Introduction to Intelligent Process Control-Seminar: “what to expect from AI technologies”

IPC-Seminar

Holger Schlarb from DESY
Hamburg, 11th of Aug. 2020

Outline:

- Introduction & goal of IPC-seminar ..
- Recap on our accelerators and perspectives
- AI & ML technologies

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

DESY.

My background

issues …

- **Studied physics Ffm & HH with some detour on math. (num. PDE/Distr. Theory)**
- **MSc & PhD in Accelerator Physics in Hamburg (Wakefield comp. / Collimation for TTF)**
- **2003 Sabbatical SLAC (SPPS commission)**
- **2005 XFEL-WP18 Leader "Special Diag."** → 2010 Ch. Gerth
- **2010 XFEL-WP02 Leader "Low Level RF controls"** → 2012 J. Branlard
- **2011 Head of "Accelerator Beam Controls" Group (MSK)**
- **2017 Involved in DESY's scientific computing strategy discussion**

Domain Expert
in acc. Physics

Disclaimer: I am not a computer scientist & not at all in AI !!!!

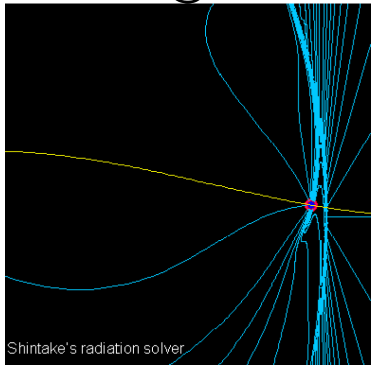
Studied quite a bit (mainly regarding methodologies / technologies)
But: little practical experience
Still in learning phase….

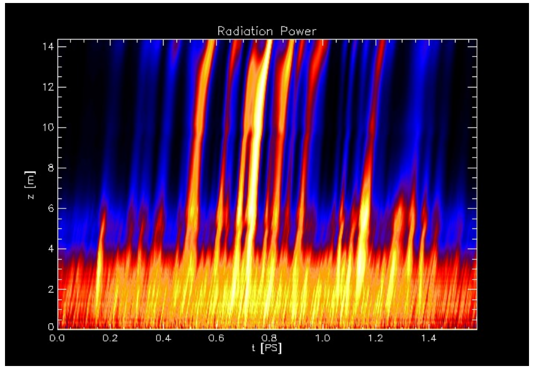
Scientific computing @ M-division

What already exists … & we are pretty strong at

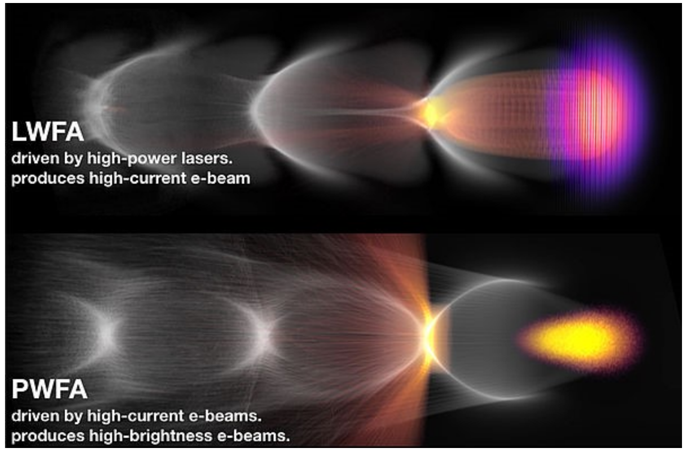
- **Numerical codes for beam dynamic simulations…(ASTRA, CSRTrack,…)**
- **Theory & Simulation of FELs (FAST, Genesis, ALICE)**
- **Electromagnetic fields simulations (e.g. ECHO) + coop. with TEMF (Darmstadt)**
- **Fast tracking & opti. codes for operation (RFTweak, OCELOT,…)**
- **Plasma Wakefield Acceleration …. (HPC required)**

@Dohlus



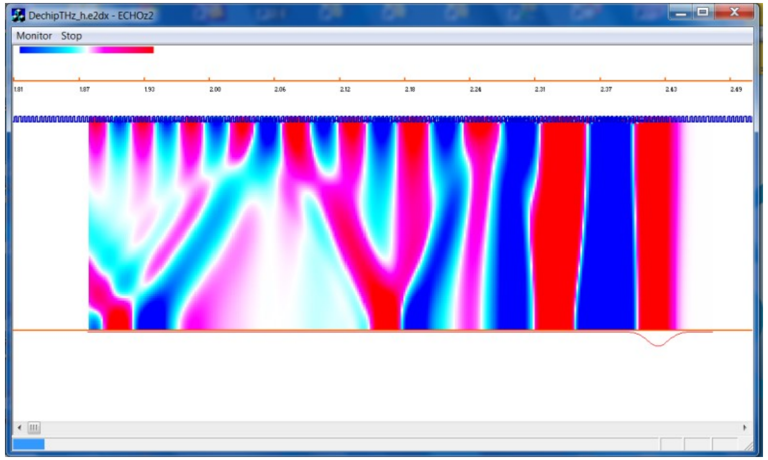


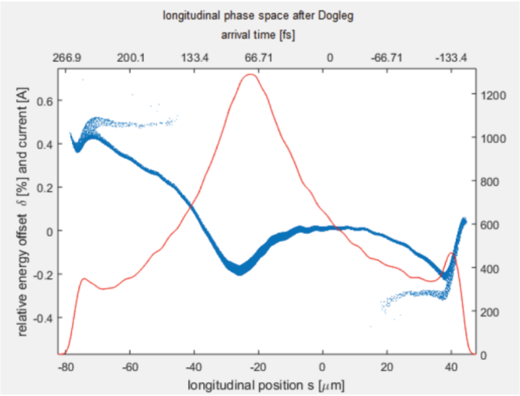
@Reiche (PSI)



@Martinez de la Ossa

@Zagorodnov





@Beutner

➔ Dedicated Sci. Comp. groups in M-Division (well covered)

& seminars: e.g. Beam Dynamics XFEL/PETRA/PWA (Igor/Mikheil/Ilya/Max)

Focus of the IPC-seminar:

Why a new seminar?

Intelligent Process Controls…

➔ **Scientific Computing related to**

- **Data mining and data analysis**
- **Advanced and data driven controls**
- **Fault diagnosis and supervisory control**
- **Automated root cause analysis / causal inference**
- **Sophisticated automation routines**
- **Information extraction technologies**
- **Advanced statistical methods and optimization algorithm**
- **Explore modern AI methods and technologies**

Has both:

Methodological & Technological

aspects!

Ties on different

scientific computing field

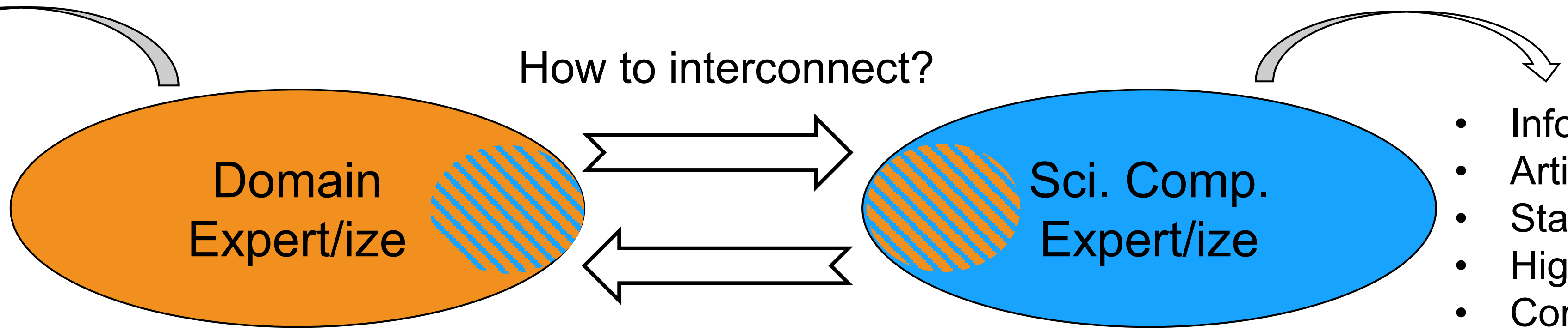
as we are used to.

applied to accelerator design, construction, commission and operation

Interdisciplinary work essential …

Both have to learn….

- Accelerator & Laser Physics
- Instrumentation
- Control systems
- Infrastructures
- RF controls
- …



- Information Technologies
- Artificial Intelligence
- Statistics & Math.
- High Performance Comp.
- Control Theory
- Software architecture
- …

To be recognized: this is a real challenge!!!

- **Different language, notation and annotation + pre-existing knowhow**
- **Different approaches, viewpoints and methodologies**
- **D-Expert: limited method overview / adequate implementation / structural analysis / identify use cases**
- **SC-Expert: formulation of goals, system design & interconnect, may own agenda (method)**

➔ **Buildup knowhow on the respective other domain (specially language)**

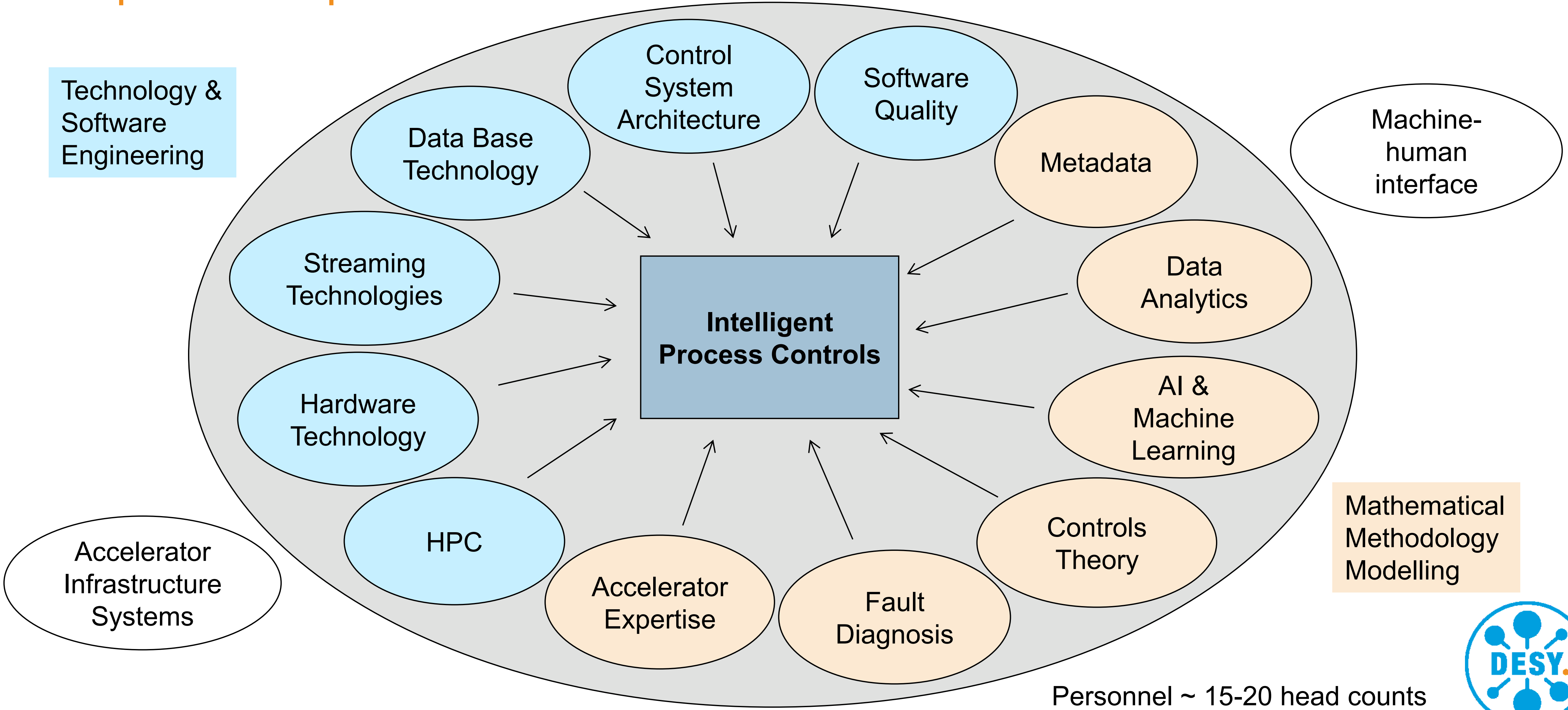
➔ **Buildup teams from both domains**

Goal of the IPC-seminar:

- **IPC-seminar should serve**
  - Bringing the acc. domain expert and computer science experts together
  - Communication platform for all types of SC problems → explore synergies / joint tool development
  - Add tutorial style talks for education and knowledge dissemination
  - Platform to get know to one another (who is working on what with which methods & expertise)
  - Opportunity to present your results or describe your problems
  - Open minded, low barrier … questions and suggests are welcome … new field for many of us
- **SC@M: concerns M-division as whole / can provide from:**
  - Across all accelerator & laser system & test facilities
  - Across M-division groups
  - Currently virtual group in M-division with members in different groups

