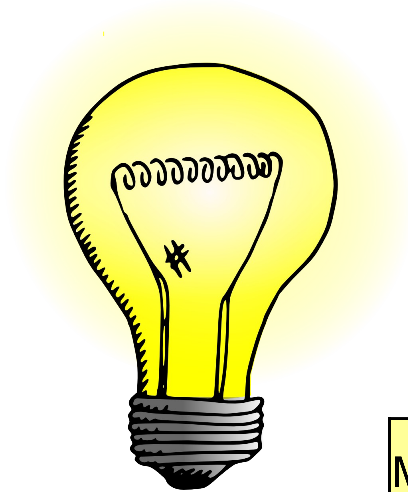
Electrical units



Unit	Description
Volt	Difference of charges between two points
Ampere	Electric charges movement rate
Ohms	Circuit opposition to current flow
Farad	Storage of energy as an electric field

S. Gibilisco, "Teach Yourself Electricity and Electronics", 4rt edition, McGraw-Hill Professional Publishing , 2006

Energy metrics



Electrical energy is transformed in heat!

Power = Voltage * Intensity

Metric	Unit	Description
Power	*Watt*	Rate of energy expenditure
Energy	*Joule*	Power dissipated over a length of time

1 Joule = 1 Watt * second = 1 Ws
3600 Joules = 1 Watt * hour = 1 Wh

Power aware computing

Introduction

Power aware computing

Power aware computing are computational techniques that consider energy consumption as one of their main constraints

Avoid wasted energy!!!

- Challenges
  - Figure out *where* and *why* waste happens
  - Determine *how* to avoid it

P. Ranganathan, "Recipe for Efficiency: Principles of Power-Aware Computing", Communications of the ACM, vol.53, no.4, pp.60-67, April. 2010

Sources of energy waste

- General-purpose systems tendency

  - Good performance for several applications
  - Union of maximum requirements of each application class

Sources of energy waste

- Optimisation for peak performance scenario

  - Average system utilisation low
  - Benchmarks stress worst-case performance workloads
    - Systems optimised for these scenarios

Sources of energy waste

- Components designed by different teams
  - No component interaction considered
- Functions of the system as independent modules
  - No function interaction considered
- Design focused on performance and availability
  - Resource waste for small improvements
  - Component or operation redundancies

Reduction of energy waste

- Common solutions
  - Replacement for more power-efficient alternative
  - Disable unused resources
  - Match work to most appropriate resources
  - Combination of multiple tasks in single energy event
  - Design for required functionality

Reduction of energy waste

- Coming solutions
  - Holistic solutions
    - Broad scope of the problem
    - Cross-layer interaction
  - Trade off some other metric for energy
  - Optimise energy efficiency for the common case
  - Spend someone else's power
  - Spend power to save power

Power aware computing

Power management

Power management

- Different levels
  - Circuits
  - **Architecture**
  - **Compilers and Systems**

- Applied to different hardware

Matthew Garret. Powering Down, Communications of the ACM 51, 9 (September 2008), 42-46

Processors

Processor

- Does not run at 100% capacity all time

- Architecture techniques to save up energy
  - CPU frequency/voltage scaling
  - Low power mode states
    - Disable functional units not needed
      - Clock gating
      - Dissociate from memory bus
      - Disable part of the cache

- Compiler and systems techniques to save up energy
  - Tickless kernel
  - Timer consolidation

Power consumption of CMOS

- CMOS integrated circuit technology
  - Theoretically only consumes energy on state changes
- CMOS power consumption
  - Dynamic power
    - Dominant power category
    - Depends mostly on
      - Voltage and Frequency
      - Capacitance and Activity Factor
  - Leakage
    - Not associated to processor activity
    - Accounts for 20-40% consumption
    - Increasing in importance as scale of integration increases

S. Kaxiras and M. Martonosi, "Computer Architecture Techniques for Power-Efficiency", Synthesis Lectures on Comptuter Architecture, Morgan & Claypool Publishers, 2008

CPU frequency/voltage scaling

- CMOS basic dynamic power equation:

$$P = C \cdot V^2 \cdot A \cdot f$$

C	Load capacitance
V	Supply voltage
A	Activity factor
f	Operating frequency

  - Voltage reduction decreases power by the square of it
    - Maximum frequency is limited by voltage
  - Potential cubic reduction in power dissipation
    - Considering *f* and *V*
  - Performance decreases linearly

Dynamic voltage/frequency scaling (DVFS)

- Dynamic adjustment of voltage/frequency
  - Tradeoff power dissipation / performance

- DVFS decision level
  - System level (OS)
    - Idleness of the system drives decision
    - Voltage/frequency scaled to eliminate idle periods
  - Program level
    - Program behaviour drives decision
    - E.g. scale down when program knows that has to wait
  - Hardware level
    - Exploits different timings of hardware components

Dynamic voltage/frequency scaling (DVFS)

- Examples of commercial systems
  - Intel SpeedStep
  - AMD PowerNow!
- Decision taken at system level
- Changes through specific CPU register

Enhanced Intel® SpeedStep® Technology for the Intel® Pentium® M Processor (White Paper)
http://download.intel.com/design/network/papers/30117401.pdf

Low power mode states

- CMOS basic dynamic power equation:

$$P = C \cdot V^2 \cdot A \cdot f$$

- Capacitance and Switching activity intertwined
  - Capacitance (*C*)
    - Fixed at design time
    - Dependant on
      - number of transistors
      - interconnections
  - Switching activity (*A*)
    - Fraction between 0 and 1
    - Factor of capacitance charged/discharged each CPU cycle
    - Dependant on transistors and interconnections enabled

Management at system level

- HLT (halt) instruction
  - Allows to indicate that there is nothing to execute
  - CPU enters halt state until next *interrupt*
  - Issued by the operating system

Signal that stops the current running task to handle some situation

- Advanced Power Management (APM)
  - CPU idle / busy calls
    - CPU in low / normal power state
    - Low power state
      - Clock stopped until next *interrupt*
      - Clock slowed down

- Advanced Configuration and Power Interface (ACPI)
  - Current specification for energy management
  - Richer low power modes and frequency/voltage scaling

History

Transition to low power states

- Power state transitions take time

> Processor must remain in idle power state more than 20 ms for getting benefit of it!!!

- Interrupts may wake up the processor too often
  - Some interrupts cannot be avoided
    - Interrupts for user interaction, e.g. keyboard
  - But other interrupts can be adjusted or disabled
    - Periodic interrupts such as timers

Timers

- Events scheduled at a specific time in the future
  - Example: cursor blinking, time clock ticking...
  - The event produces a timer interrupt
- Timer interrupts have a big impact in consumption
  - Regularly wake up the processor
  - System has plenty of interrupts
- Two examples of optimisation
  - Linux tickless kernel
  - Consolidation of timers

PowerTOP

```
    PowerTOP version 1.10      (C) 2007 Intel Corporation

Cn                Avg residency       P-states (frequencies)
C0 (cpu running)        (20.2%)         3.07 Ghz     9.3%
polling           0.0ms ( 0.0%)         2.14 Ghz    20.1%
C1                0.0ms ( 0.0%)         1.60 Ghz    44.6%
C2                0.1ms ( 1.3%)          800 Mhz    24.7%
C6                1.1ms (78.5%)

Wakeups-from-idle per second : 823.8    interval: 10.0s
no ACPI power usage estimate available

Top causes for wakeups:
  28.2% (365.2)       firefox-bin : futex_wait (hrtimer_wakeup)
  18.9% (244.5)      <kernel IPI> : Rescheduling interrupts
  18.9% (244.3)       <interrupt> : extra timer interrupt
   7.9% (102.6)   plugin-containe : schedule_hrtimeout_range (hrtimer_wakeup)
   6.2% ( 80.0)          acroread : schedule_hrtimeout_range (hrtimer_wakeup)
   4.7% ( 61.3)   USB device 1-4.2 : USB Laser Mouse (Logitech)
   4.1% ( 53.7)       <interrupt> : ehci_hcd:usb1, uhci_hcd:usb5, uhci_hcd:usb8
   3.6% ( 47.0)       <interrupt> : uhci_hcd:usb4, uhci_hcd:usb7, HDA Intel
   2.0% ( 26.2)       soffice.bin : schedule_hrtimeout_range (hrtimer_wakeup)
   1.8% ( 23.2)       firefox-bin : schedule_hrtimeout_range (hrtimer_wakeup)
   0.6% (  8.2)       <interrupt> : ahci
   0.4% (  5.1)       <interrupt> : eth0
   0.4% (  5.0)   plugin-containe : futex_wait (hrtimer_wakeup)
   0.3% (  4.4)     <kernel core> : ehci_work (ehci_watchdog)
   0.3% (  4.0)      work_for_cpu : usb_hcd_poll_rh_status (rh_timer_func)
   0.3% (  3.7)   thunderbird-bin : futex_wait (hrtimer_wakeup)

Suggestion: Enable USB autosuspend by pressing the U key or adding
usbcore.autosuspend=1 to the kernel command line in the grub config

 Q - Quit   R - Refresh   U - Enable USB suspend
```

