Understanding
Data Center Engineering

SWINOG-26, Berne, 23. May 2013

André Oppermann <oppermann@gmail.com>

Contents

(Only 42 slides)

About Data Center Engineering

- Data Center Engineering has become its own overarching discipline
  - Stringent uptime, availability and reliability requirements
  - High power/current electrical systems
  - Efficient cooling and mechanical systems
  - Very high volume air handling systems
  - Backup engine-generator systems
- Tight integration of all of the above into a complete and coherent package

About Data Center Engineering

- It is all about trade-off‘s
  - The right ones of course
  - On the overall goals
  - Many individual technical decisions
  - Even small and simple things can have a large impact
- Design must be beautiful in its simplicity
  - Complex design leads to frequent (human) errors
- Good Engineering is the skill and art of:
  - making the right trade-off‘s
  - at the right time
  - in the right place

About the presenter

- Built first data center in 1998
  - Known as TIX in Zürich (Hardstr. and Josefstr.)
  - Telehouse Zürich -> IXeurope -> Equinix
- DeepGreen (Walensee)
  - Principal engineer and CTO
  - 50MW total power
  - PUE_{est} 1.1 with lake source cooling
  - 12 months for full government permit incl. HV substation (!)
- EBM Telecom „Datacube“ (Basel Münchenstein)
  - Principal project manager
  - 6MW total power
  - PUE_{est} 1.2 with „conventional“ cooling
  - 3.5 months for full government permit (!)
- On-time and **on-budget** (!)

About the presenter

- Electrical engineer by training
- Well versed in cooling and air handling systems
- Lots of experience
  - How things *should* be done
  - How things should *not* be done
- „Accredited Tier Designer“ certification from Uptime Institute
  - Like CCIE but for data center engineering

ATD TIER™
André Oppermann
Accredited Tier Designer
February 7, 2013
573 UPTIME INSTITUTE

Classification of Data Centers

- General purpose corporate IT
  - Scheduled downtime allowed outside office hours
  - „Spare room“ type at SME businesses
- Always-on general purpose
  - No downtime
  - Internet, Global operation, Banks, Trading, ...
  - Standard servers and systems
  - No control over IT equipment
- Always-on specialized
  - No downtime
  - Google, Facebook, Amazon, ..., Number crunching
  - Special purpose built servers
  - Control over IT equipment

Goals for data center engineering

- Requirements from the client / owner
  - Highly reliable and available (24x7 forever)
    - No scheduled downtime ever
    - Not for maintenance nor capacity extensions
  - Highly scaleable in initial vs. final buildout
    - Build capacity to actual demand
    - Retain the ability to significantly increase capacity
  - High power density per unit (m2 or rack)
    - More cpu‘s and cores in the same package
    - To be able to actually fill the rack with servers
  - Inexpensive (!)
    - Limited budget
- Choose any three of the above

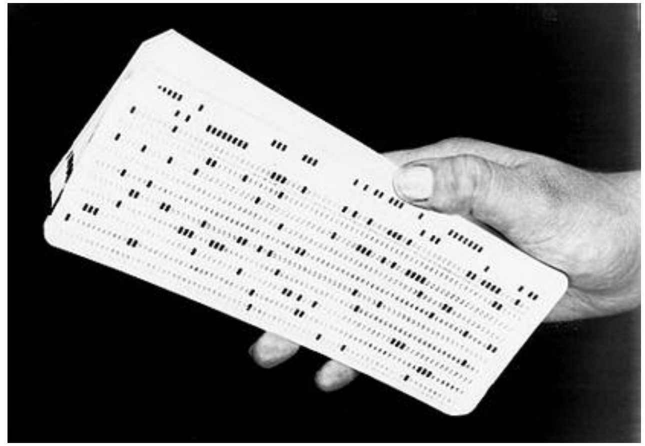
Goals for data center engineers

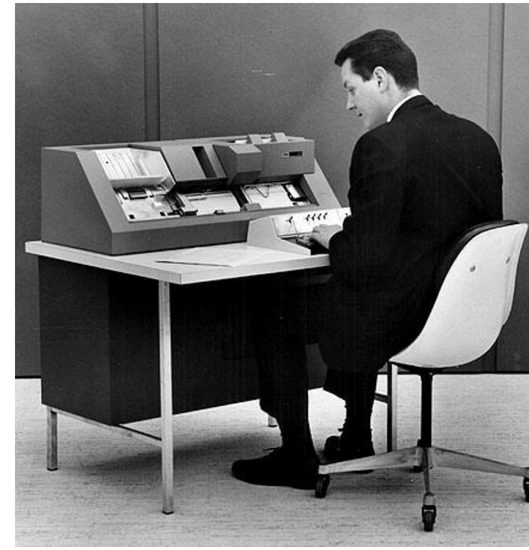
- It *is* possible to have all four
  - With only a small increase in cost
  - Requires the engineer(s) to treat it as a whole system, not separate parts in the same place
- Involves
  - In-depth analysis of the exact requirements
  - Environmental situation of the data center site
  - Doing lots of physics and math calculations to find the optimal balance of all moving parts
- Specialized knowledge and experience
  - Normal sub-system specific engineers do *not* have it