Intelligent process control of distributed systems

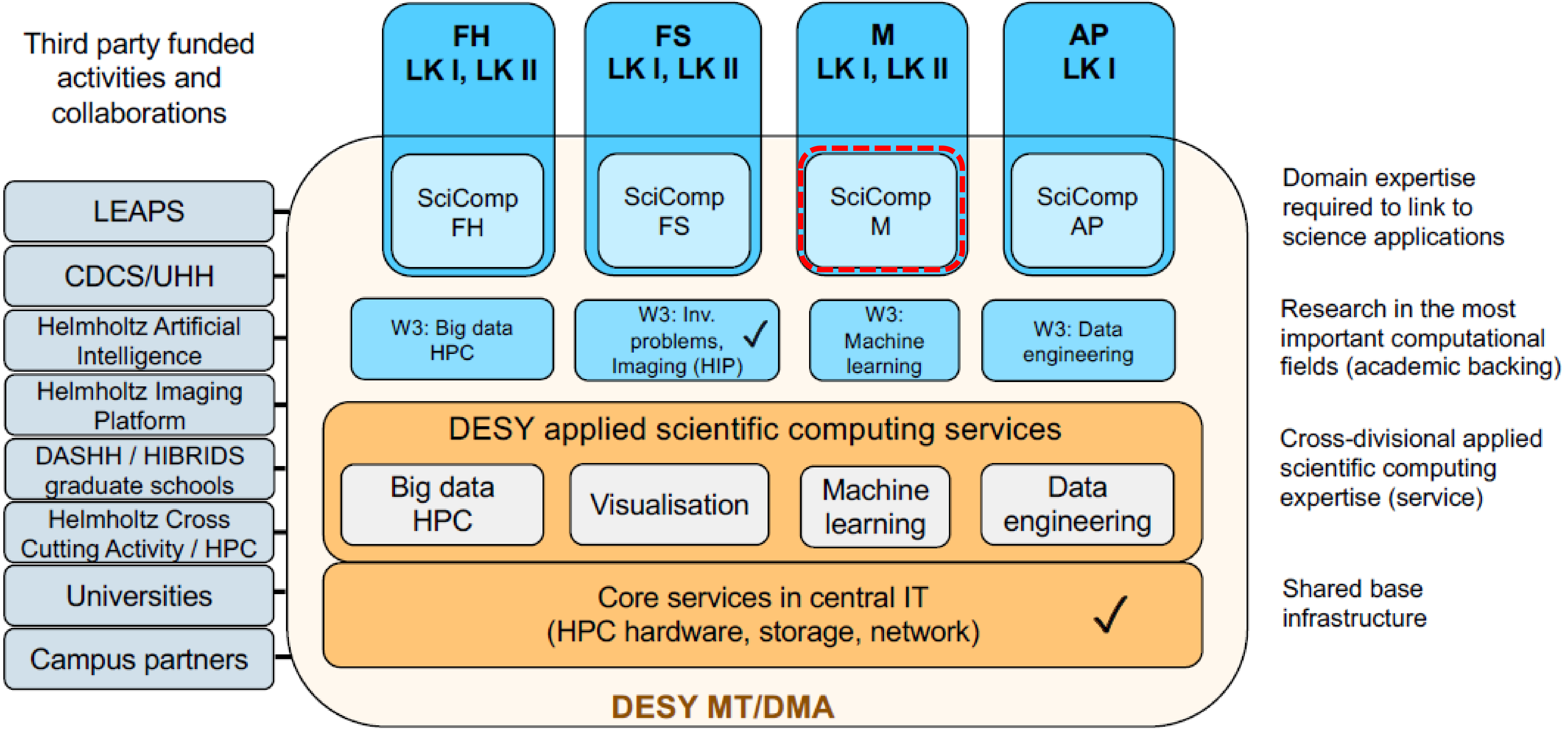
… requires wider expertise …



Personnel ~ 15-20 head counts
LKI/LKII & 3rd party funds

Data and computing science – strategy update

First steps towards a DESY wide scienctific computing strategy



DMA = Data Management Analysis

Short recap:

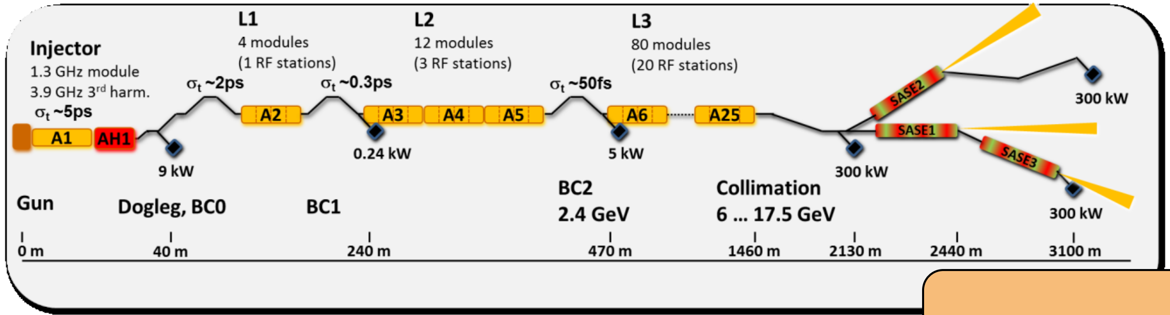
DESY’s Accelerators & test facilities

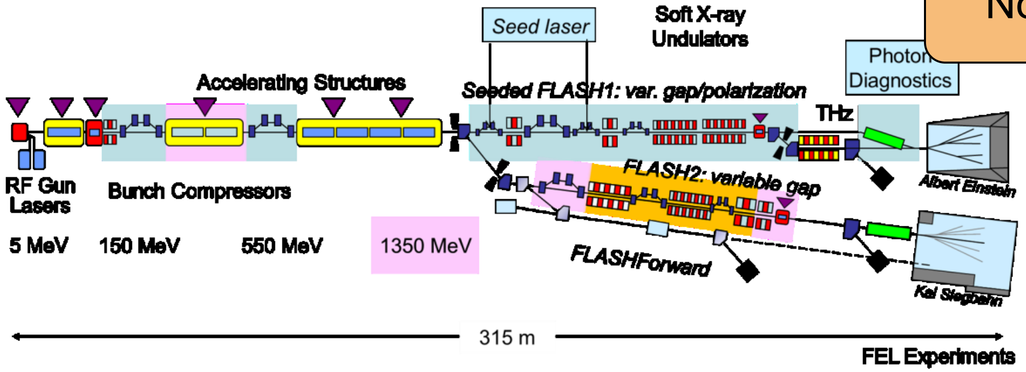
DESY accelerators & test benches: birds view …

Different accelerators … different challenges

Accelerators at DESY …

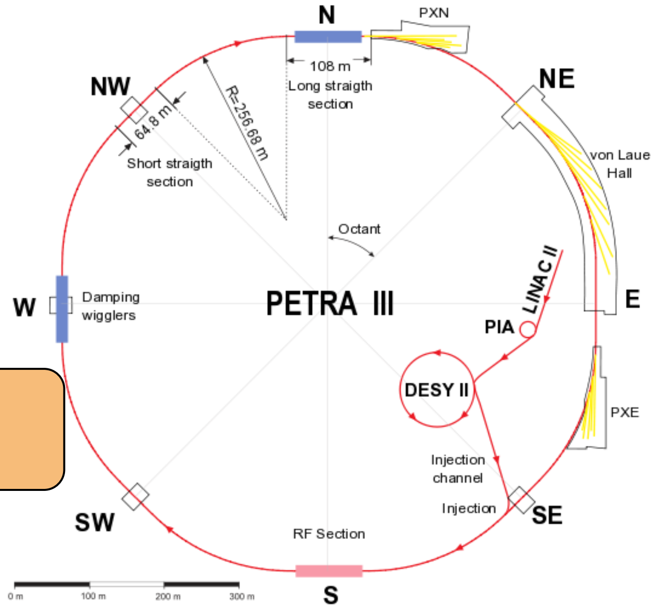
FELs…. (reproducibility / optimization / flexibility / …)





No equilibrium

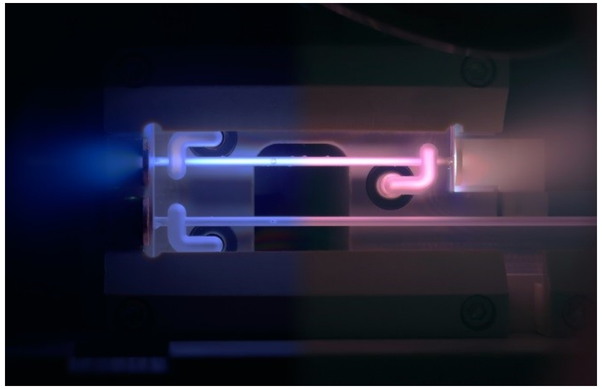
Storage ring (lattice / beamlines / optic &orbit / injection / availability/...)



Equilibrium but non-linear

Highly non-linear

Plasma accelerators (lasers control/ HPC /in-situ FB)



Challenges and degree of matureness very different, but

→ Increased complexity of controls

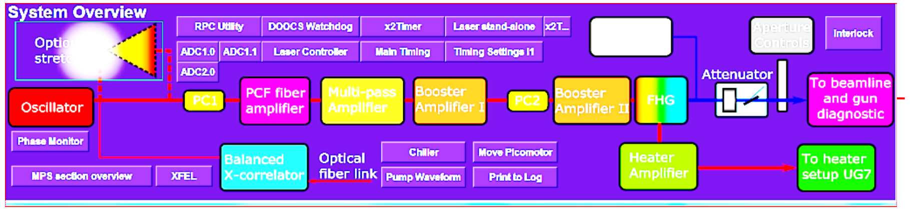
→ Higher demands on accelerator operations

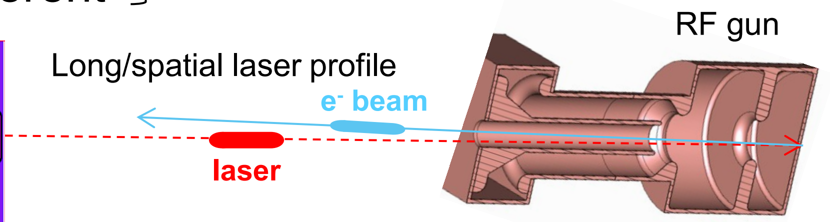
→ Push on the limits regarding performance & flexibility & availability

Beam dynamics of an FEL drive beam

issues …

- **First of all:** it is a linear accelerator! → no equilibrium state
→ every shot is different — High control effort!!!
- **Photo injectors:**





- **Complex beam transport** (up to second order)

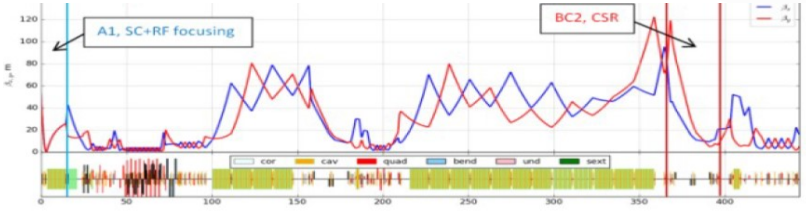
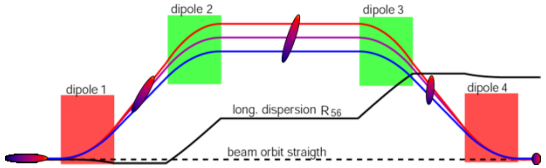


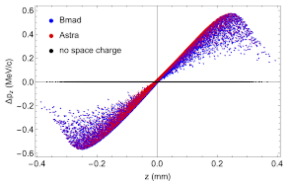
Figure 1: European XFEL lattice up to BC2.

- **Bunch compression**

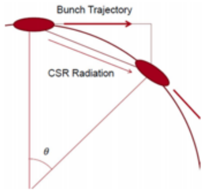


- **Collective effects**
  - Space charge
  - Coherent Synchrotron Radiation
  - Wake fields

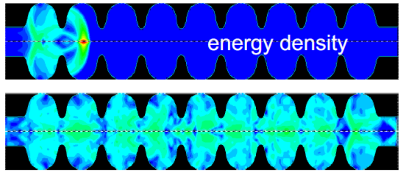
Space charge



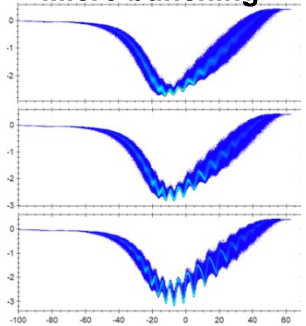
Coh. Synchr. Rad.