Information about what causes CPU wake ups

Hard drives

Traditional hard drive

- Composed of electronic and mechanical parts

Most of solutions exploit reduction of consumption of the mechanical parts

Spin down

- Switch off the platter spindle motor when inactive
  - Supported by most operating systems
- Costs
  - Reduces hard-drive life expectancy
  - Uses a lot of energy to spin up
  - Creates delays (order of seconds)
- Smart management of I/O to
  - minimise spin transitions
  - reduce delays

Solid state drives (SSD)

- Composed only of electronic parts

- No mechanical parts
  - Lower consumption than regular HDs
  - Faster read operations

Networks

Network devices

- Ethernet is the dominant wired communication technology
  - Common supported speeds 10-1000 Mbps
  - Similar energy consumed with and without data transmission
  - Idle mode prevents any kind of reception
  - New standard IEEE 802.3az for low power modes

- Wake on LAN
  - Technique to wake up a slept machine
    - Network keeps physical interface enabled
    - “Magic packet” tells the interface to wake up machine

- Wireless LAN
  - Physical and routing protocols to optimise consumption

Advanced Configuration Power Interface (ACPI)

Advanced Configuration and Power Interface (ACPI)

- Specification
  - Holistic approach for computer system energy conservation
  - Created by HP, Intel, Microsoft, Phoenix and Toshiba
- OS-directed configuration and power management (OSPM)
  - Governs all system and device power state transitions
  - Aware of user preferences and applications requirements
- Defines
  - Hardware/software interfaces and data structures
  - State definitions
  - ASL & ASM languages

Andrew Grover. Modern System Power Management. ACM Queue vol. 1, no. 7 (October 2003), 66-72

Advanced Configuration and Power Interface Specification, Revision 4.0a
http://www.acpi.info

Goals

- Configuration and power management
  - For desktop, laptops, workstation and servers
- Enhanced functionality and robustness at OS-level
  - Inexpensive power managed hardware
  - Better power management decisions
  - Reduction of conflicts between OS and firmware
- Robust interface for configuring motherboard devices
- Promotion of industry-wide adoption

Functional areas

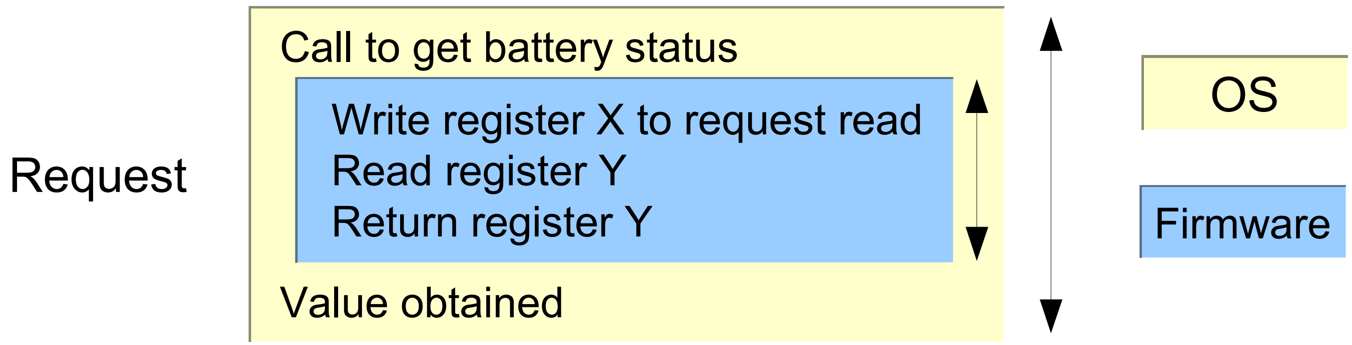
- Power management
  - System, device, and processor
- Performance management
  - Device and processor
- Configuration / Plug & Play
- System events
- Battery management
- Thermal management

Everything controlled by the OSPM!!!

Architecture

Firmware former usage

- Initial power management mechanisms
  - Implemented on device firmware
  - Firmware did most of the work
  - Example:



- Firmware reliance not good
  - A call can take long time or never return
  - OS stability is dependent on firmware quality

Andrew Grover. Modern System Power Management. ACM Queue vol. 1, no. 7 (October 2003), 66-72.

ASL & AML Languages

- ACPI defines two languages
  - ACPI Source Language (ASL)
    - Human readable source code
  - ACPI Machine Language (AML)
    - Interpreted language
    - Describes hardware and steps to reach it

- ACPI defines abstract control methods for devices
  - Example: _BST retrieves battery status

- System firmware provides AML code
  - Include information about the devices
  - Implements control methods

ACPI Firmware usage

System initialisation

Create ACPI Description Tables

Request

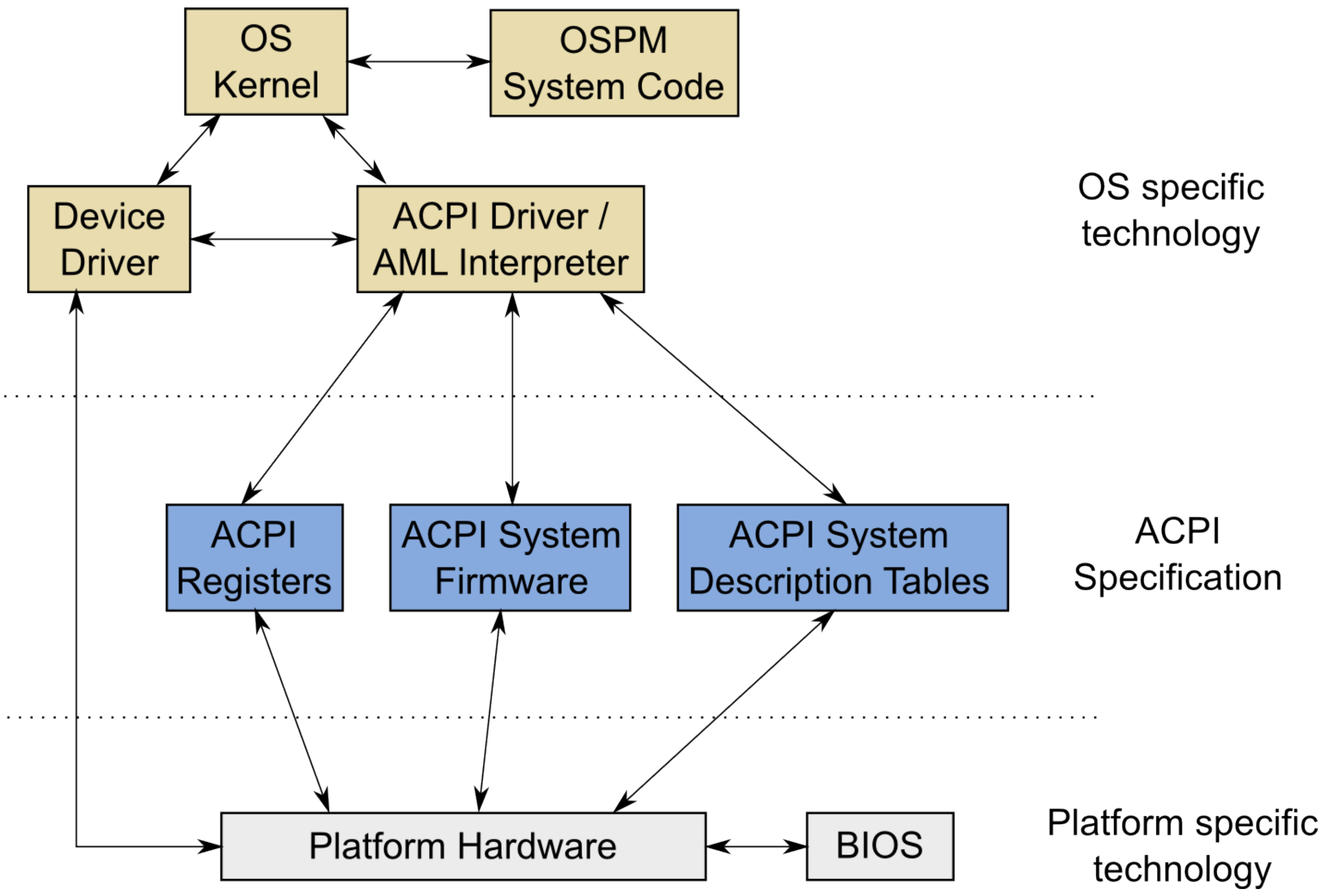
Get battery status
Call _BST:
Write register X to request read
Read register Y
Return register Y

OS

Firmware

- Advantages
  - Less opaque and debugable
  - Can run concurrently and does not block the system
  - Less constrained in size than firmware code