History of data centers

- There are actually two histories
  - Business computing
    - Punch cards
    - Mass printing of letters and invoices
  - Pure computing systems
    - Typically military related
    - Used to be top secret and not know to the public
- Important to understand where certain requirements and concepts came from
  - Paper and card board is very sensitive to moisture
  - Vaccum tubes are sensitive to overheating
  - Large and bulky, difficult to control air flow
- These resulted in a tight temperature and moisture control envelope
  - and entire room cooling with mixed air

History of data centers: Punch cards





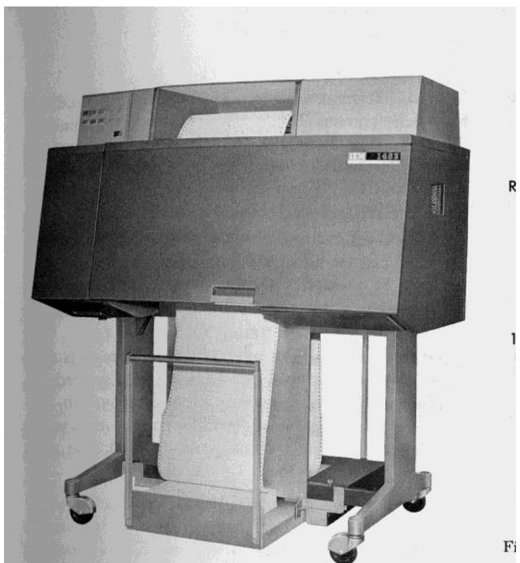


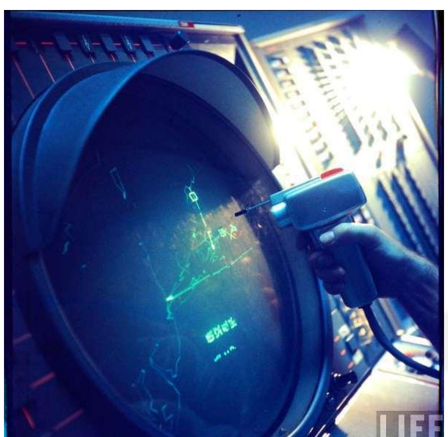
Figure 7. IBM 1403 Printer

History of data centers: SAGE

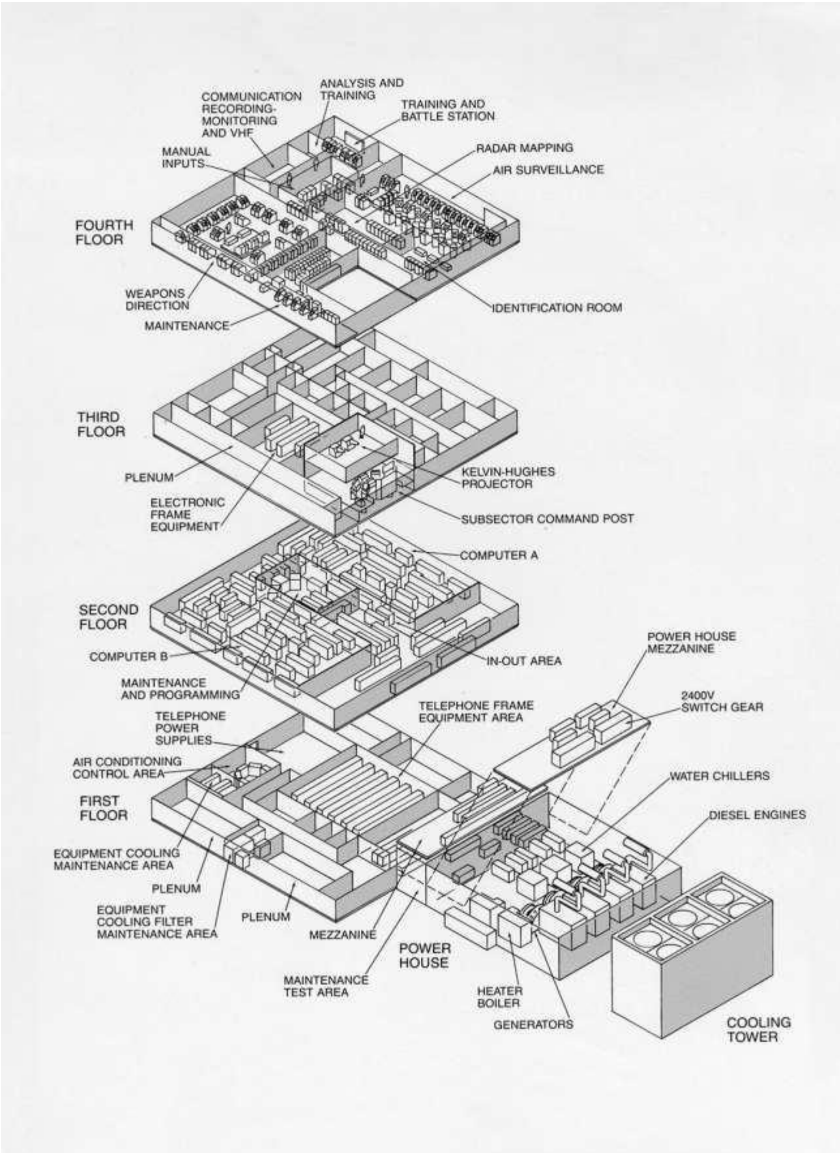
- Semi-Automatic Ground Environment
  - Automated air defense for north America during cold war
  - One SAGE Direction Center per Air Defense Sector
  - 24 direction centers, 134 radar sites
  - Started in 1953, first two ready in 1958, fully operational from 1963 until 1983
  - Predecessor of NORAD, civil air traffic control
- First data center building block built the way as nowadays
- Impressive figures
  - 3 MW power consumption
  - 2x 60‘000 vaccuum tubes (2 redundant computers hot standby)
  - 2‘000 m2 per computer, 15‘000 m2 total
  - 32 bits, 4x64K words magnetic core memory with parity
  - 166kHz clock speed
  - 150 display consoles