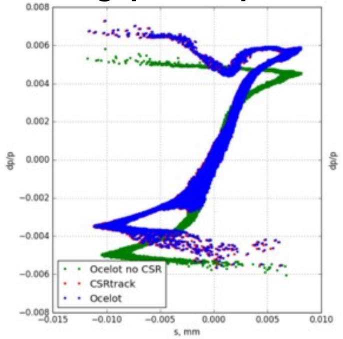
Wake fields



Micro bunching



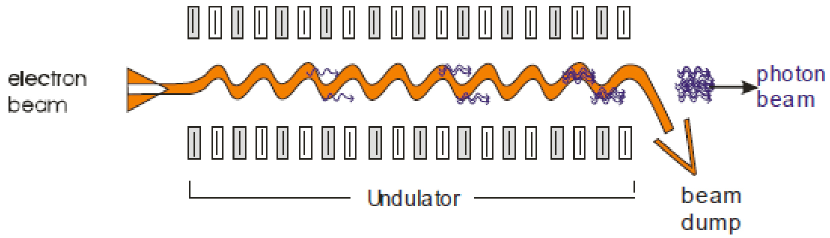
Long. phase space



- **Micro bunching instability …**
- **… many more**

Beam dynamics of an FEL drive beam

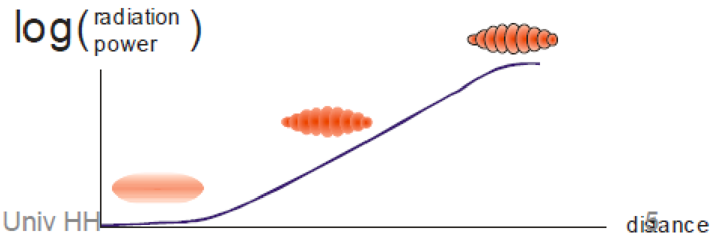
Self-Amplified Spontaneous Emission (SASE)
(Kondratenko, Saldin 1980)
(Bonifacio, Pellegrini 1984)



Resonance wavelength:

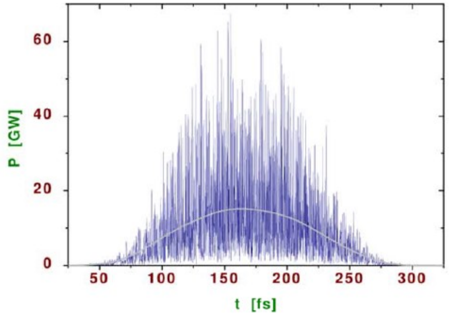
$$\lambda_{ph} = \frac{\lambda_u}{2\gamma^2}\left(1 + \frac{K^2}{2}\right)$$

Undulator parameter ≈ 1



ARD-ST3 workshop DESY-Zeuthen, 2017   Jörg Rossbach, Univ HH

Stochastic process



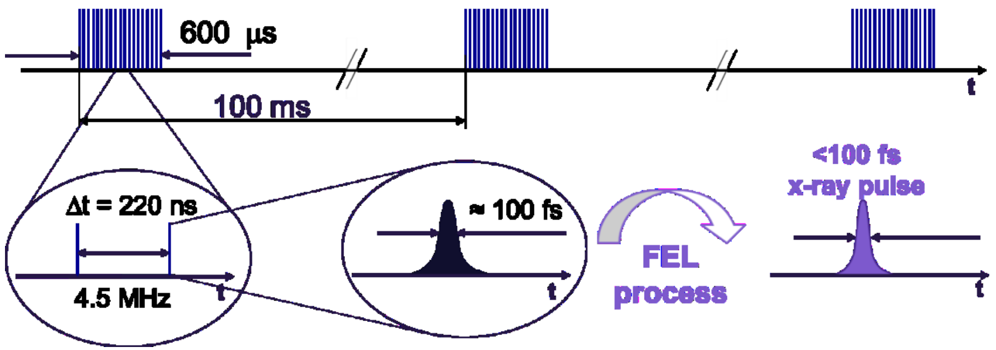
L ~ 200m

- ➔ Beam parameters challenging to achieve
- ➔ Control over 6D phase space mandatory
- ➔ kA peak currents causing strong collective effects

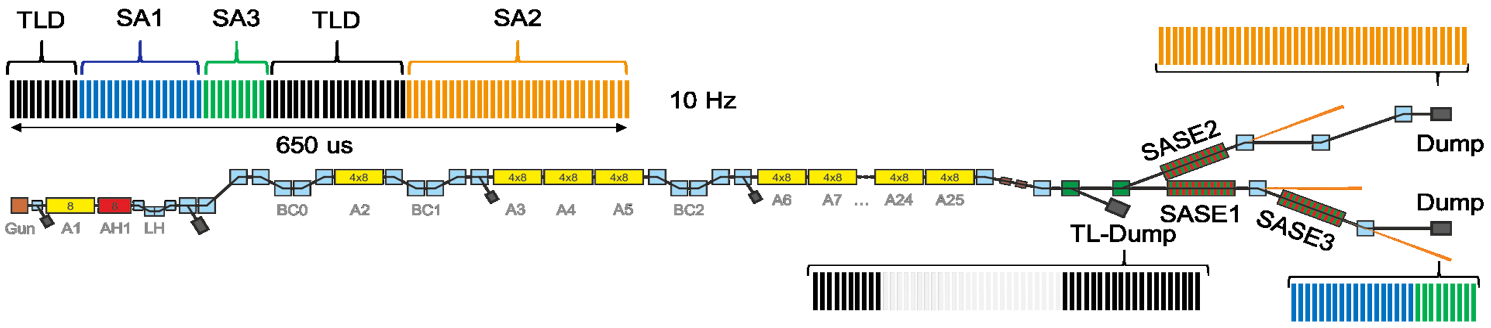
Wide dynamics range & different operation modes

Time structure & multiple beamline (SA1/SA2/SA3)

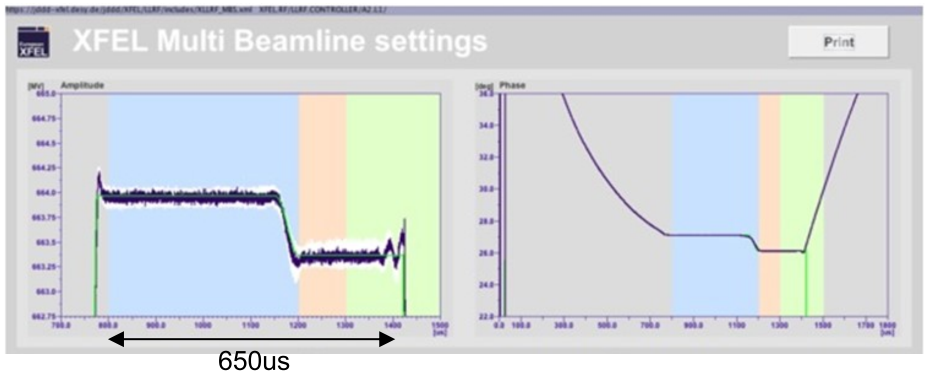
- **Pulsed operation (RF ~ 1.4ms, 10Hz)**
  - ➔ 27000 / sec
  - ➔ e- bunches 220ns spaced
  - 100ms separation



- **Bunch distribution & pattern**
  - Fast Kicker System
  - Precise (Slow) Flat-Top Kicker
  - 5/10Hz splitting



- **Change of beam properties for different beam lines**
  - Transients processes
  - Subsystems in different states
  - Large reconfigurations



Challenge:
3 machines in one…..
Coupling…

HELMHOLTZ
SPITZENFORSCHUNG FÜR GROSSE HERAUSFORDERUNGEN

Large and complex Infrastructure & Subsystem

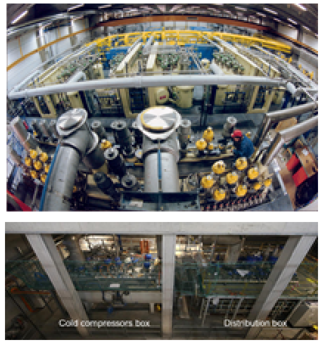
Infrastructure systems (Cryo/Cooling/Power) & large and diverse components ~ 10000 typically

Infrastructure

Large number of sensors & components

Large supply infrastructures: Cooling Systems/Power Distribution/Climatisation/Cryogenic system …

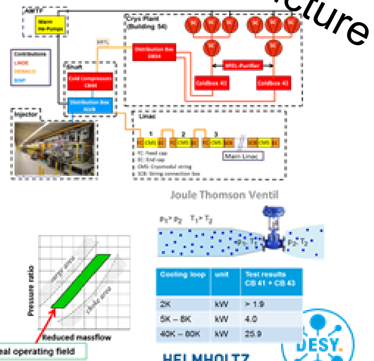
Here at the example of the Cryogenic-System:



XFEL cryogenic system:

- 671 control valves
- 2647 temperature sensors
- 800 pressure sensors
- 212 flow sensors
- >100 level sensors
- 433 regulation loops
- > 22000 records
- > 220000 properties

4-stage cold compressors:



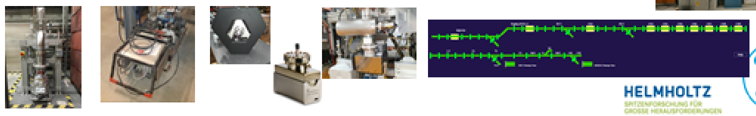
Similar complexity for water cooling & power distribution…

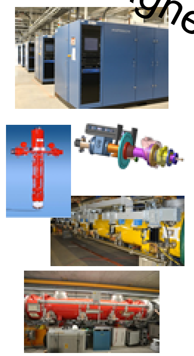
Magnets & RF

Large number of sensors & components

Beam line devices …

Magnets	103 Dipoles, 495 Quads, 59 Multipoles, 403 Correctors, 103 Quad-Movers, 2 Solenoids
Modules	101 x 1.3 GHz, 1 x 3.9 GHz, 27 RF Stations 808 SRF cavities & HP Coupler & HOMs …
Fast Devices	34 kicker magnets, 3 transverse deflecting structure
Undulator	1 Laser Heater Undulator, 91 SASE Undulator Segments
Vacuum	Total 4400 m of cold & warm vacuum in various sections: beam (1), coupler (99), laser (2), dump (3), iso (21) adds to 760 sputter ion pumps /140 titan sublimation pumps/ 40 pump carts/ 54 valves / 8 fast shutter / + vacuum gauges
	And many more …………………

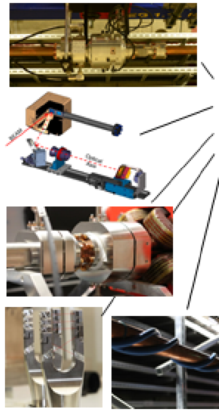




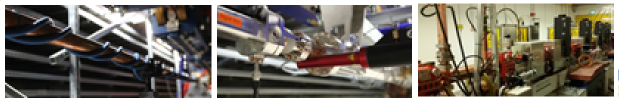
Diagnostics

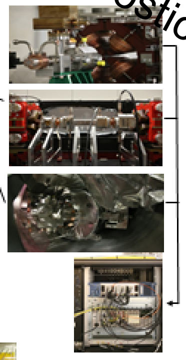
Large number of sensors & components

For beam diagnostics … mostly bunch-to-bunch resolution required



Installed Diagnostics Items	Number
Beam Position Monitors	453
Charge Monitors	51
Imaging stations	67
Dark current monitors	9
Wire Scanners	12
Loss Monitors	474
Dosimetry Systems	630
Transverse Deflecting Structures	2
Bunch Compression Monitors	4
Beam Arrival Time Monitors	7
Electro-Optical Systems	3
THz spectrometer	1

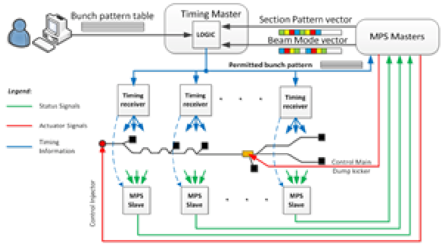




Machine Protection

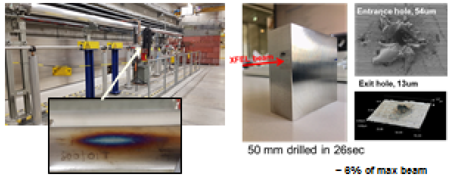
High power beam & beam density

Machine protect to prevent acc. & photon sub-system damages

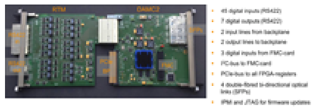


e⁻ beam beam …

Photon beam …



- 160 collectors boards
- ~ 4000 different signals (magnets/BLMs/toroids/vacuum/ diagnostic/couplers/waveguides/..