Implementation Structure

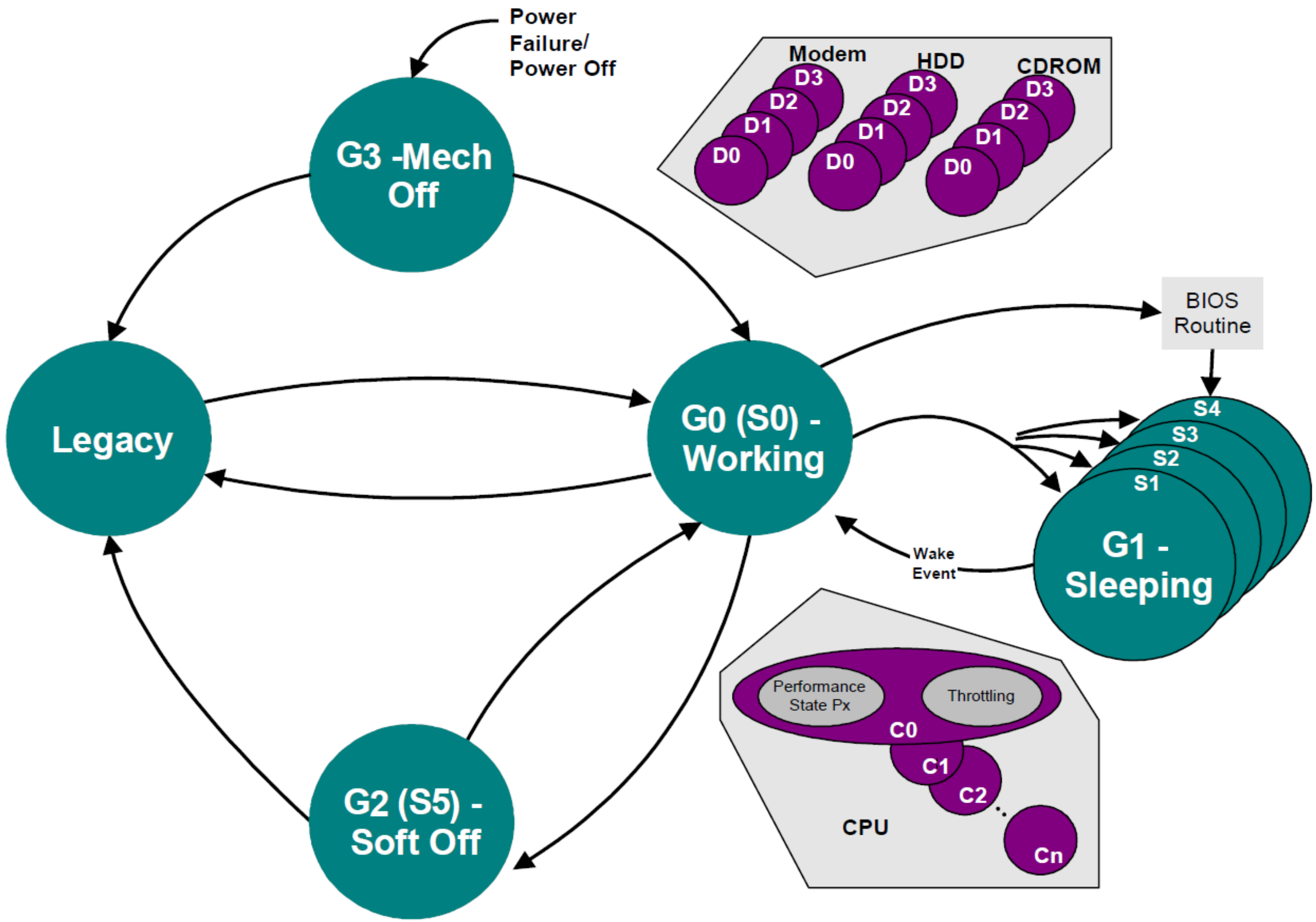


State definitions

State definitions

- ACPI saves energy by switching to low power states
- Different states at different levels
  - Global system state (G)
  - Sleeping state (S)
  - Device power state (D)
  - Processor power state (C)
  - Processor throttling state (T)
  - Device and processor performance states (P)

Main view of power states



Device management

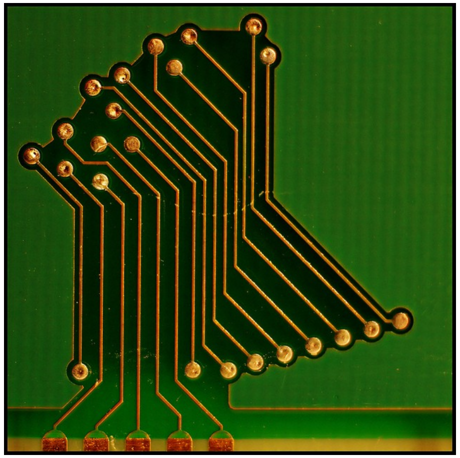
Processor

- Main controls over the processor
  - Power states (C-states)
  - Clock throttling (T-states)
  - Performance states (P-states)

- OSPM must balance
  - Performance
  - Power consumption and battery life
  - Thermal requirements
  - Noise-level requirements

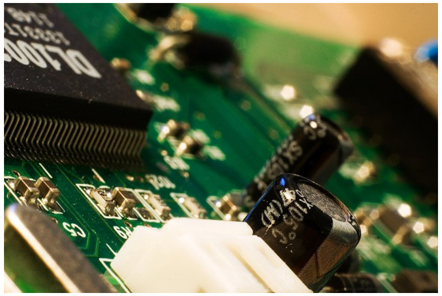
I/O Interconnect

- Standard way to connect a device to the system
  - PCI, PCI Express, USB, IEEE 1394...
- Provides operations to manage power of attached devices
  - Power capabilities not defined by ACPI
  - Baseline power management support OS can use
  - OS can track power state of all devices in I/O interconnect



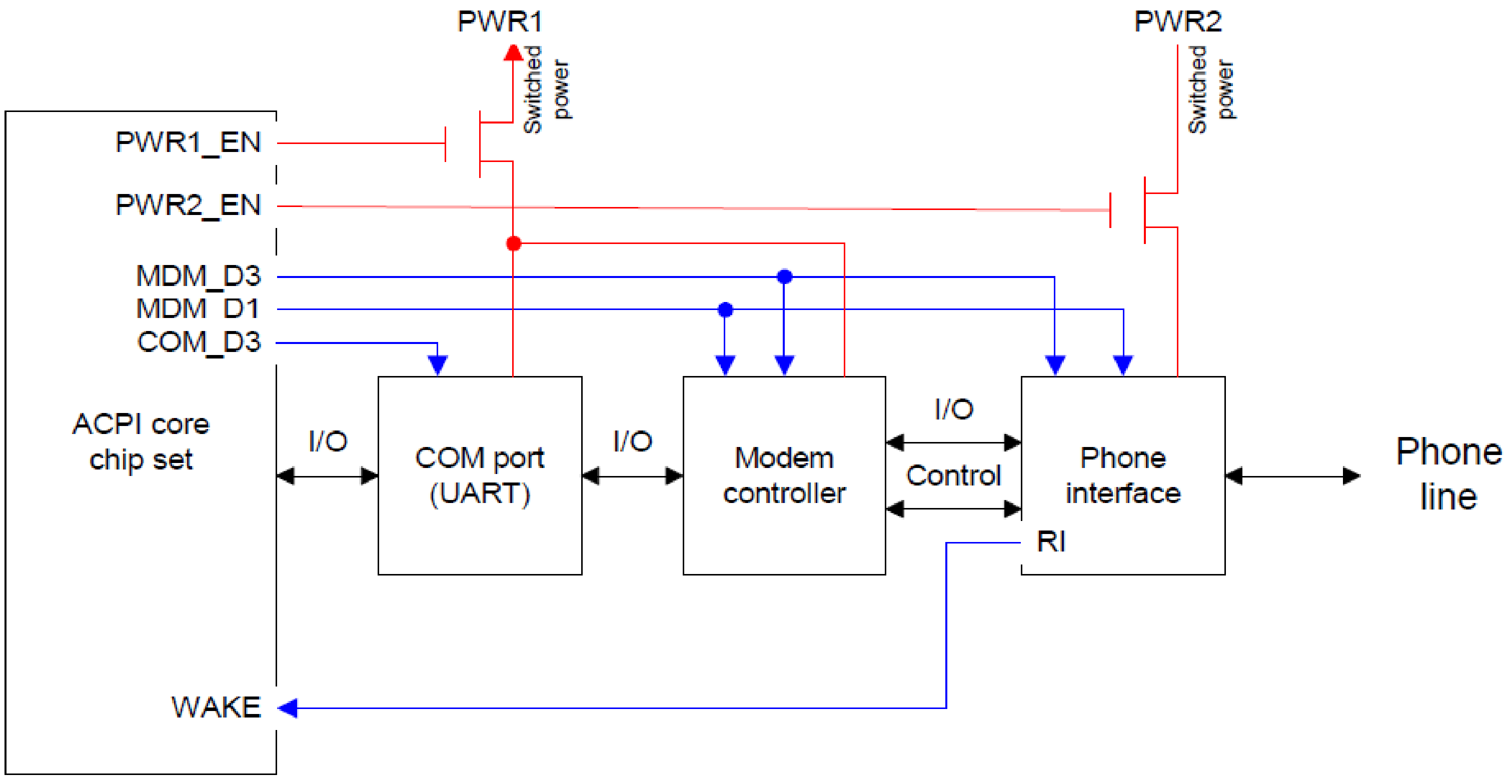
Generic devices

- Provides device independent power state definitions
- Defines classes of devices
  - Audio, COM Port, Display, Input, Modem, Network, PC Card, and Storage device class
- For each class ACPI defines device specific power characteristics
  - Device Power State Characteristics
  - Minimum Device Power Capabilities
  - Device Functional Characteristics
  - Device Wakeup Characteristics