- See http://scottlocklin.wordpress.com/2013/03/28/the-largest-computer-ever-built/

History of data centers: SAGE



History of data centers: SAGE



Engineering Standards

- Uptime Institute
  - The Tier level classification
  - Based on failure and detailed risk analysis
  - Defines outcomes, not methods or parts
  - Rigerous and formal certification
- ECO, Bitcom, TÜV IT
  - Pretty system graphs
  - Very flexible
  - Only partially helpful
  - Mostly vendor dominated
  - Can certify parts and different build standards
    - Have to carefully read the fine print
  - *No outcome specification*

Engineering Standards

- Simple statements are misleading
  - „N+1“ does *not* define an engineering standard
- The objectively observable outcomes have be defined
  - The „how“ is up to the engineer, state of the art and available components
  - The „why“ and „what“ is specified
- These parts have to be described
  - Capacity components
  - Distributions paths
  - Concurrent maintenance
  - Fault tolerance
  - True Net capacity ratings

Uptime Institute Tier Levels

- Focused on
  - Outcomes => what actually happens
  - Preventive maintenance without interruption
- Certifications
  - Design certification of the plans and schematics
  - Full certification as actually built on site
  - Operational sustainability
    - All preventive maintenance processes in place
    - Routine inspection of all measured values Trained operators
    - Trained operators
    - A bit like ISO9000 quality processes specifically for data center operations
- Very strict, no compromises allowed

Uptime Institute Tier Levels

- Tier I: Basic Site Infrastructure
  - Non-redundant capacity components and distribution path
- Tier II: Redundant Capacity Components
  - Redundant capacity components (N+1 UPS, Generator, Cooling)
  - Non-redundant distribution path
- Tier III: Concurrently Maintainable
  - Redundant capacity *and* distribution paths (N+1 *everywhere*)
  - *Each and every* capacity and path component can be removed from service without downtime (both electrical *and* cooling)
- Tier IV: Fault Tolerant
  - Two independent and phyiscally isolated capacity and distribution paths
  - „N“ capacity after *any* failure (including generators)
- All Tier Levels:
  - True rating of capacity components
  - Dual utility feeds or grid connections *not* required

Standards: Efficiency

- The Green Grid
  - PUE Power Utilization Efficiency
    - The lower the more efficient (can‘t be lower than 1.0)
  - WUE Water Utilization Efficiency
    - Fresh water consumption when using evaporative cooling
  - DCcE Data Center compute Efficiency
    - Energy per compute unit
  - CuE Carbon usage Efficiency
    - Where your energy comes from
  - ERE Energy Reuse Efficiency
    - Waste heat used for building heating
- The government
  - Wants to see that you‘ve taken all reasonable measures to reduce and re-use energy
  - How far you have to go depends on canton and the official reviewing your permit application (!)

Physics: Electrical and Cooling

- Laws of Nature
  - Creative accounting doesn‘t work here
  - Hand waving doesn‘t work either

- Embrace, understand and work with them
  - Don‘t rely on old look-up tables or methods
  - Many classic engineering rules-of-thumb do *not* apply to data center engineering

Physics: Electrical

- The typical electrical load in a data center is very different from a commercial building or industrial building
  - Only electronic computer power supplies
  - Constant power draw (wrecks overbooking strategy of local grid)
  - Cooling system (high jump loads)