→ up 500kW electron beam power

→ Damages within ~ few us possible

Challenge:

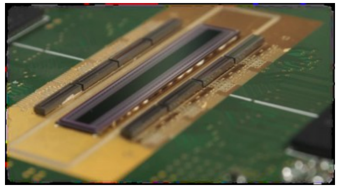
- Large number of different inputs
- Fast reaction time ~ us
- Re-configurable
- Not too restrictive … but still save

Changes due to digitalization…

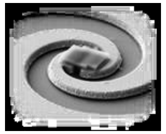
Driven by performance requirements and needs for higher flexibility…

Very complex sensor technologies …

40 GHz pickup

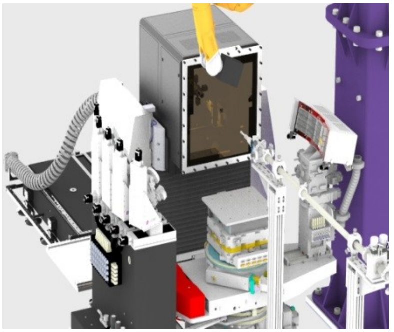


THz det.

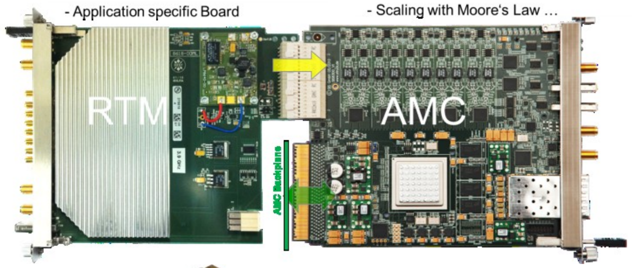


1D detector KALYPSO

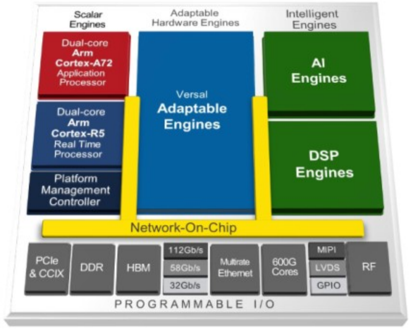
2D detectors / high-speed cameras



& analog / digital processing units ….

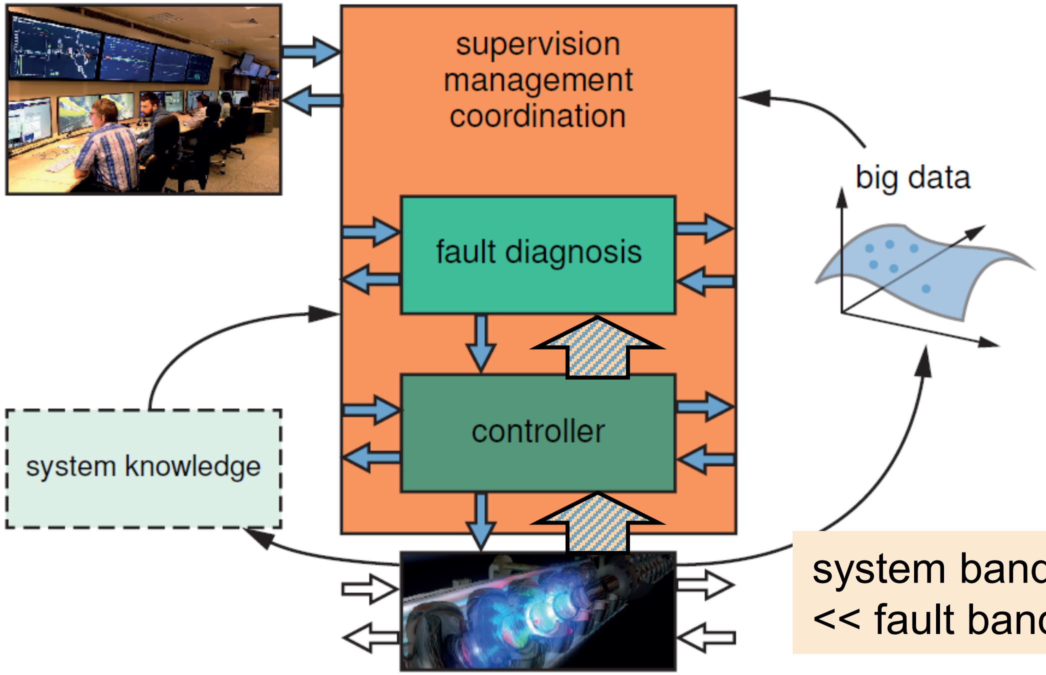


…where next evolution is already on the way (Network on Chip = NoC)



MicroProc. → DSP's → FPGA's → SoC → MPSoC → NoC

Separation from human controls …



system bandwidth
<< fault bandwidth

Courtesy: A. Eichler

- Configuration → data quality
- Demand on supervisory systems
- New HW technologies

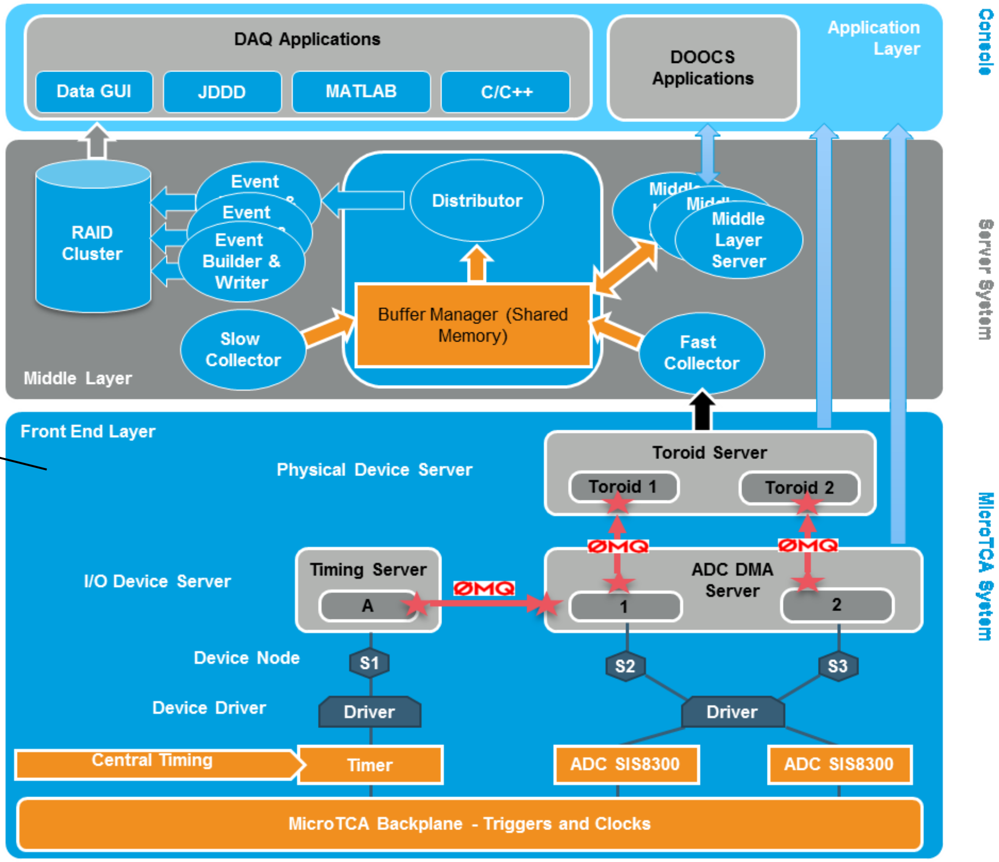
What changed over the past decade…

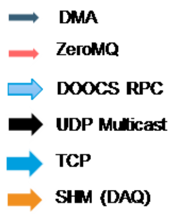
Digitalization changes …

- **Classical 3 tier structure + in cooperated Data Acquisition System at EuXFEL**

Standardize hardware & software

MicroTCA .4





tesy: T. Wilksen

- > **10 million control parameters**
- > **700.000 local archives**
- > **20.000 high data rate channels**
- > **30 TB/day DAQ** (compressed)

➔ **Configuration management** already challenging

➔ **Large amount of data available**
- data mining
- information extraction
- machine / system optimization
- failure detection
- predictive maintenance
- …

We would like to better make use of this data

➔ **< 1% sent from front-ends**

- Requires new techniques & technologies → AI/ML & automation
- Integration into controls architecture → needs carefully to be reviewed

What makes these machine complex …

Challenge caused by combination of performance / flexibility & availability…

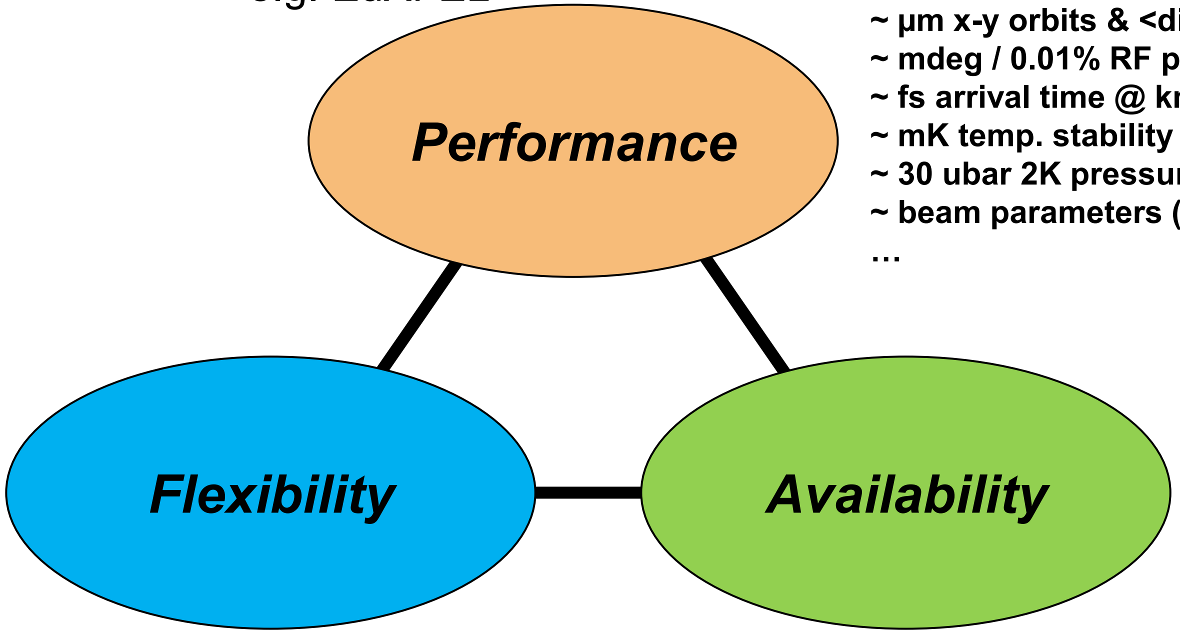
Particle accelerators are ...

- Largely distributed …
- Various type of systems …
- Highly non-linear processes …
- Strongly coupled (sub-)systems …
- Timing varying disturbances …
- High dimensionality ~ few 1000
- Use of cutting-edge technologies… (sensors / front-end electronics / real-time processing)
- Limited access to key observables

Facing:

Increasing demand on **Performance, Flexibility & Availability**

e.g. EuXFEL



~ µm x-y orbits & <dispersion
~ mdeg / 0.01% RF phase/ampl.
~ fs arrival time @ km scale
~ mK temp. stability in parts
~ 30 ubar 2K pressure
~ beam parameters (λ,σ_t, E_{ph}, chirp..)
…

- **Bunch pattern (flexible, on demand)**
- **Switching bunch pattern modes (10Hz)**
- **Multi-beamline operation**
- **Fast tuneability within trains**
- **Special modes**
- **…**

- **Ideally 99%**
- **Small setup times (reproducibility)**
- **Small tuning times (operability)**
- **Reduced scheduled down**
- **…**

Future operation & future machines …. perspectives

Questions

- can we tailored beams with **high precision** for user experiments
- how far can we push accelerator **stability** and **reproducibility**
- how can we **predict control** over sufficiently long time horizons
- can we develop **autonomous accelerator** operation
- can we establish new user-modes to **scan the right physics**
- can we build accelerators that **predicts failures** and **repair itself**
- can we **enable new accelerator designs** not accessible today

Performance

Flexibility

Availability

by applying Advanced Controllers, Data Analytics and AI technologies for *intelligence process control* of accelerators

Artificial Intelligence…

What is artificial intelligence …

Some definition …

Artificial intelligence is the attempt to simulate rational or cognitive human intelligence on (technical) machines in order to use it in a profitable and beneficial way for humans @Ralf Otte

Classification:

1. Appropriate intelligence - adequate reaction to environmental stimuli (I1)
2. Learning intelligence - independent acquisition of new knowledge (I2)
3. Creative intelligence - generation of knowledge outside of formal induction and deduction (I3)

schwache KI

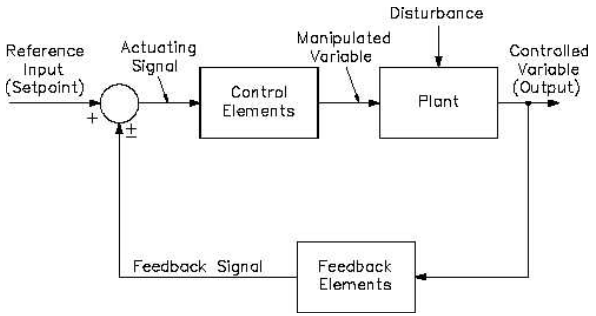
---

4. Conscious intelligence - Conscious understanding of knowledge (I4)
5. Self-confident intelligence - conscious understanding of the “ICH” concept (I5)

starke KI

Still science fiction

Example I1:



Example I2: Technical AI systems that are able to apply complex knowledge and at least partially autonomously generate new knowledge

- speech processing systems,
- image recognition systems,
- expert systems,
- rule-based systems,
- automatic evidence systems,
- …

Remark: AI-definition in Europe (how its done) differs from Anglo-Saxon (what can it do)

Methodology to gain knowledge

Some definition …

Deduction: ("Inferring Forward", "Applying Forward" a rule base)

- Premise (axiom) + inference rules (known knowledge) → rules (new knowledge)
- Cause (observed facts) + rules of conclusion (known knowledge) → effect (new facts)

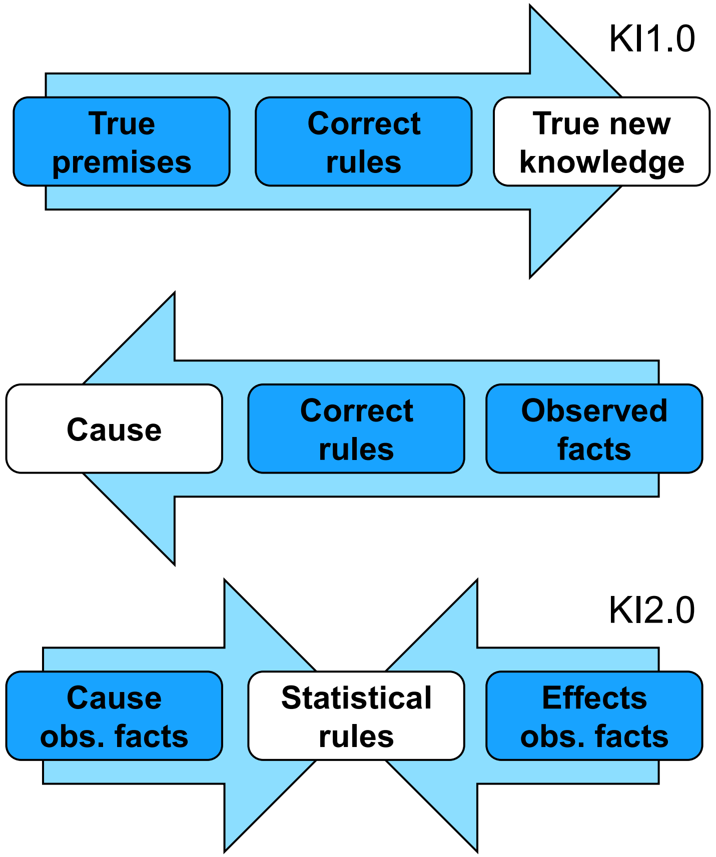
E.g. Statement logic, predicate logic PL1/2, fundamental nature / engineering laws, math. contruction..

Abduction: ("inferring backwards", "applying backwards" a rule base)

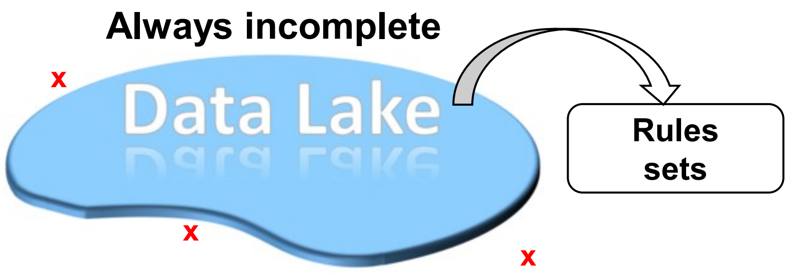
- Effect (observed facts) + rules of conclusion (known knowledge) → cause (new facts)

Induction: (»learning from data«, creating a non-generally applicable rule base)

- Cause (observed facts) + effect (observed facts) → rule (new statistical knowledge)



Remark: in math.
full induction is correct since complete!

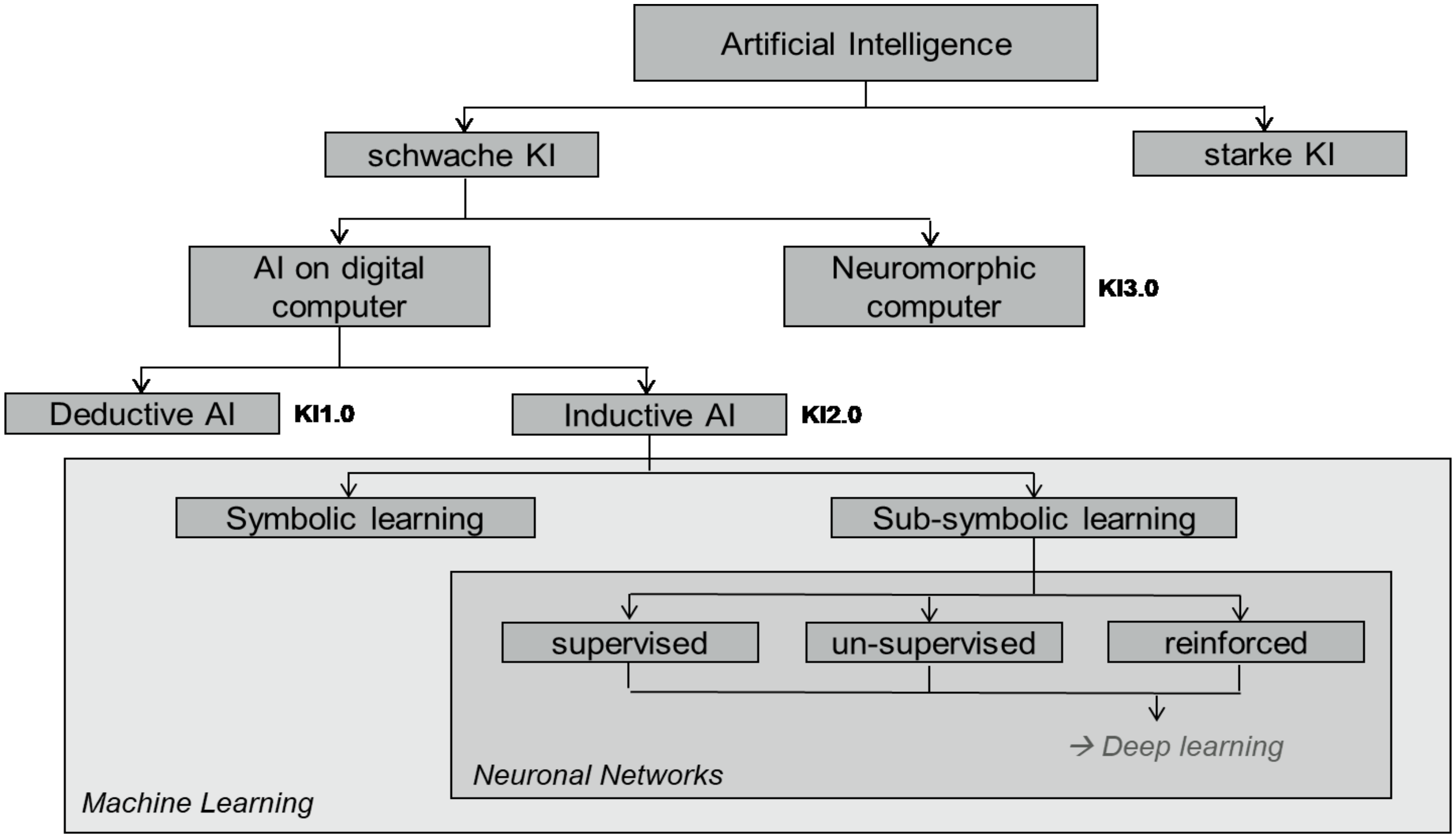


➔ Only of statistical nature
➔ Cannot extrapolate! x
➔ Generalization … good interpolation

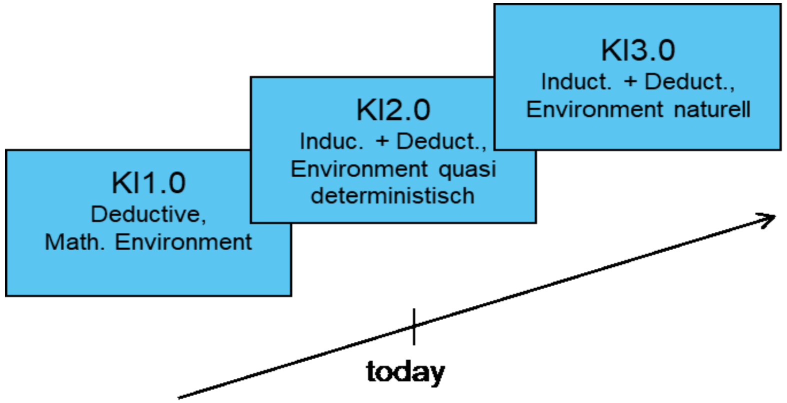
Safety systems:
NO GO!

AI landscape

A possible way of categorization …



@Ralf Otte



(EN: narrow / general, super
ANI / AGI , ASI)

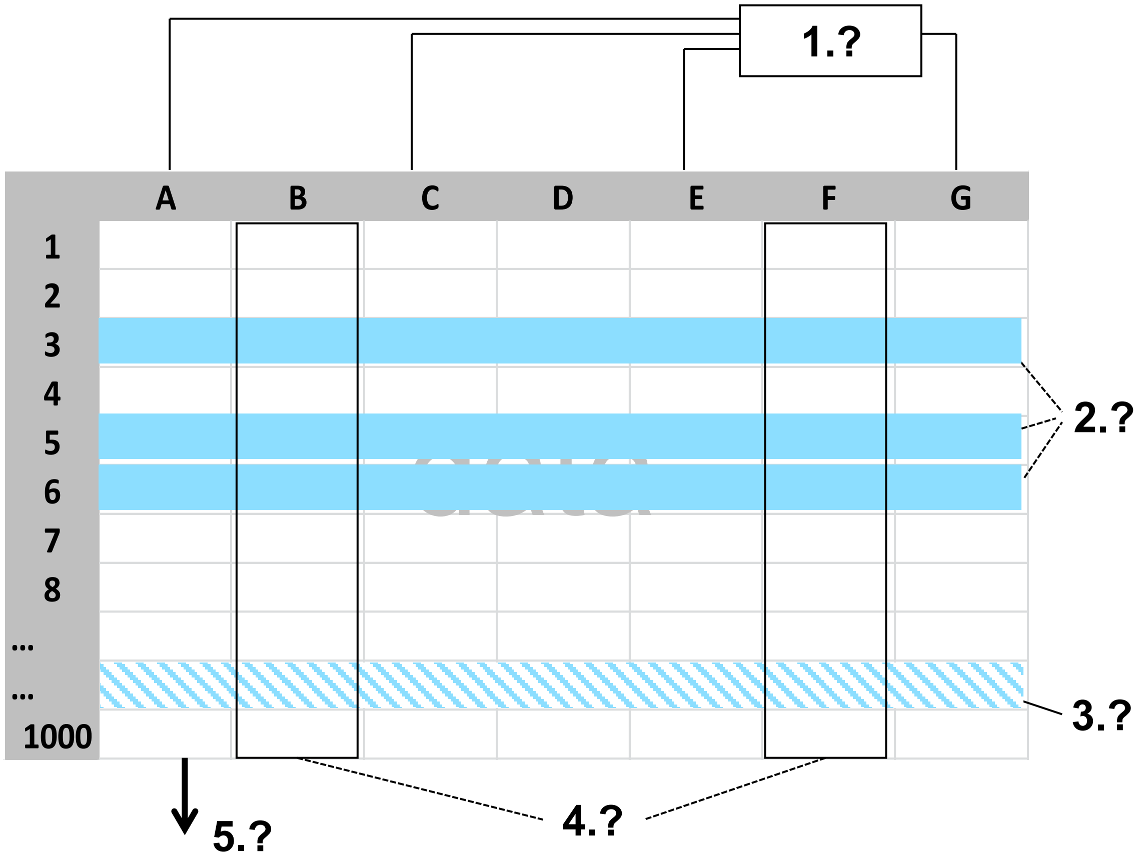
← electronic analog circuits to mimic neuro-biological architectures

Big hype:
Knowledge Discovery in Data Base (KDD)

Supervised: Y exist
Un-supervised: no Y
Reinforced: receives reward

What can you do with data using ML

issues …



Columns: Variables A….G (called features)
Row: set of variable

1. Columns: G = f(A,C,E) ask for functional relation (regression) variable to each other?
   → directed relation
2. Rows: how many clusters exist?
   → structural analysis
3. Individual rows: affiliation to a cluster?
   → structural analysis
4. Individual columns to one-another: correlation analysis
   → undirected relation
5. Succeeding rows: prediction for following rows
   → trend analysis