Modem example

- Hypothetic modem diagram



Modem example

Power States

Device power state	State
D0	Modem controller on Phone interface on Speaker on Can be on hook or off hook Can be waiting for answer
D1	Modem controller in low-power mode (context saved by device) Phone interface powered by phone line or in low-power mode Speaker off Must be on hook
D2	Same as D3
D3	Modem controller off (context lost) Phone interface powered by phone line or off Speaker off On hook

Power Policy

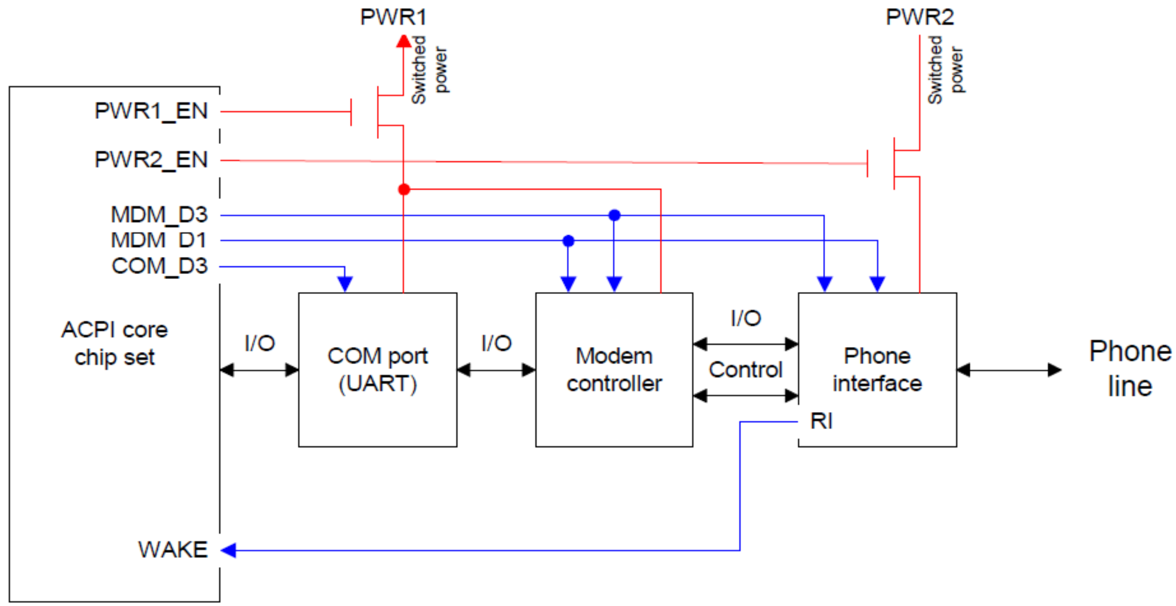
Transition	Action
D3 → D0	COM port opened
D0, D1 → D3	COM port closed
D0 → D1	Modem put in answer mode
D1 → D0	Application requests dial or phone rings

Wake Policy

Wakes the machine when phone rings (if wake enabled)

Defined by ACPI

Modem example

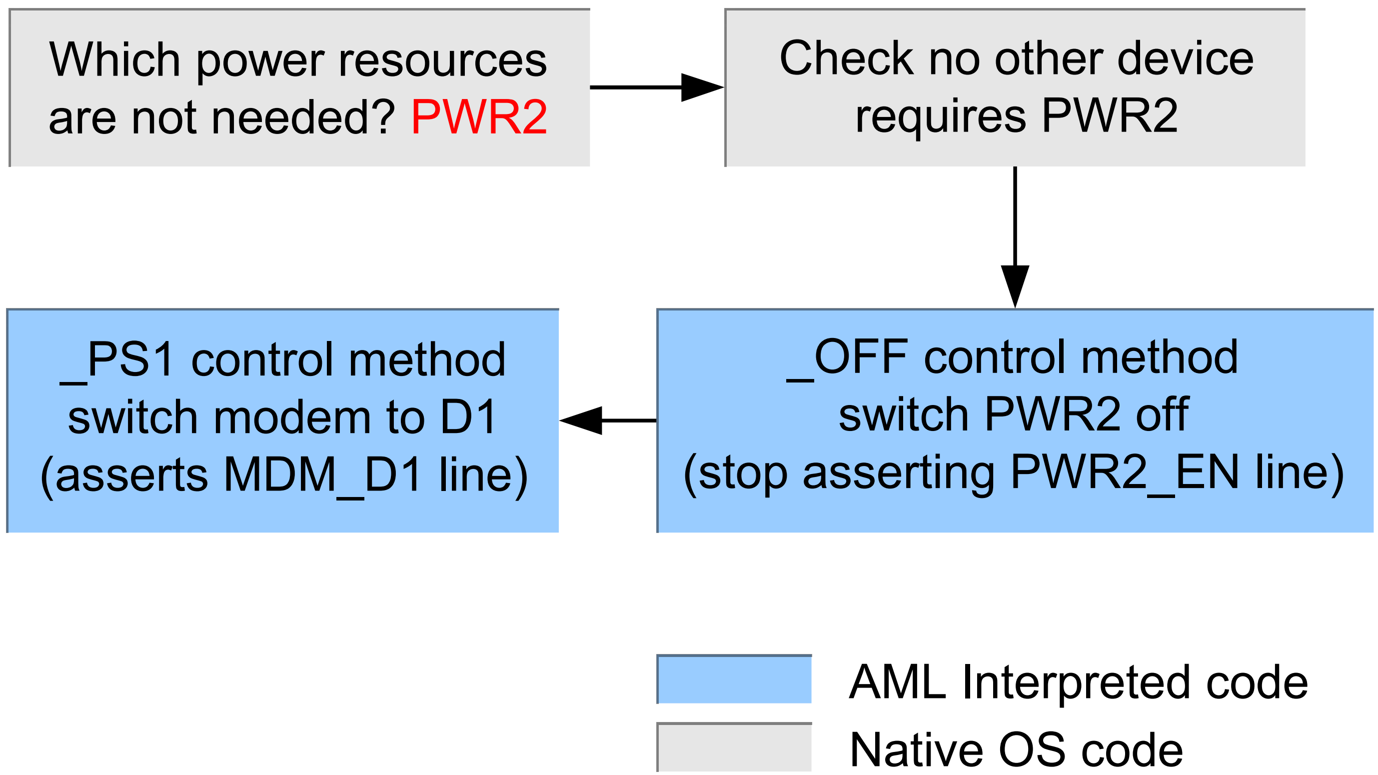


- Obtaining modem capabilities
  - D0 requires PWR1 and PWR2 power resources
  - D1 requires PWR1 power resource
  - (D3 implicitly requires no power resources)
  - To wake machine no power resources are needed

} Given by firmware

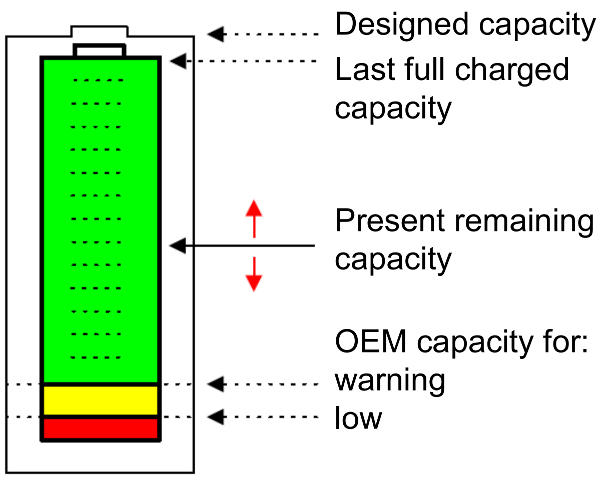
Modem example

Putting modem in answer mode to wait for a call (D0 → D1)



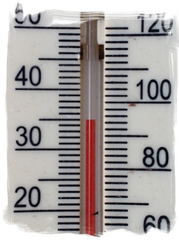
Battery

- Provides mechanisms to query
  - Designed capacity
  - Latest full charged capacity
  - Present remaining capacity
  - Rate of discharge
- Provides notification
  - Insertion or removal of battery
  - Start or stop discharging
- Battery levels
  - Warning: notification to user
  - Low: energy to go to sleep mode
  - Critical: no energy, emergency stop



The image belongs to the ACPI 4.0a specification and it is used with the permission of Intel Corporation.