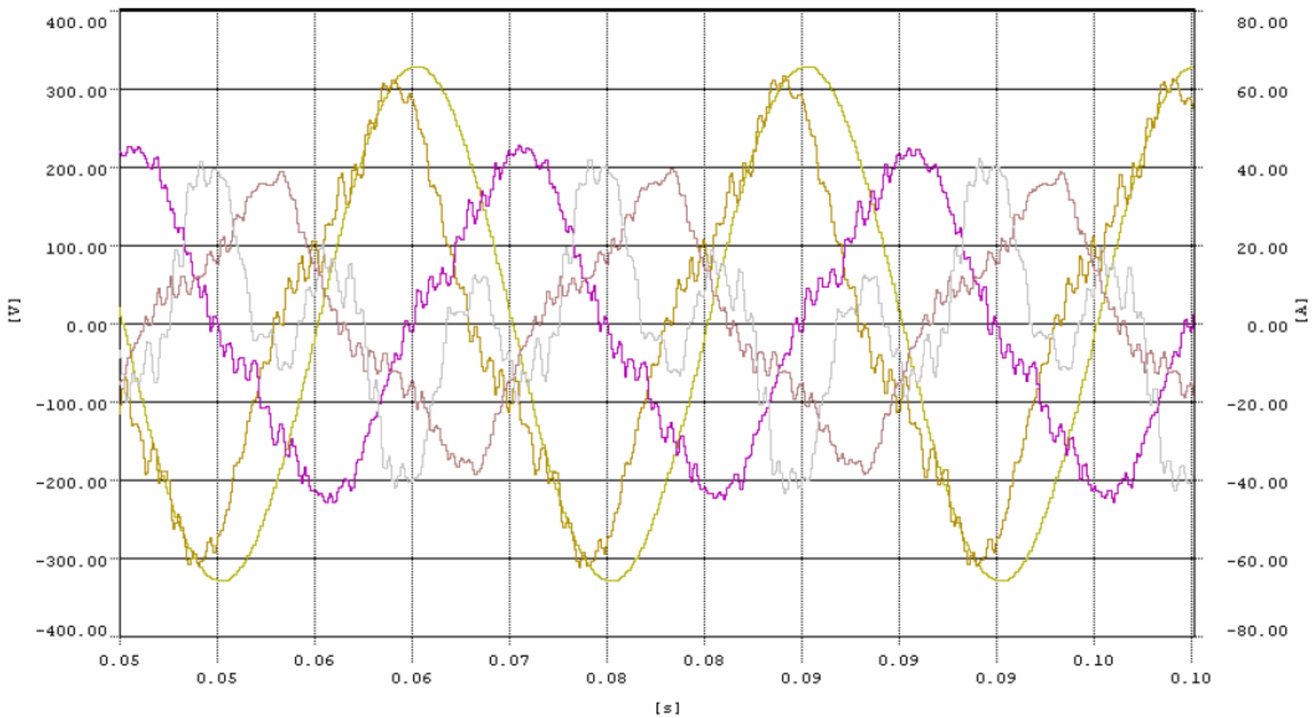
- Characteristics
  - Mostly single phase 230VAC *switching* power supplies
    - Only few servers have 3-phase power supply options
  - Severe phase load imbalances
    - Up to 50% or more due to little coordination
  - Load factor is highly capacitive (leading current)
    - Cos-φ of 0.75_{cap} not uncommon
  - Distored non-linear loads with high harmonics
    - Up to 50% of the total current in the harmonics
  - High in-rush currents on power up (!)
    - Easily trips loaded fuse breakers after outage

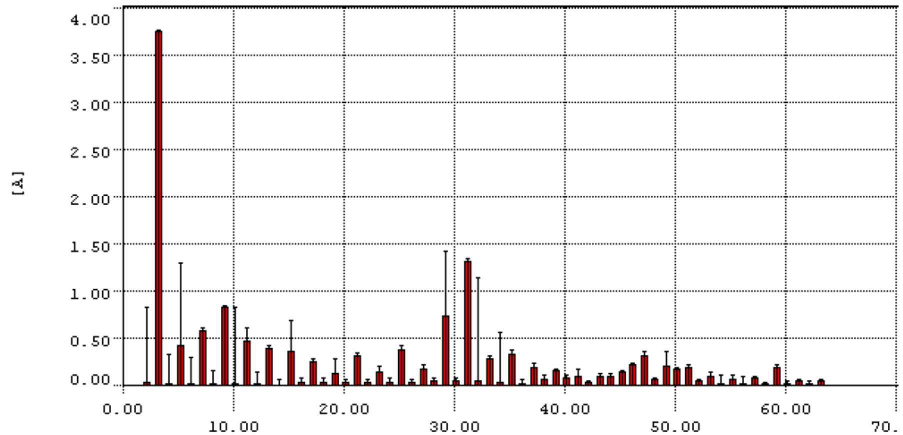
Physics: Electrical

- Important Parameters in an AC Power System
  - Loads
    - Linear (resistant loads, inductive motor loads)
    - Non-linear (any form of electronic power supply)
  - Power
    - Active Power (kW)
    - Apparent Power (kVA)
    - Reactive Power (kVAr)
  - Power factor
    - Phase shift angle between voltage and current (cos-φ)
    - Displacement power factor (λ)
    - Distortion power factor
  - Currents
    - RMS current (A)
    - Fundamental current (A)
    - N-th Harmonic current (A) [*this is the evil stuff*]
    - Total harmonic current distortion (THiD)
  - Voltage
    - Frequency (Hz)
    - Total harmonic voltage distortion (THvD)

- Summary
  - You can‘t simply compare data sheet values
    - **1kW at source ≠ 1kW at sink**
  - Causes de-rating of the systems name-plate capacity
  - Unproductive current flows creating additional losses

Physics: Electrical





Physics: Electrical

- UPS systems
  - Lies and misleading information
    - True capacity kW vs. kVA vs. Power factor vs. Distortion
    - Efficiency is measured and reported with linear loads
  - Different types
    - Static (electronic) with batteries
      - Battery placement (air conditioned) and maintenance is critical
      - Energy storage for 5 to 30 minutes (second chance for generator)
      - Typically modular with smaller and larger increments
    - Dynamic (and hybrid)
      - Rotating mass based with flywheel on- or off-axis
      - Energy storage only for 20-60 seconds (no second chance)
      - Typically expandable in large increments
  - The UPS bypass design is extremely critical