ML: first thing you learn …linear regression…

which is kind of standard for natural scientist & engineers…

Best estimate for $y = f(\mathbf{x})$:

columns: *input features* x_n

$x \in \Re^N$; $f: \Re^N \to \Re$

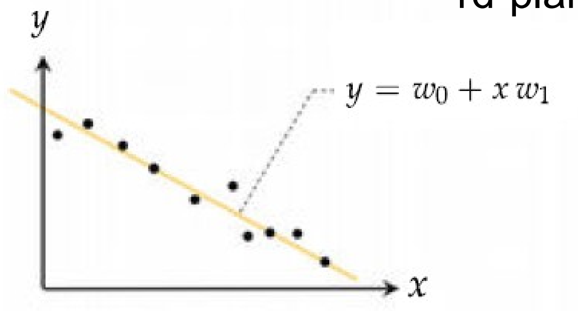
$$\boxed{w_0 + x_{1,p}w_1 + x_{2,p}w_2 + \cdots + x_{N,p}w_N \approx y_p, \quad p = 1,\ldots,P.}$$

Rows of data sheet

bias

feature touching weights w_n

1d-plane



Least Square cost function $g(\mathbf{w})$:

Metric required (here just l_2: p $\|\cdot\|_2^2$)

use x_0=1

$$\underset{\mathbf{w}}{\text{minimize}} \ \frac{1}{P}\sum_{p=1}^{P}\left(\mathring{\mathbf{x}}_p^T\mathbf{w} - y_p\right)^2$$

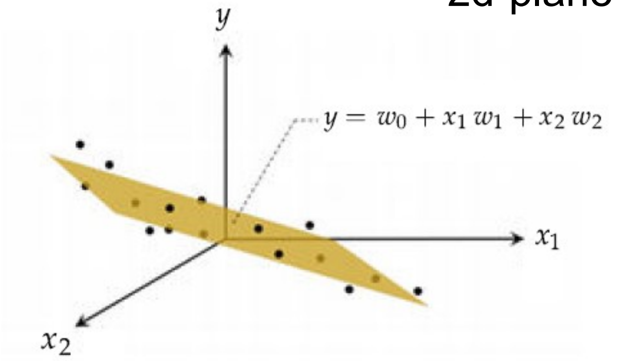
➔ Local optimization problem (convex cost function)

Hessian of $g(\mathbf{w})$ thus $H_{ij}(g(\mathbf{w})) = \partial_i \partial_j\, g(\mathbf{w})$

$\nabla g(\mathbf{w}) = 0$ ➔ $$\left(\sum_{p=1}^{P}\mathring{\mathbf{x}}_p\mathring{\mathbf{x}}_p^T\right)\mathbf{w} = \sum_{p=1}^{P}\mathring{\mathbf{x}}_p y_p$$

Can be solved in one Newton step

2d-plane



Remarks:

- Concept easily extends to non-linear regression (add new features $x_i^n\,x_j^m$)
  ➔ if feasible **preferred way**, coefficients have defined meanings
- l_2 norm overfits outliers → l_1 "last absolute deviation" $|\cdot|$
- For larger N … gradient descents method used to compute **w**

ML: short detour to “Optimization Problems”

most cases finally end up as…

General formulation:

nonlinear program

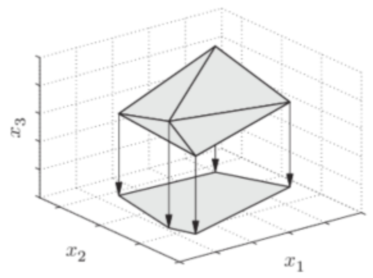
$$\begin{array}{lll} \inf_z & f(z) & \\ \text{subj. to} & g_i(z) \le 0 & \text{for } i = 1, \dots, m \\ & h_j(z) = 0 & \text{for } j = 1, \dots, p \\ & z \in Z, & \end{array}$$

$f : \mathbb{R}^s \to \mathbb{R},\ g_i : \mathbb{R}^s \to \mathbb{R},\ h_i : \mathbb{R}^s \to \mathbb{R}$

Cost function *f,g,h* might be explicit, implicit or as oracles

Constrains

Goal to find z* “the optimizer”



e.g. polytope for linear inequalities

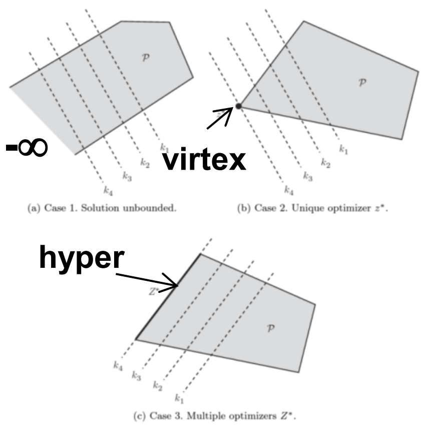
Convex: $\lambda z_1 + (1-\lambda) z_2 \in S \qquad f(\lambda z_1 + (1-\lambda) z_2) \le \lambda f(z_1) + (1-\lambda) f(z_2)$ ➔ Global optimum $f(z^*) = f^*$

linear program

$$\begin{array}{ll} \inf_z & c'z \\ \text{subj. to} & Gz \le w \\ & Az = b, \end{array}$$

$G \in \mathbb{R}^{m \times s},\ w \in \mathbb{R}^m$

$A \in \mathbb{R}^{p \times s},\ b \in \mathbb{R}^p,$

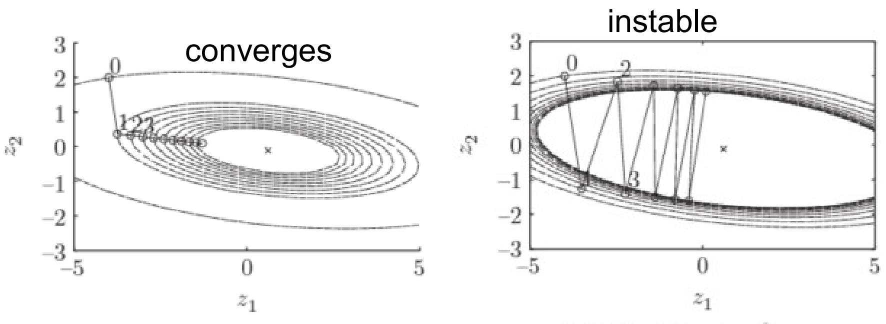


quadratic program

$$\begin{array}{ll} \min_z & \frac{1}{2} z' H z + q' z + r \\ \text{subj. to} & Gz \le w \\ & Az = b. \end{array}$$

$z \in \mathbb{R}^s,\ H = H' \succ 0 \in \mathbb{R}^{s \times s},\ q \in \mathbb{R}$

Unbound: $\boxed{z^* = -H^{-1} q}$



(a) GM. Step size $\frac{1}{L}$. (c) GM. Step size $\frac{2}{L}$.

H^{-1} computational expensive ➔ gradient method
(obviously high order methods are more efficient)

Initial iterate $z^0 \in \mathbb{R}^s$

$$z^{k+1} \leftarrow z^k - \frac{1}{L} \nabla f(z^k)$$

ML: Nonlinear Feature Engineering…

First steps towards universal approximators…

Inputs (features)

But use now non-linear fct.

$$\mathrm{model}\left(\mathbf{x},\Theta\right) = w_0 + f_1\left(\mathbf{x}\right)w_1 + f_2\left(\mathbf{x}\right)w_2 + \cdots + f_B\left(\mathbf{x}\right)w_B$$

Represent all possible parameters (weights), also inside fct.

Generalization:

Degree of freedom &
Structure correctly chosen

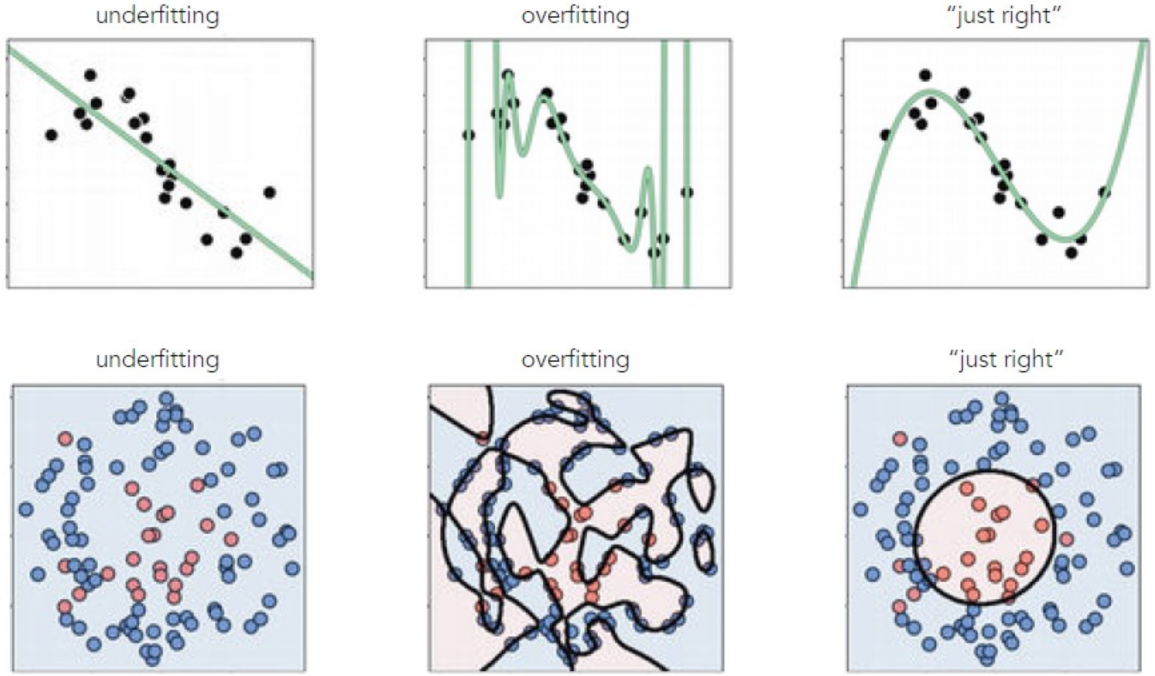
- Wavy appears ➔ use trigonometric functions

$$f_1\left(\mathbf{x}\right) = \sin\left(v_{1,0} + v_{1,1}x_1 + v_{1,2}x_2\right)$$
$$f_2\left(\mathbf{x}\right) = \sin\left(v_{2,0} + v_{2,1}x_1 + v_{2,2}x_2\right)$$

- Elliptical boundaries

$$\mathrm{model}\left(\mathbf{x},\Theta\right) = w_0 + x_1w_1 + x_2w_2 + x_1x_2w_3 + x_1^2w_4 + x_2^2w_5.$$

- Neural network family $f_b(x) = \tanh\left(w_{b,0} + w_{b,1}x\right)$

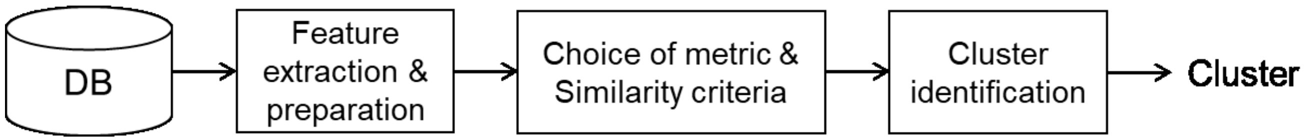


ML: clustering methods

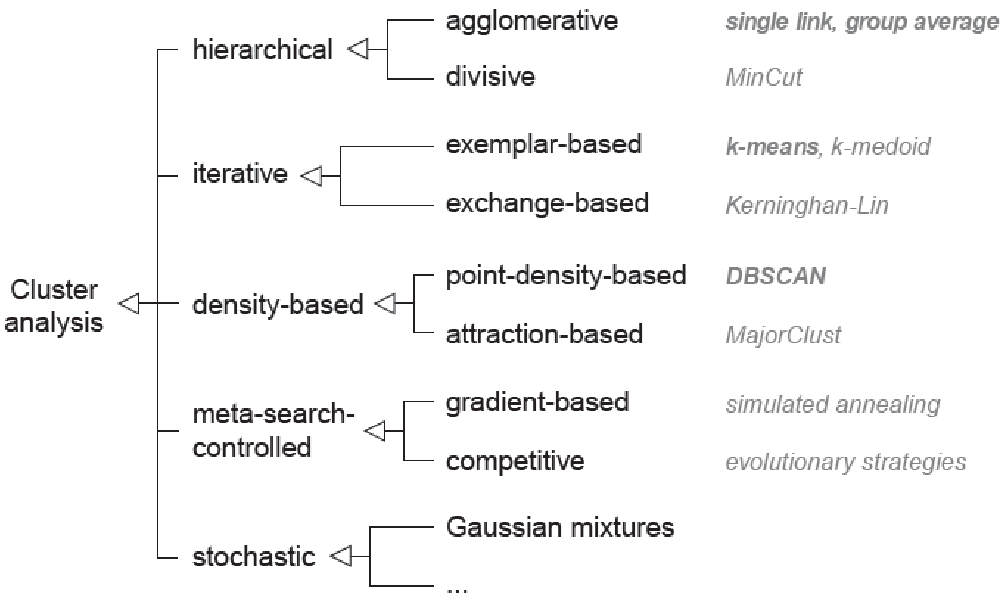
… for data structure analysis, outlier identification, …

Basic sequence of a cluster analysis

Unsupervised!