Thermal management



- System divided in thermal zones
  - Usually the whole system is one zone
  - Includes devices, thermal sensors, and cooling controls
  - Trade off among performance, consumption, and noise
- Cooling modes
  - Active
    - Increase consumption to cool
    - Can involve acoustic noise
  - Passive
    - Decrease consumption and/or performance to cool
- Devices critical trip points
  - Define temperature to trigger system shutdown

Additional information

- ACPI Specification
  - http://www.acpi.info
- The ACPI Component Architecture Project
  - http://www.acpica.org/
- ACPI FAQ for Linux implementation
  - http://www.columbia.edu/~ariel/acpi/acpi_howto.txt

Data centres

Data centres

Design and architecture

Datacentres

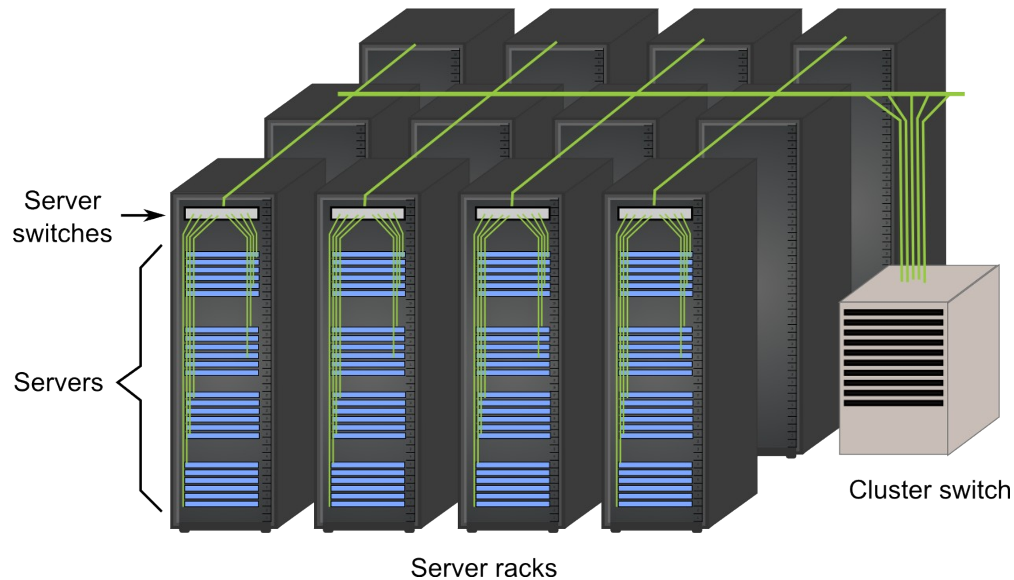
“Datacenters are buildings where multiple servers and communication gear are co-located because of their common environmental requirements and for physical security needs,and ease of maintenance.”

Barroso and Hölzle

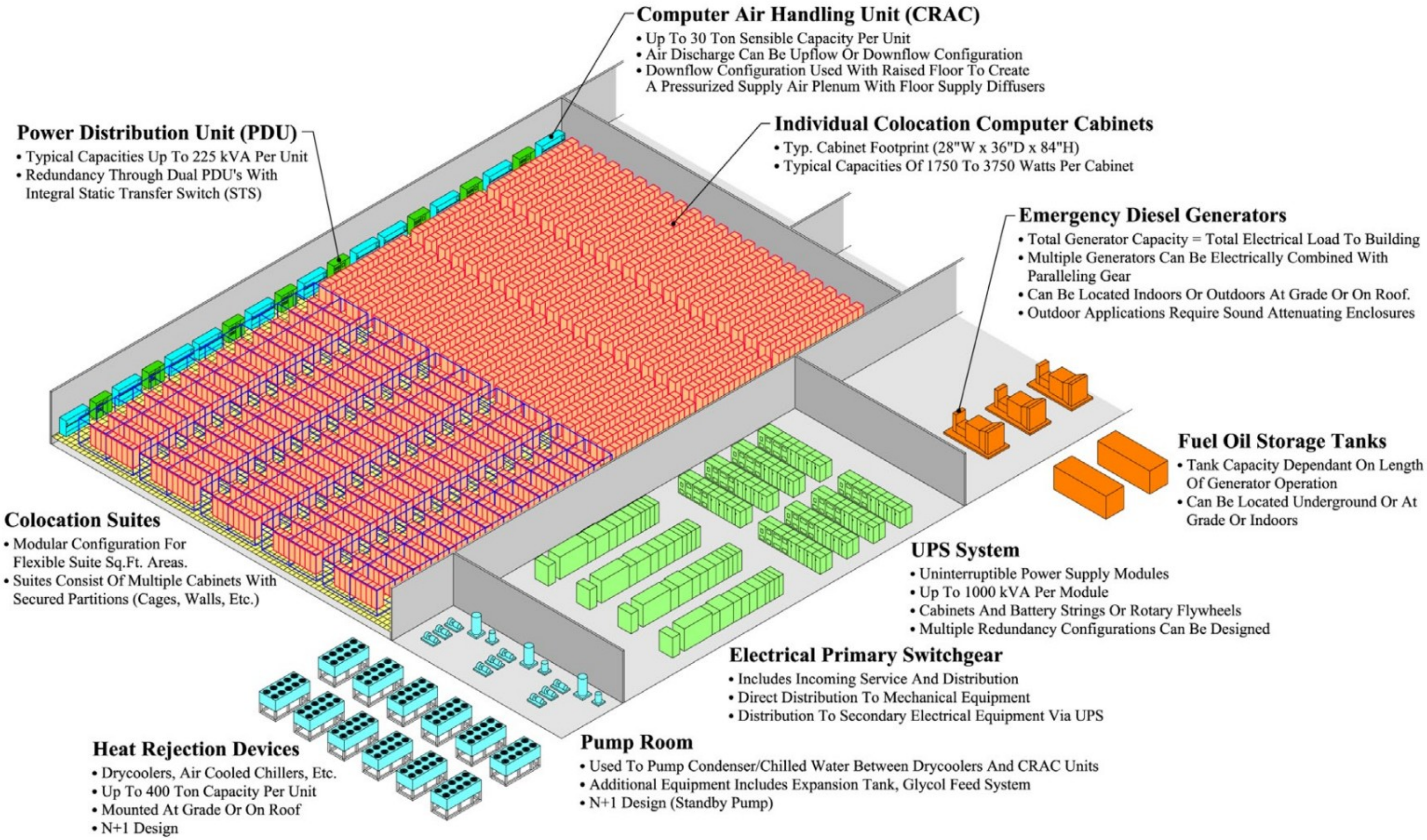
Luiz André Barroso and Urs Hölzle. The Datacenter as a Computer - An Introduction to the Design of Warehouse-Scale Machines. Morgan & Claypool Publishers, 2009.