- About 60% of all unscheduled outages come from power + UPS failures!
- Be wary of magnetic field effects from parallel high current cables and power rails

Physics: Electrical

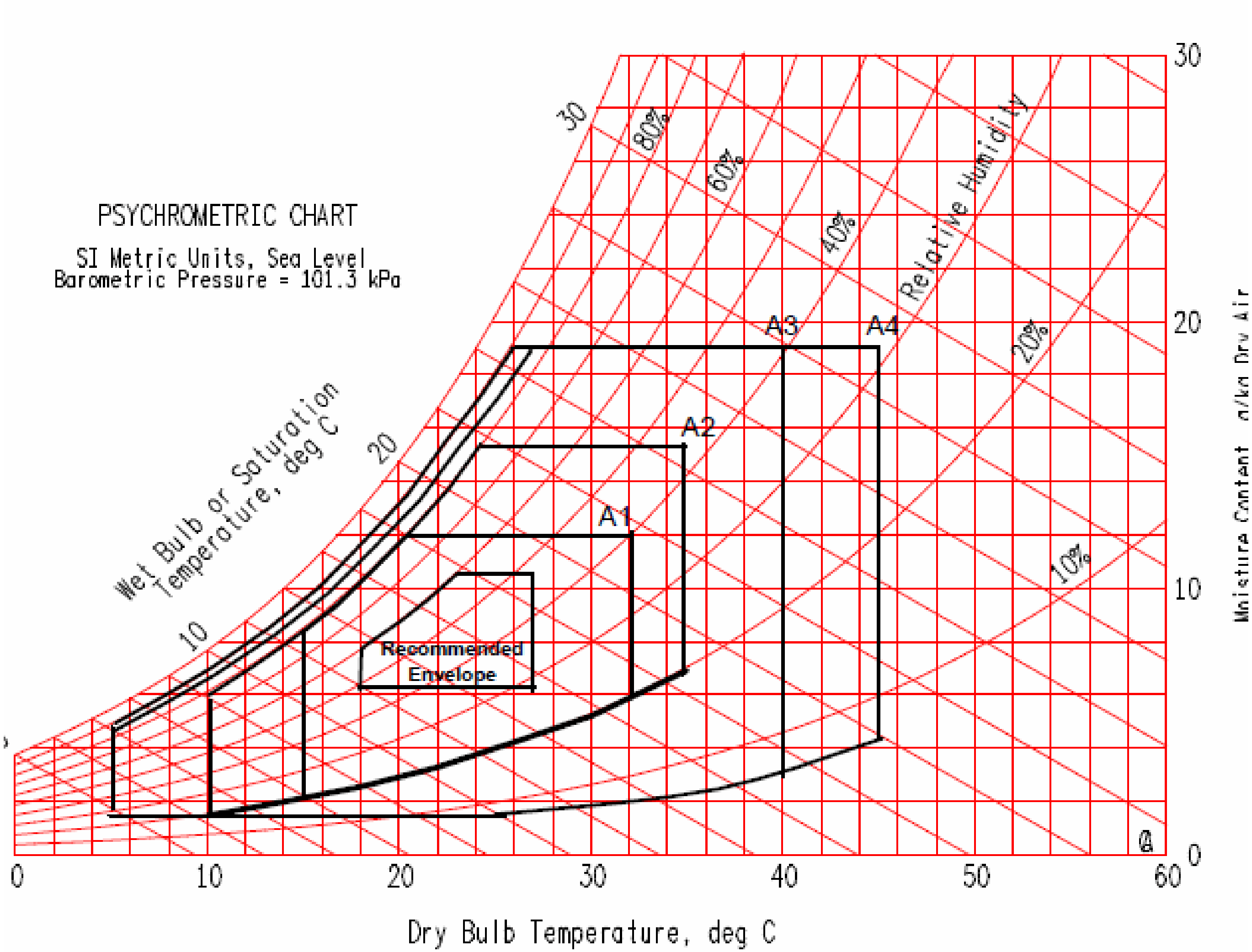
- Engine-generator systems
  - Power rating
    - Standby vs. Prime vs. Contiguous
    - Grid-parallel operation? => Special permits from local grid operator required
    - Pre-heating, from zero to full load in 10 seconds
  - Noise and exhaust emissions
    - Complex and stringent regulations depending on jurisdiction
    - Exhaust stack height requirements may be surprising
  - Fuel stability
    - Fouling (fungii, bacteria, algae)
    - Moisture takeup
    - Instability of fuel additives
    - Suppliers only specify 6 months of storage (!)