Classification scheme



➔ Likely dedicated talk/tutorial on clustering methods

Test cases using DBSCAN

Density-Based Spatial Clustering of Application with Noise

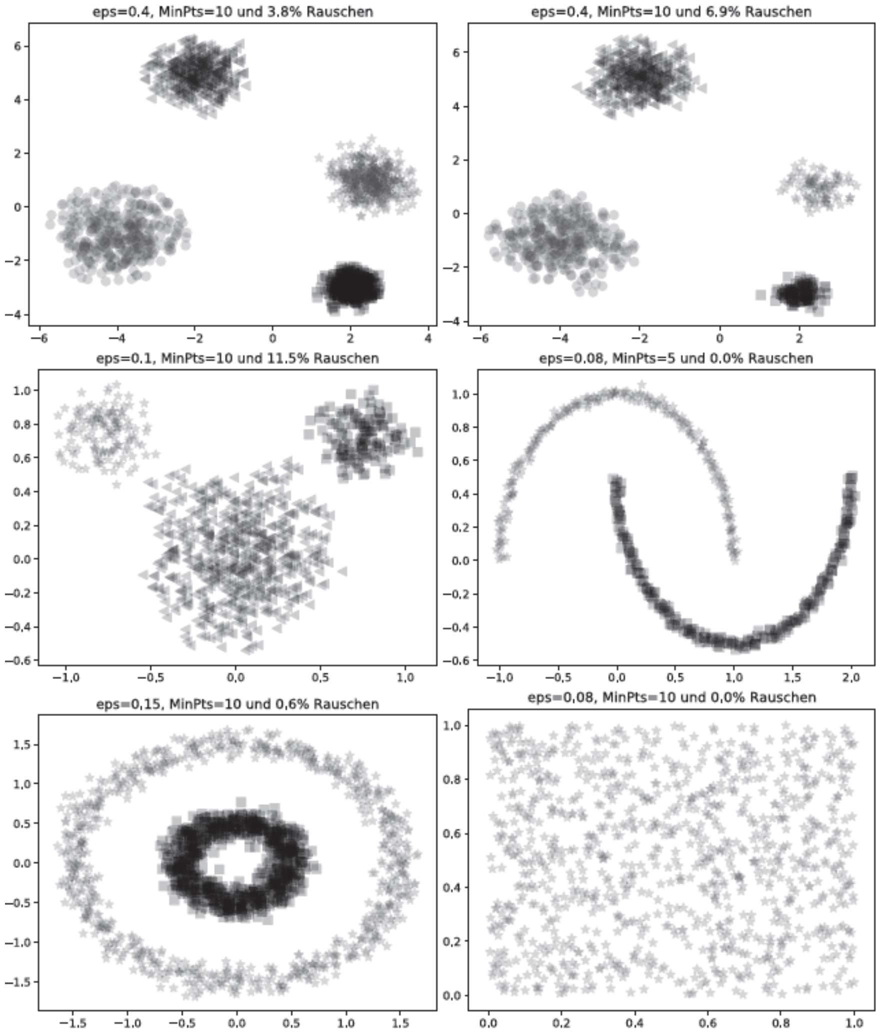


Abbildung 11.10 Clusterbildung mit DBSCAN auf den Testproblemen

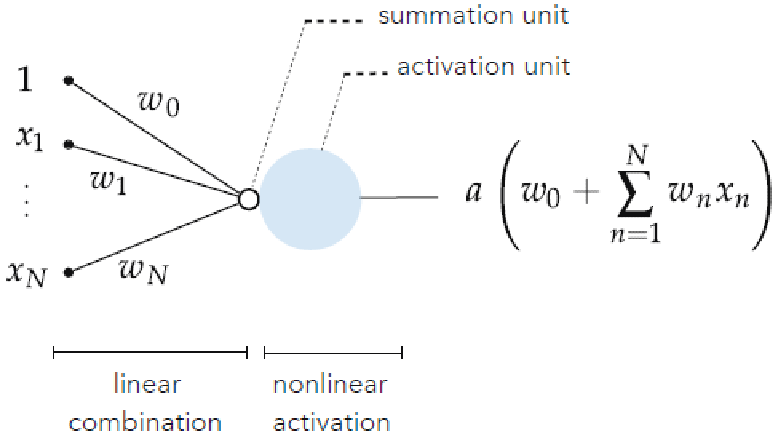
@Frochte

Artificial neuronal networks as universal approximators

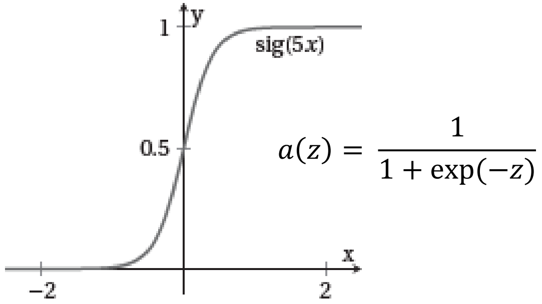
Basic intro…

- **Single neuron: algebraic represented by**

$$f^{(1)}(\mathbf{x}) = a\left(w_0^{(1)} + \sum_{n=1}^{N} w_n^{(1)} x_n\right)$$



a(·) e.g. sigmoid



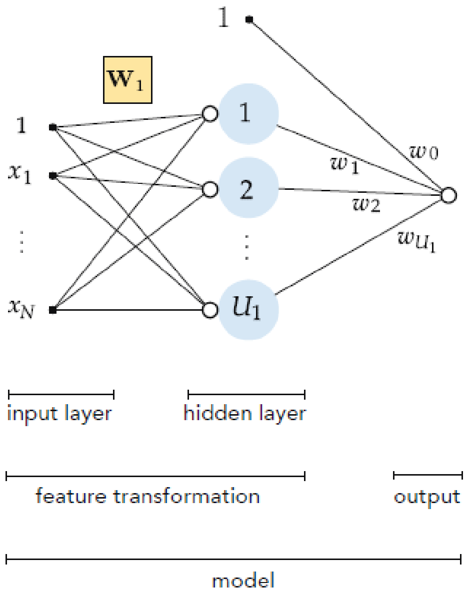
- **1 hidden layers: algebraic represented by**

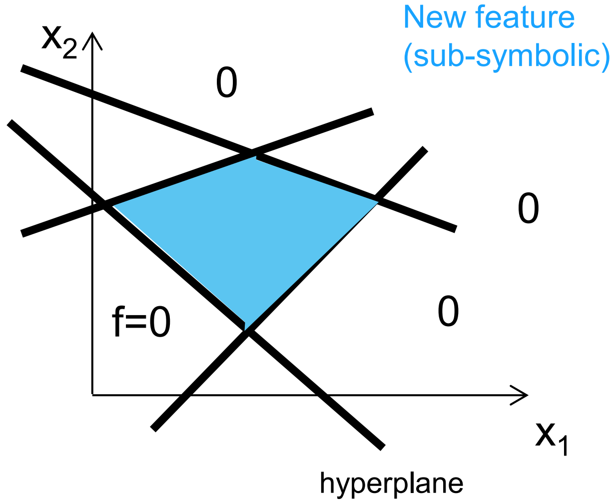
$$\text{model}(\mathbf{x}, \Theta) = w_0 + f_1^{(1)}(\mathbf{x})\, w_1 + \cdots + f_{U_1}^{(1)}(\mathbf{x})\, w_{U_1}$$

$$f_j^{(1)}(\mathbf{x}) = a\left(w_{0,j}^{(1)} + \sum_{n=1}^{N} w_{n,j}^{(1)} x_n\right)$$

weights (N+1) · U_1

$$\mathbf{W}_1 = \begin{bmatrix} w_{0,1}^{(1)} & w_{0,2}^{(1)} & \cdots & w_{0,U_1}^{(1)} \\ w_{1,1}^{(1)} & w_{1,2}^{(1)} & \cdots & w_{1,U_1}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N,1}^{(1)} & w_{N,2}^{(1)} & \cdots & w_{N,U_1}^{(1)} \end{bmatrix}.$$





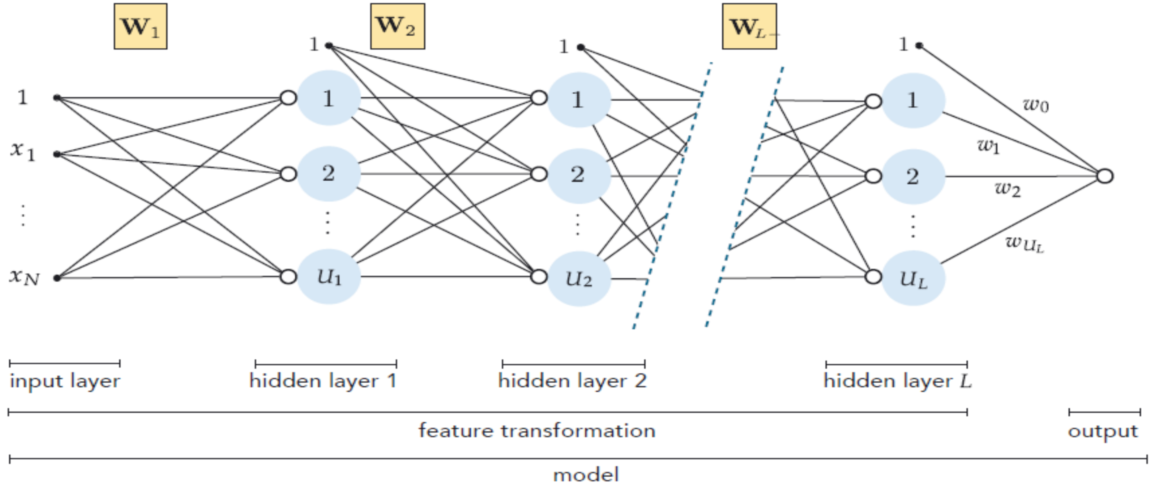
Artificial neuronal networks as universal approximators

Basic intro…

- **Multi-layer neural networks:**

$$\mathrm{model}\,(\mathbf{x},\Theta) = \mathbf{w}_{L+1}^T\,\mathring{a}\left(\mathbf{W}_L^T\,\mathring{a}\left(\mathbf{W}_{L-1}^T\,\mathring{a}\left(\cdots\mathring{a}\left(\mathbf{W}_1^T\mathring{\mathbf{x}}\right)\right)\right)\right)$$

➔ concatenated chain of multi-dim. function
➔ allows to approx. arbitrary functions

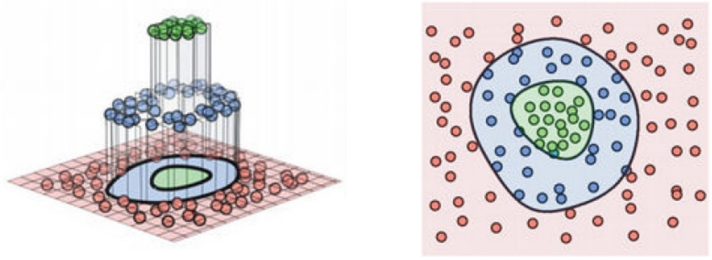


fast — Input to output information flow

slow — Error backpropagation to tune weights using gradient decent

$$\Delta w_i = -\eta\frac{\partial E}{\partial w_i} = -\eta\frac{\partial E}{\partial u}\frac{\partial u}{\partial w_i} = \eta\delta x_i$$

- **Specialties of ANNs**
  - Structure: <u>parallel distributed processing</u>
  - Natural propensity for <u>storing experiential knowledge</u>
- **Break through:**
  - **Efficient training methods** (initialization, regulation techniques, batch-training, dropout, efficient optimizers, avail. libraries....)
  - **GPUs:** massive improvement on parallel processing
  - **Remark:** large ANN still takes a long time to train
- **ANN strength:** Enhancement due manipulation of sub-symbolic elements
- **ANN criticism:** Missing interpretability of sub-symbolic elements