IT Architectural overview

Server



Power and cooling systems



Graphic provided courtesy of DLB Associates

Power system

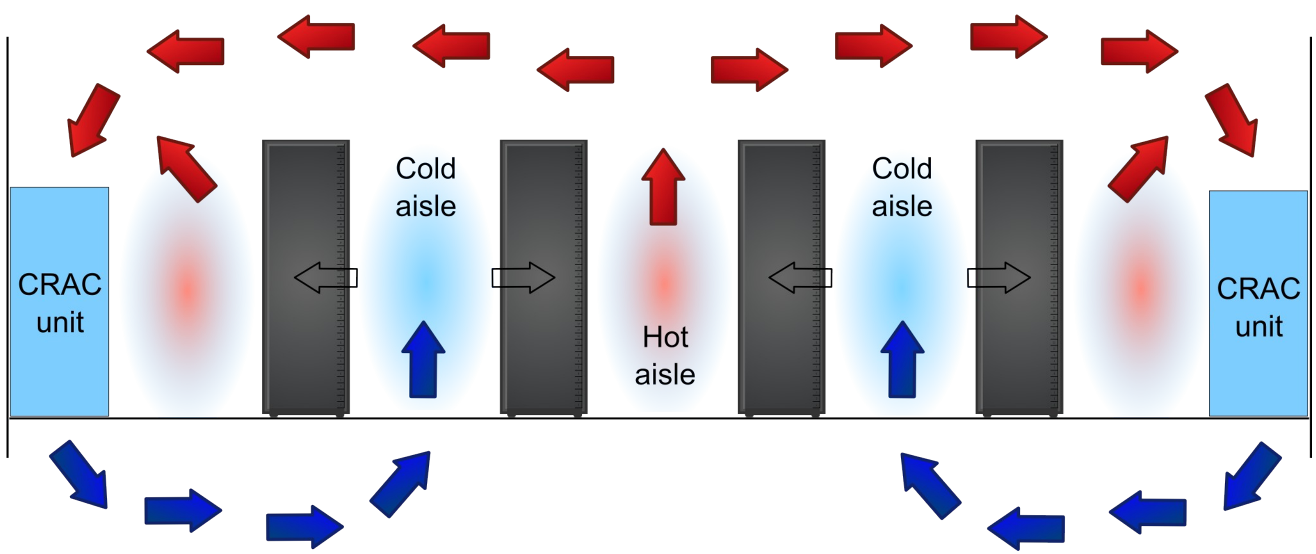
- Primary switchgear
  - Breakers for protection
  - Transformers (10-20kV to 110-600V)
- Diesel generators
  - Switched on in case of utility power failure
- Uninterruptible Power Supply (UPS)
  - Batteries for short term energy provision
  - Functionality
    - Switch energy source
    - Sustain system during generators start up with batteries
    - Power feed conditioning

Power system

- Power Distribution Units (PDUs)
  - Break a higher voltage line into several circuits
    - 200-480V line to many 110-220V circuits
    - 75-225 kW of total load at 6 kW circuit
  - Distributes energy to each rack
  - One breaker for circuit

Cooling system

- Computer Air Room Conditioning (CRAC)
  - Blows cold air under the floor plenum
  - Cold air moves to front of server racks (cold aisle)
  - Cold air flows through server racks
  - Warm air is expelled in the back (warm aisle)



Cooling system

- CRAC units cool the room's air
  - Liquid coolant is pumped from chillers or cooling towers
  - Coils are kept cool (12-14 ºC) with liquid coolant
  - Warm air is pushed through the coils by fans
  - Cold air (16-20 ºC) is moved to the floor plenum
- Air reaches the servers at 18-22 ºC
- Data centre cannot operate without cooling
- Cooling can account for 40% of the data centre energy consumption

Free cooling

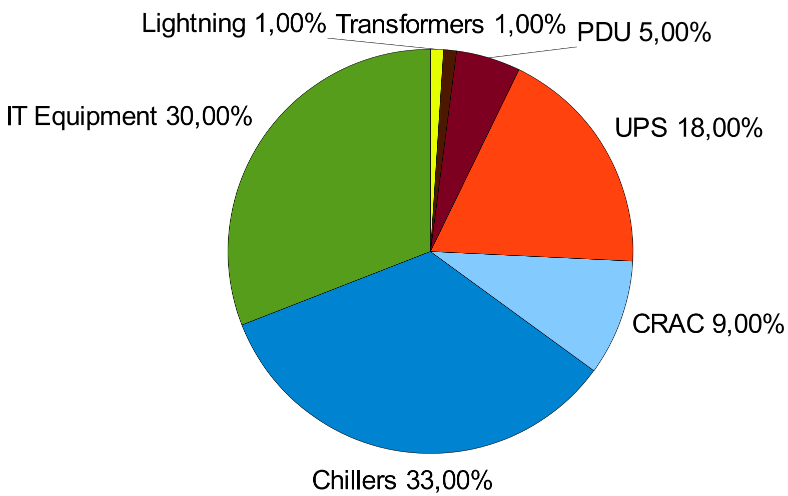
- Methods much more efficient than chillers

Cooling tower

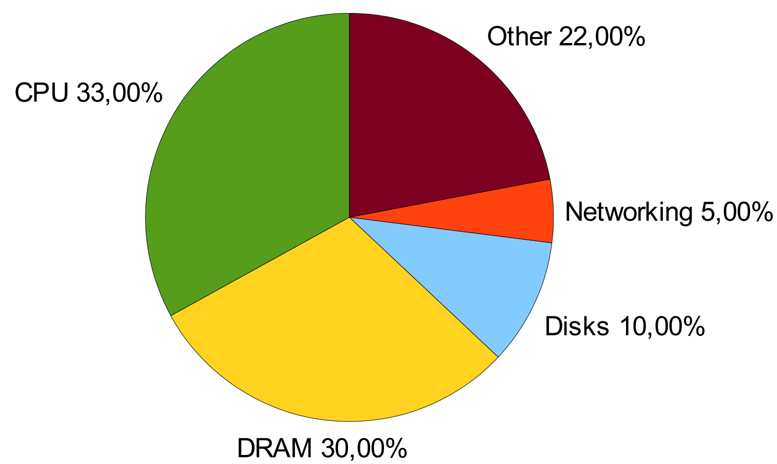
Glycol-based radiator

Fans to push air from outside

Consumption of a data centre



Datacentre overheads



Server consumption

Data based on Luiz André Barroso and Urs Hölzle. The Datacenter as a Computer - An Introduction to the Design of Warehouse-Scale Machines. Morgan & Claypool Publishers, 2009.

Redundancy and reliability

- Redundancy for fault tolerance and maintenance
  - Applied to energy and cooling systems

- Configurations: N+1, N+2, 2N

Tier Classification (ANSI/TIA 942)

Type	Availability	Description
Tier I	99.67%	Single path for power & cooling. No redundancy (N)
Tier II	99.74%	Single path for power & cooling. Redundancy (N+1)
Tier III	99.98%	Multiple paths for power & cooling (only one active). Concurrently maintainable. Redundancy (N+1)
Tier IV	99.995%	Multiple active paths for power & cooling. Redundancy in both paths (min (N+1))

W.Pitt Turner IV, J.H.Seader, K.G.Brill. Tier classifications define site infrastructure performance, Uptime Institute, White Paper

Container-based datacentres

Server racks, power distribution and heat exchange inside a container!!

Highly efficient cooling
High server density

Image from Wikimedia by user Sun Microsystem

Data centres

Efficiency metrics

Sources of efficiency losses

- Power systems

Element	Losses
Transformers	0.50%
UPSs	7-12%
Highly efficient UPSs	3%
Low-voltage power (110-220V) cables	1-3%

- Cooling
  - Fans that move cool and warm air
  - Mix of cool and warm air during long paths
  - Too low temperature selection
    - 25-27 ºC better than traditional 20 ºC

Energy efficiency metrics

- Power Usage Effectiveness (PUE)
  - Proposed by The Green Grid association
  - Efficiency of the IT support infrastructure

$$PUE = \frac{Total\ facility\ power}{Total\ IT\ equipment\ power}$$

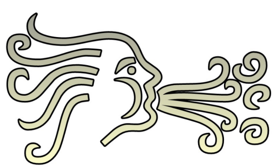
Current data centres have PUEs between 1.5 and 2.0

Google and Microsoft have reported PUEs around 1.2!!!

C. Beladi. Green Grid datacenter power efficiency metrics: PUE and DCiE. White paper. 2008

Energy efficiency metrics

- PUE has to be used carefully
  - Example: Server fans in IT-PAC Microsoft data centre container module



$$\frac{\textit{Facility power}}{\textit{IT equipment power}}$$

IT-PAC airflow management is improved → Server fans are no longer required → PUE incresases

Overall energy consumption reduced despite PUE increase

- Other aspects must be taken into account

http://www.datacenterknowledge.com/archives/2011/01/31/microsoft-eliminates-server-fans-despite-pue-hit/

Data centres

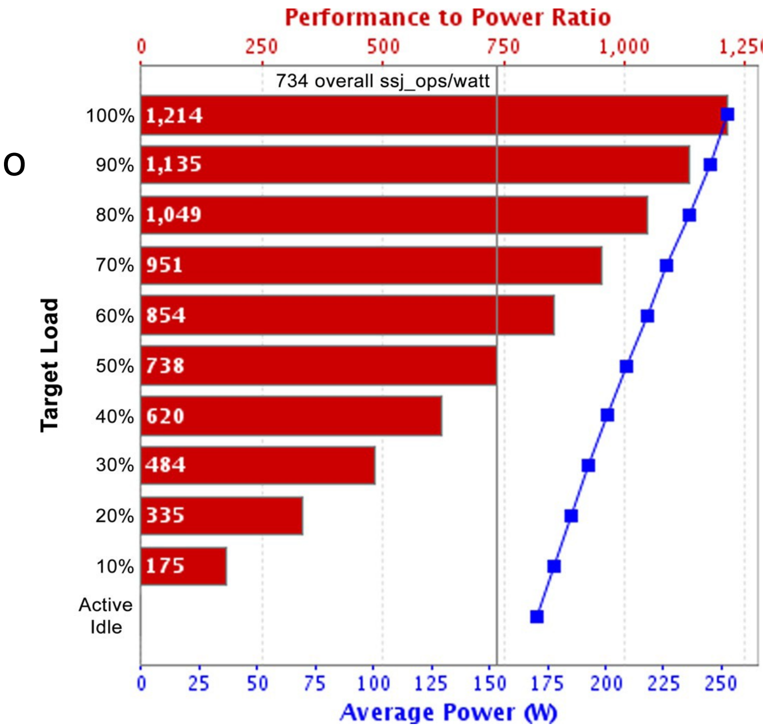
Energy-proportional computing

Benchmarks

- SPECpower_ssj2008
  - Benchmark with a broad range of workloads
  - Tests performance at 11 levels of load
  - Two metrics
    - Average system power
    - Performance-to-power ratio