Physics: Electrical and Cooling

Electrical input = Heat output

Physics: Cooling

- What is acceptable to modern IT equipment

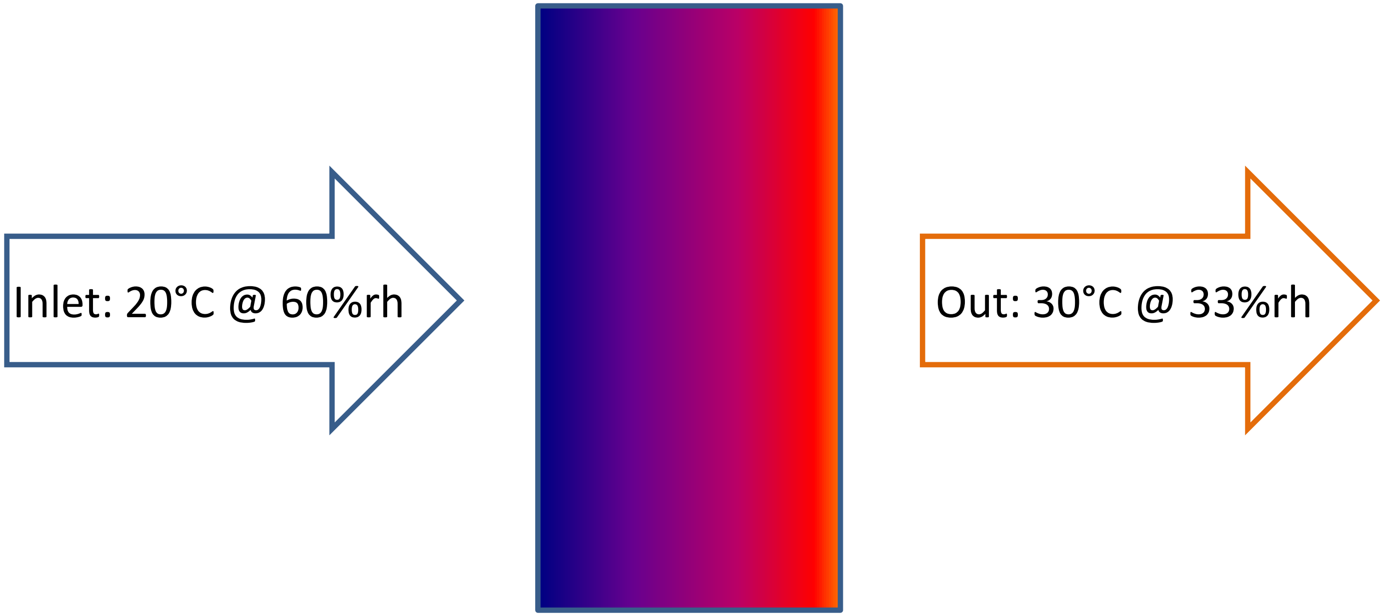


Physics: Cooling

- The allowed range is specified by
  - ASHRAE (standards organization)
  - IT Equipment Manufacturers
    - Sometimes more generous
- Generally we can say
  - Wide (inlet) temperature range
    - Class A2: 10° – 35° C
  - Wide (inlet + outlet) humidity range
    - Class A2: 20 – 80 % relative
    - Humidity is relative to temperature
  - Temperature rate of change is limited
    - Class A1 to A4: 5°C over 20 hours
    - Thermal stress on materials
  - Be aware of high altitude de-ratings!
    - Starting from 1‘000m

Physics: Cooling

- Be aware of inlet and outlet temperature/humidity



- Condensation outside of air handlers must be avoided
  - Inside air handlers it is very in-efficient
and requires re-humidification
  - Want to avoid humidification entirely

Physics: Cooling

- Air is our primary heat transport media
  - Unfortunately air is a horrible heat conductor
    - But abundant and easy to handle
- Water is our secondary heat transport media
  - Excellent heat conductor

	Water	Air
Specific heat capacity	4.18 kJ/(kg•K)	1.01 kJ/(kg•K)
Spec. heat transfer capacity	0.55 W/(m•K)	0.026 W/(m•K)
Volume of 1 kg	1 Liter	773 Liters
Transporting 1kWh @ Δt 10K	86 kg / h	356 kg / h
	86 Liters / h	275‘188 Liters / h

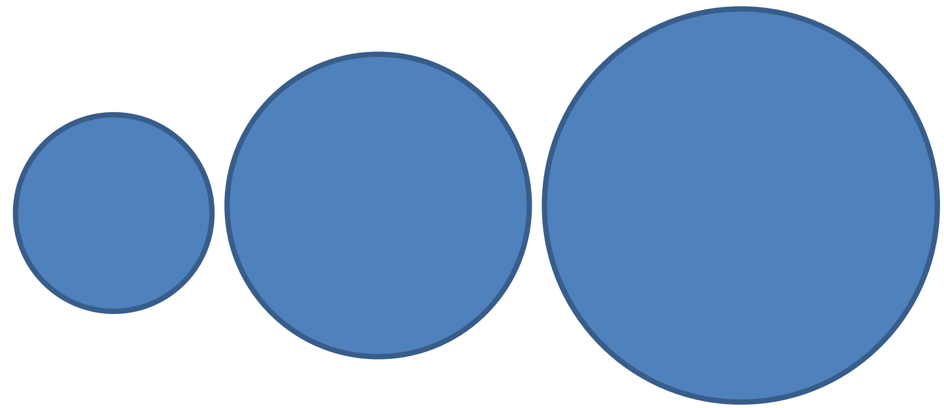
Physics: Cooling

- Formulas

Volume $\frac{V_1}{V_2} = \frac{n_1}{n_2}$

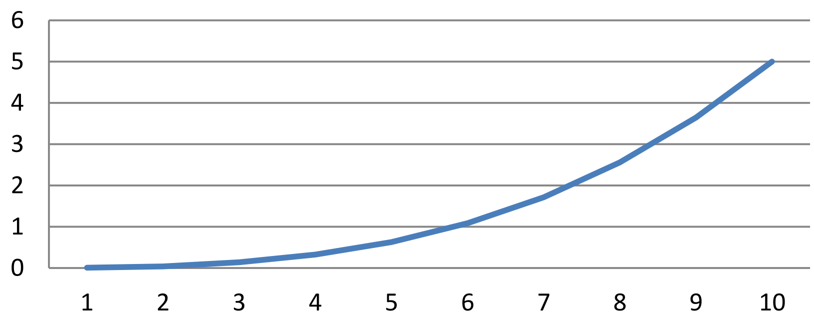
Pressure $\frac{p_1}{p_2} = \left(\frac{n_1}{n_2}\right)^2$

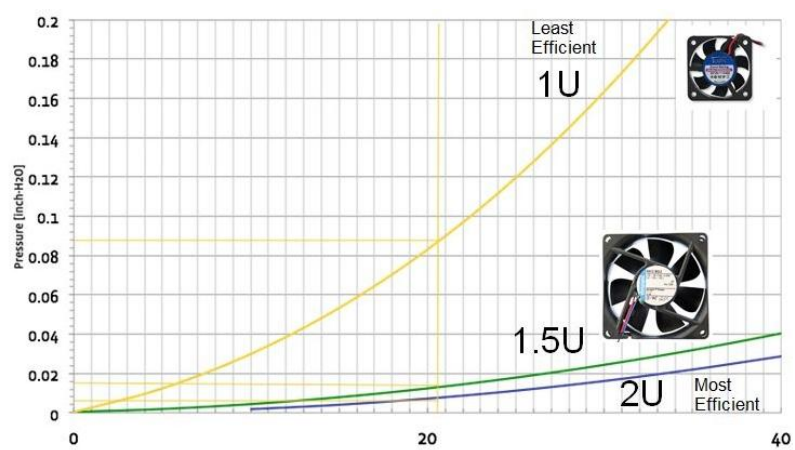
Power $\frac{P_1}{P_2} = \left(\frac{n_1}{n_2}\right)^3$



d=100mm	d=150mm	d=200mm
0.0078mm2	0.0177mm2	0.0314mm2
100.0 Pa	-13.4 Pa	--3.3 Pa

Air volume vs. Power consumption





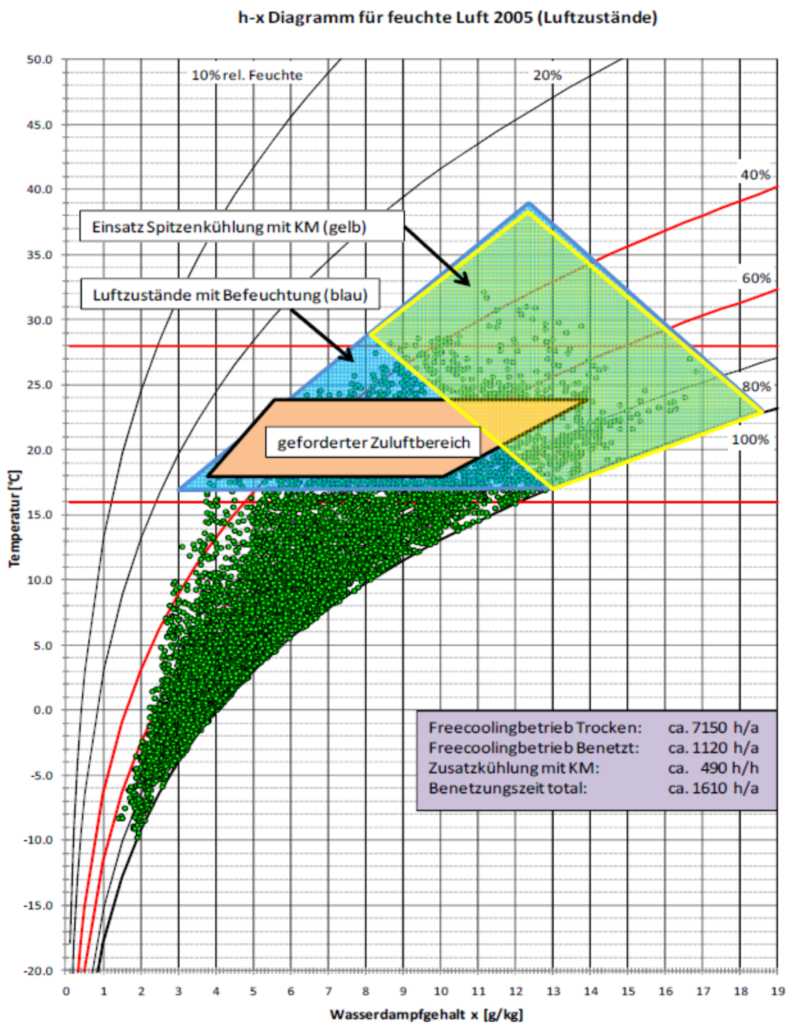
Understanding Data Center Engineering

Physics: Cooling

- Thus we want
  - **High Δt for lower air volume**
    - **Colt / hot aisle separation is a MUST**
  - *Low air speed* for low fan power consumption
    - Less volume or large cross-section
  - *Large fans* for efficient air movement
  - Always *below 2m/s* air or chilled water velocity
    - Energy consumption vs. Equipment cost
  - Higher pump speeds for most efficient operation point
    - Pump design and non-compressible media
  - Optimal balance
    - Capex vs. Ongoing energy cost (Opex)
    - DCF and future energy cost model