<u>Example of benchmark for:</u>

2.83 Ghz Quad-core Intel Xeon
4 GB DRAM
7.2 k RPM 3.5” SATA HD

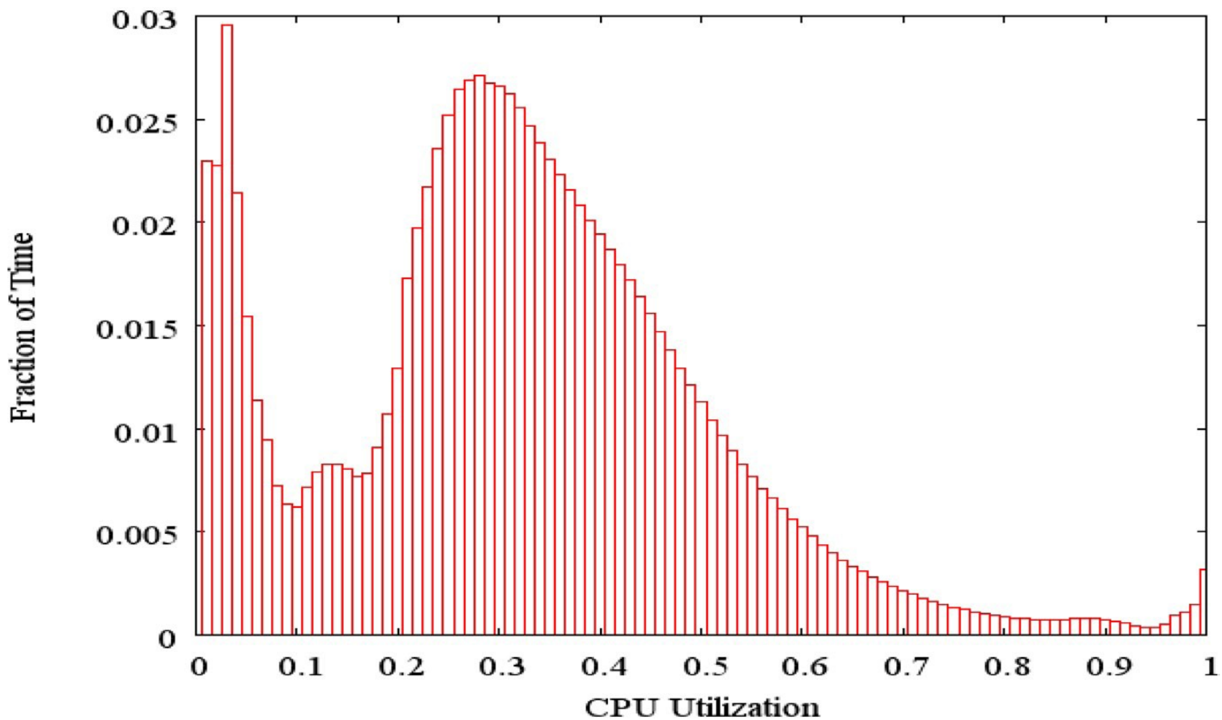


Results: http://www.spec.org/power_ssj2008/results/power_ssj2008.html

Image from Luiz André Barroso and Urs Hölzle. The Datacenter as a Computer - An Introduction to the Design of Warehouse-Scale Machines.

Load and energy efficiency

- Most of the time spent in no energy-efficient load regions



Activity profile of 5000 Google servers in 6 months

Image from Luiz André Barroso and Urs Hölzle. The Datacenter as a Computer - An Introduction to the Design of Warehouse-Scale Machines.

Load and energy efficiency

- Almost no time in idle state
  - In low load there are several hundreds of queries
  - Load spread out over all available servers
    - Optimised for performance and/or availability

Sustainable approach

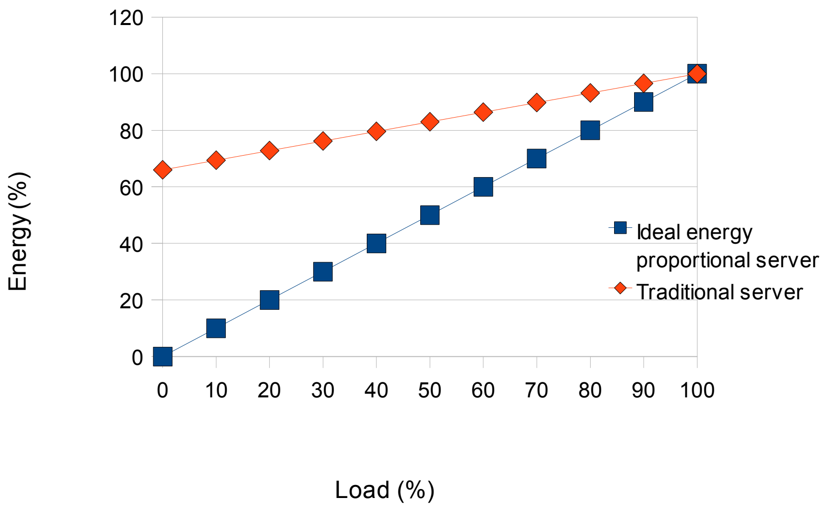
Concentrate load in a few servers
Switch off unused servers

But!!!

More software complexity
Energy for data distribution

Energy proportional computing

- Consumption proportional to the load
  - Ideal linear function without constants



Traditional server data based on data provided by Barroso and Hölzle.

Energy proportional computing

- Capacity to adapt consumption to load

Device	Dynamic power
CPU	3.0x
Memory	2.0x
Disks	1.3x
Network switches	1.2x

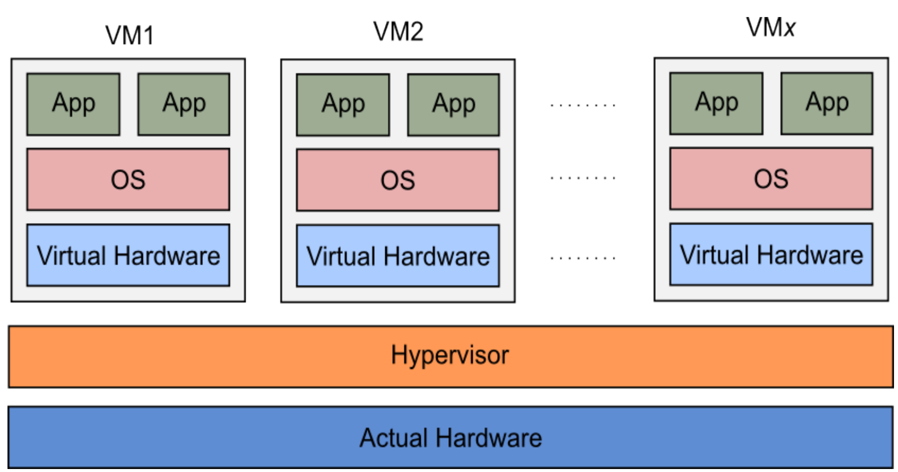
More hardware improvements are required!!!

↓

Meanwhile only option is to switch off hardware!!!

Virtualisation

- Software emulation of the hardware of a machine
  - Virtual machine
    - Emulated machine
    - Composed of data structures stored in memory and disk
  - Hypervisor or Virtual Machine Monitor
    - Software that manages and runs the virtual machines
    - Examples: VMWare, VirtualBox, Parallels...