Physics: Cooling

- Where to go with the heat?
  - Typically to outside air
    - Dry free cooling
    - Wet free cooling
    - Mechanical cooling
  - Optimize for energy efficiency
  - Chilled water loop system
    - „Split“ system
  - Direct outside air system
    - Large air volume
    - Heavy filtering



Physics: Cooling

- Chilled water loop system
  - It‘s all about Δt
    - The higher the better
    - Small spread from air to water loop
    - Large heat exchanger surfaces
  - Maximise free cooling
  - Freedom to bring chilled water into any part of the building
  - Independence from outside air pollution

- Direct outside air use
  - Constantly large temperature and humidity swings (day/night)
  - Too cold in winter, too hot in summer
    - Requires air re-circulation and adiabatic cooling
  - Heavy air filtration
    - Frequent filter maintenance
    - Good filters have high pressure drop
    - Large air volumes

Permit Process

- Has become very complex
  - Many government offices involved
  - High scrutiny on data center projects
  - Requirements on waste heat re-use
    - Problems: Low quality, lots in summer, less in winter
  - Clear air requirements for generators
    - Additional filters necessary
    - Exhaust stack height a non-obvious problem
  - Environment protection
    - Special fuel storage tank and piping requirements
    - Glycole and refridgerant fluid in cooling loops may have special leak protection requirements
  - Fire marshall
    - Emergency exit distances
    - Hot/cold aisle housing can cause unexpected trouble
  - Building architecture & design
    - Building facade has to be approved by city council
    - Perimeter fences can cause problems

Risk Modeling

- Cost of outages
  - Loss of productivity and business (online service)
  - Image and reputation impact
  - Planned or ad-hoc response
- Mean time to repair
  - Critical replacement parts on-site or made to order
  - The main cost and time sink is bringing servers up again with filesystem/database repair or even restore
- Site risks
  - Natural risks: lightning, heavy snow and rain, river and lake flooding, land slides, earth quakes, meteor impact
  - Infrastructure risk: car or truck burning on street outside, nearby building and wild fires, dust and pollen, railway line accidents, heavy and chemical industry accidents, airport approach/takeoff routes
- Other secondary and indirect risks
  - Extended emergency fuel supply contracts likely worthless

Further items

- Access control
  - Multi-factor biometric
- Fire detection and surpression
  - Early detection systems
  - Various methods of extinguishing
    - Actually not that easy
- Video surveillance
  - IP cameras
    - Backdoor into control network (!)
- Monitoring
  - Detailed power consumption
  - Temperature, humidity, air flow

Sub-optimal practices

- Almost all designs fail concurrent (interruption free) maintenance
  - Insufficient path redundancy
  - Energized cross connects
  - Not enough valves in ring pipe for full isolation
- No cold / hot separation
- Too high air speeds
  - In particular under raised floors
  - Can suck air away
  - Small or crowded raised floors blocking air flow
- No free cooling bypass

Best practices

- Best practice for general purpose data center
  - Large Δt configuration
    - 24°C inlet and 34 – 40°C return air temperature
      - Above 40°C we come into conflict with worker safety rules
    - Minimum Δt 10°C between colt/hot aisle
    - 1:1 delta-t air/water heat exchange with 1-2°C spread
    - Fully regulated large fans with large air-flow cross-section
      - Cold air via raised floor is difficult for high densities
  - Chilled water system
    - Extra-large surface air handler heat exchangers
    - High loop temperatures at 22 – 32°C
  - Free cooling
    - Large surface air handler heat exchangers
    - Mechanical chiller runs only in summer during hot day time
      - Can use simple, inexpensive and „inefficient“ chillers

Best practices:

- Static UPS with batteries
  - Better modular scalability
  - Longer bridge times
  - Compensate distortion, harmonics and power factor
  - Easier to integrate into buildings
- True Tier III or Tier IV design (*not* Tier bla whatever plus)
  - For concurrent maintenance and capacity extensions
  - Up to 1MW Tier IV can be more cost effective than Tier III
  - Have it certified either by Accredited Tier Designer or official Uptime Institute
    - There are many non-obvious mistakes that can be made resulting in loss off concurrent maintainability
    - Almost nobody gets it right on first try (or second)
- You don‘t know what you don‘t measure!!
  - Measure and record everything you can

Recommendations

- Hire an experienced professional

- Read this book
  - I was the technical reviewer and have written preface
  - Title: IT-Räume und Rechenzentren planen und betreiben
  - ISBN 978-3-7640-0553-5, 471 pages

- Read „Data Center Site Infrastructure, Tier Standard: Topology“ from the Uptime Institute
  - URL

Questions?

- Thank you for your attention

- Don‘t hesitate to contact me!
  - Email: oppermann@gmail.com