Virtualisation – Consolidation of servers

Traditional servers

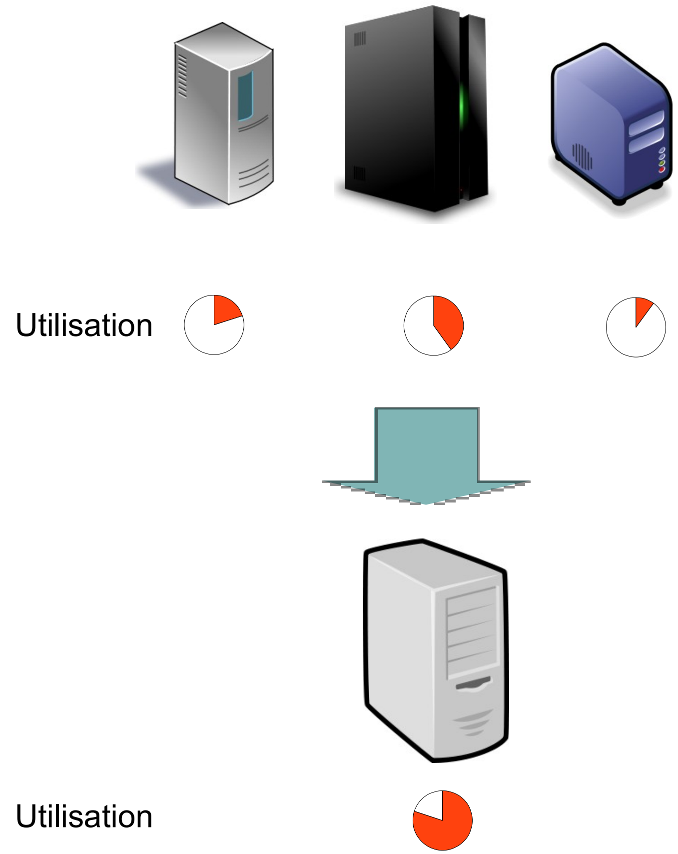
Non homogeneous servers

- Different functionality and types
- Different requirements
- Possibly underutilised

Virtualisation

A few homogeneous servers

- Run several heterogeneous virtual servers
- Higher utilisation



Virtualisation – Consolidation of load

Cloud

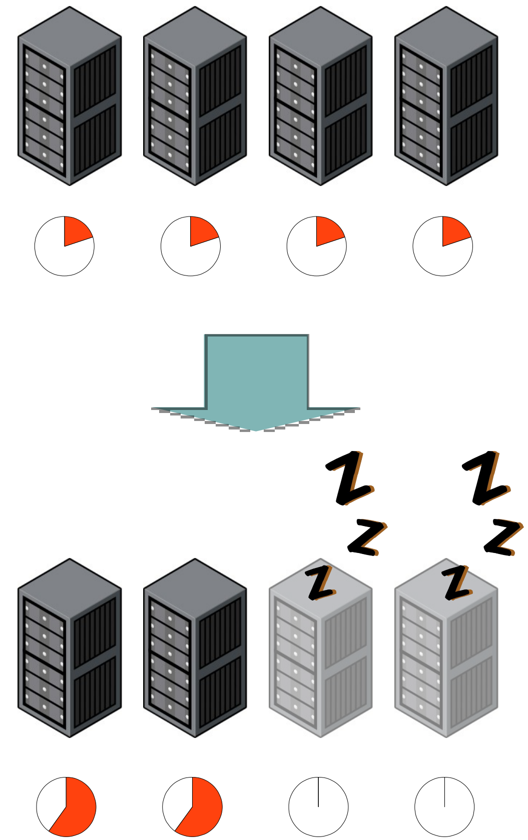
Thousands of on-line servers

- Same type workload
- Same requirements
- Underutilised when no peak load

Virtualisation

Thousands of servers, subset of them on-line

- Run several virtual servers
- Migrate and consolidate virtual load when demand is low



Virtualisation - Challenges

- Virtual machine provisioning costs
  - Creation overheads
  - Reutilisation
- Matching job requests and requirements
  - Appropriate virtual machine configuration to job requests
  - Avoidance of overprovisioning
- Migration of tasks
  - Associated costs of migration
  - Tradeoff between migration and task completion

Data centres

Initiatives and standards

ENERGY STAR

ENERGY STAR

- Program Requirements for Computers Servers 1.0 (15th May 2009)
  - Defines different terms
    - Basics, types of servers, complementary data equipment, server components, and other additional terms
  - Mainly focused on servers up to 4 processors
    - Power supply efficiency requirements
    - Idle power requirements
  - Enforces provision of system information
    - *Version 1.0 Power and Performance Data Sheet*
  - Defines test procedures to qualify products

http://www.energystar.gov

ENERGY STAR

- Power supply efficiency requirements

Table 1: Efficiency Requirements for Computer Server Power Supplies

Power Supply Type	Rated Output Power	10% Load	20% Load	50% Load	100% Load
Multi-Output (AC-DC & DC-DC)	All Output Levels	N/A	82%	85%	82%
Single-Output (AC-DC & DC-DC)	≤ 500 watts	70%	82%	89%	85%
	> 500 - 1,000 watts	75%	85%	89%	85%
	> 1,000 watts	80%	88%	92%	88%

Comprise loads different than 100%

$$Efficiency = \frac{Power\ Output}{Power\ Input}$$

Table from ENERGY STAR Program Requirements for Computers Servers 1.0

http://www.energystar.gov

ENERGY STAR

- Idle power requirements

Table 3: Base Configuration Idle Power Requirements

Computer Server Type	Idle Power Limit
Category A: Standard Single Installed Processor (1P) Servers	55.0 watts
Category B: Managed Single Installed Processor (1P) Servers	65.0 watts
Category C: Standard Dual Installed Processor (2P) Servers	100.0 watts
Category D: Managed Dual Installed Processor (2P) Servers	150.0 watts

Managed Server

High available server in a highly managed environment

- Capability to operate with redundant power supplies
- Dedicated management controller

No idle power limit defined for more than 2 processors!!!
They just enforce to provide low power mode for idle.

Table from ENERGY STAR Program Requirements for Computers Servers 1.0

http://www.energystar.gov

ENERGY STAR

- Idle power requirements

Table 4: Additional Idle Power Allowances for Extra Components

System Characteristic	Applies To:	Additional Idle Power Allowance
Additional Power Supplies	Power supplies installed explicitly for power redundancy[1]	20.0 watts per Power Supply
Additional Hard Drives (including solid state drives)	Installed hard drives greater than one	8.0 watts per Hard Drive
Additional Memory	Installed memory greater than 4 GB[2]	2.0 watts per GB[2]
Additional I/O Devices	Installed Devices greater than two ports of 1 Gbit, onboard Ethernet [3]	< 1Gbit[4]: No Allowance = 1 Gbit[4]: 2.0 watts / Active Port[5] > 1 Gbit[4] and < 10 Gbit[4]: 4.0 watts / Active Port[5] ≥ 10 Gbit[4]: 8.0 watts / Active Port[5]

Table from ENERGY STAR Program Requirements for Computers Servers 1.0

ENERGY STAR

- *Version 1.0 Power and Performance Data Sheet*
  - Model and configuration identification
  - System characteristics
  - System configuration(s) (maximum, minimum and typical)
  - Power data
  - Additional power and performance data
  - Available and enabled power saving features
  - Power measurement and reporting capabilities
  - Thermal information from the ASHRAE thermal report
  - Additional qualified configurations

Example of datasheet for server Dell PowerEdge:

http://www.dell.com/downloads/global/products/pedge/en/PowerEdge_R210_250W_Energy_Star_Data_Sheet.pdf

ENERGY STAR

- Future specifications
  - Program Requirements for Computers Servers 2.0
    - Currently first draft available
    - Will include active consumption
      - as opposed to just idle consumption
  - Program Requirements for Data Center Storage 1.0
    - Currently second draft available
    - Will establish maximum consumption of storage products

Storage product

> “Fully-functional storage system that supplies data storage services to clients and devices attached directly or through a network”

Definition from referred specification

http://www.energystar.gov

EU Code of Conduct for Data Centres

EU Code of Conduct for Data Centres

- European action to reduce consumption of data centres
  - Best practices for data centres operators
    - V3.0 Guide (Feb 2011)
  - Operators register and commit to their application

- Roles of applicant
  - Participant
    - Operator of data centre or equipment in it
    - Commitments
      - Annual report of energy consumption
      - Implementation of some of the best practices

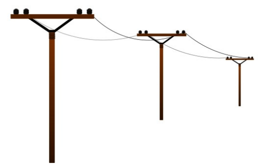
  - Endorser
    - Support the initiative and participants

EU Code of Conduct for Data Centres

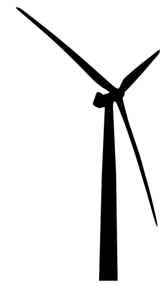
- Scope of the best practices
  - Data centre utilisation, management and planning
  - IT equipment and services
  - Cooling
  - Data centre power equipment
  - Other equipment
  - Building
  - Monitoring

- Best practice expected implementation
  - Immediate
  - During software install or update
  - During new IT introduction or replacement
  - During building of datacenter or retrofit
  - Optional

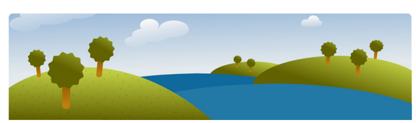
Sustainability influenced by data centre location



Short distance to power plant reduces losses



Renewable energies reduces contamination



Sources of free cooling reduce energy spent

Infrastructures that require heat can reuse wasted heat

Some advantages may change along time!!!