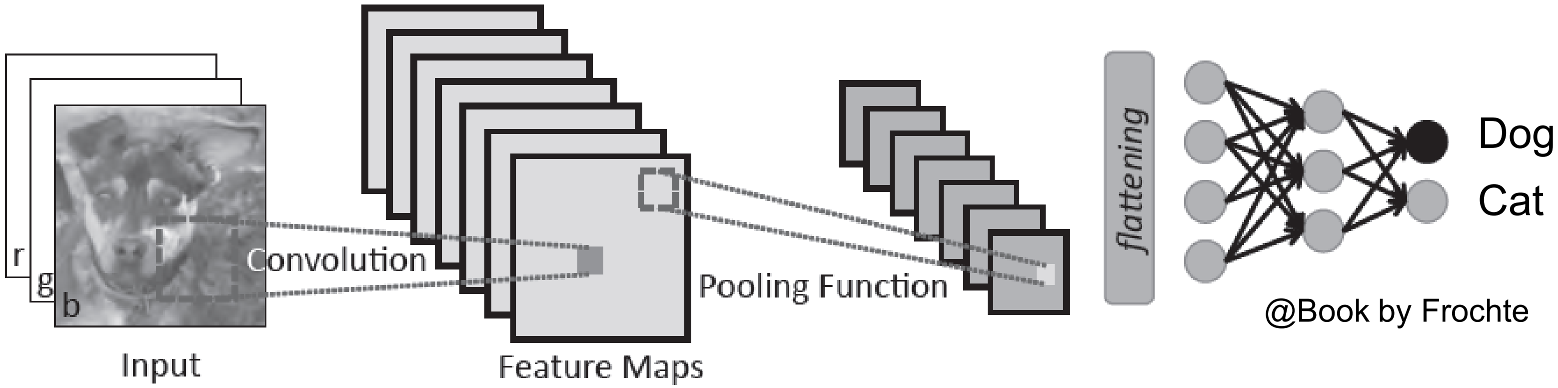
Example full-connected neural network for two-class dataset

Some special types of ANN:

Basic intro…

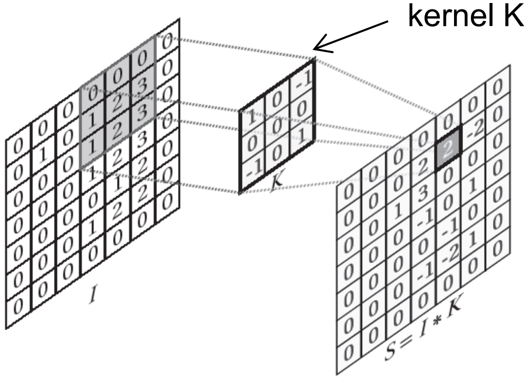
- **Convolutional NN:**
  - Mainly of image processing
  - Feature maps using convolution

typical architectures



@Book by Frochte

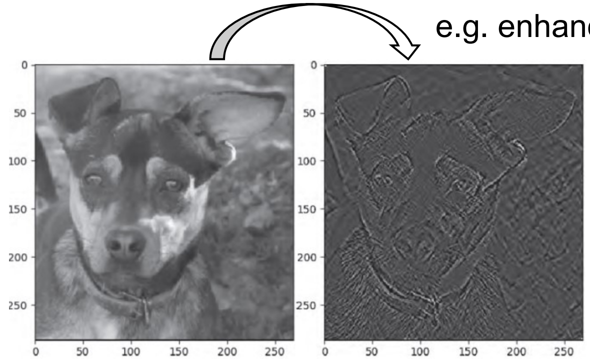
kernel K



$$S[i,j] = (K * I)[j,j] = \sum_m \sum_n I[i+m, j+n] \cdot K[m,n]$$

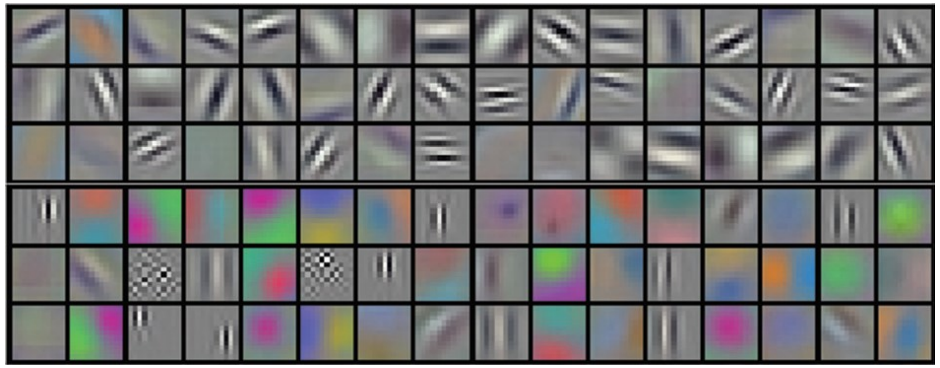
Typically 3x3 / …/11x11

e.g. enhance edges



- **Specialty:**
  - learns kernel function for feature maps
  - Sparse interactions
  - Parameter sharing
  - Equivalent representations

96 kernels 11 x 11 first layer feature map (60M param/650k Neur.)



Ref: Hinten2017 DOI:10.1145/3065386

Some special types of ANN:

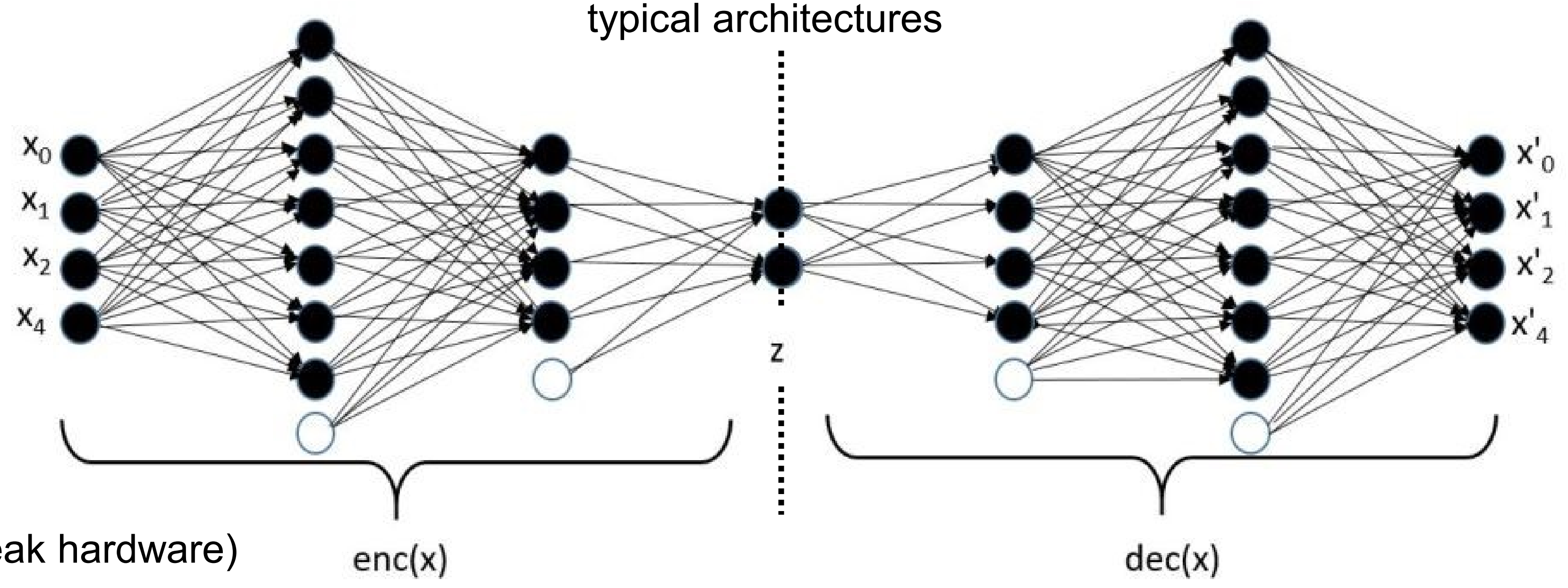
Basic intro…

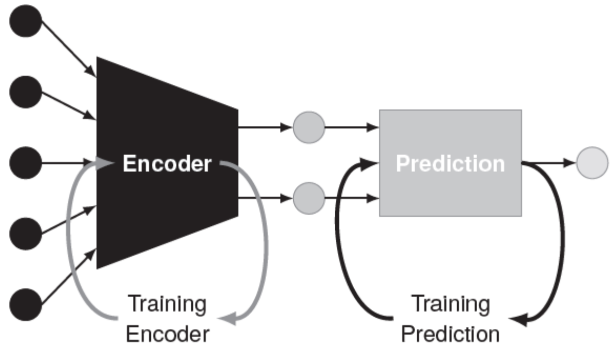
- **Autoencoder-NN:**
  - For feature reduction / extraction.
  - Target value is input value y = x
  - Ask for a function of type

    $x = f_{dec}\ (\ f_{enc}(x)\ )$

- **Specialty:**
  - Use for data reduction
  - Noise removal
  - Non-linear dimension reduction
  - Decoupling of learning cycles (weak hardware)

typical architectures





784 → 24 features

Abbildung 9.6 Output des Autoencoders

@Book by Frochte

Some special types of ANN:

Basic intro…

- **TM-Polynomial-NN:**
  - Activation functions are polynomials
  - E.g. solving non-linear ordinary differential equations

$$\boxed{\frac{d}{dt}\mathbf{X} = \mathbf{F}(t,\mathbf{X}) = \sum_{k=0}^{\infty} P^{1k}(t)\mathbf{X}^{[k]}}$$

F as Taylor expansion

  - Evolution of system from t_0 → t expressed by Matrix Lie maps

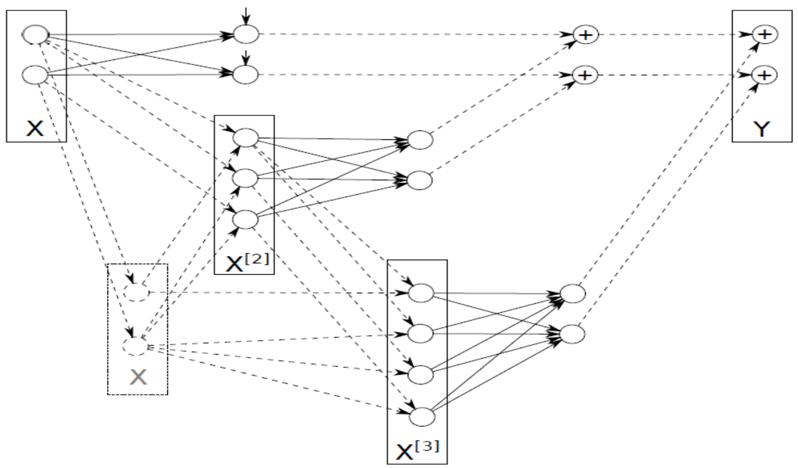
$$\mathbf{X}_{i+1} = \mathcal{M} \circ \mathbf{X}_i = W_0 + W_1 \begin{pmatrix} x_i \\ y_i \end{pmatrix} + W_2 \begin{pmatrix} x_i^2 \\ x_i y_i \\ y_i^2 \end{pmatrix} + W_3 \begin{pmatrix} x_i^3 \\ x_i^2 y_i \\ x_i y_i^2 \\ y_i^3 \end{pmatrix}$$

1) ODE ➔ calc. W ➔ efficient numerical solver for ODL
2) Trajectory ➔ train W ➔ extract ODL

- **Specialty:**
  - Shallow but very deep nets
  - Clear interpretation of weights
  - Trick: Regularization includes symplecticity of W_i

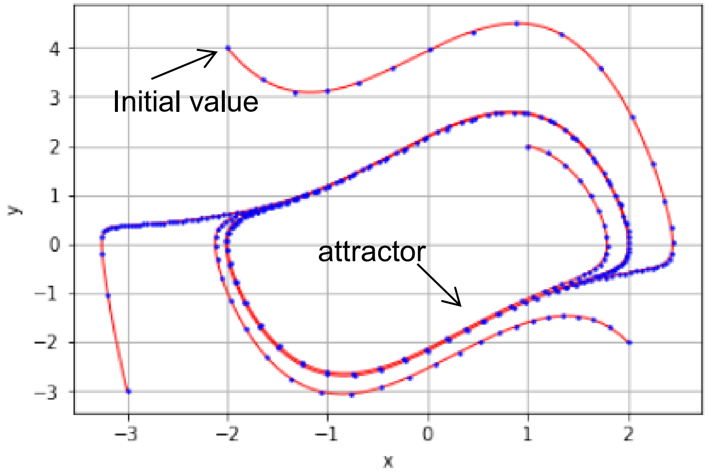
➔ Talk by Andrei Ivanov

typical architectures



Example: Van der Pol Osc.

$$x' = y,$$
$$y' = y - x - x^2 y.$$



Symplecticity:

$$\left(\frac{\partial \mathbf{X}}{\partial \mathbf{X}_0}\right)^T J \frac{\partial \mathbf{X}}{\partial \mathbf{X}_0} - J = 0, \quad \forall\ \mathbf{X}_0, \qquad J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

@Andrei Ivanov

To wrap up

- **Digitalization** and **complexity increase** of accelerator → new tools for controls required
- **Supervisory control** of processes mostly requires to build **models** of your systems:

Deductive	White box	Inductive or Identification	Black box	← Just from data ← Excite system	Grey box	Deductive & Inductive	Model $y = f(x)$, but often $x = f^{-1}(y)$ required

- **AI & ML** (I2)**:** allows to construct models / rules from data, but they are of **statistical nature**
- **Strength of ANN:** learning and manipulation of sub-symbolic elements derived from features
- **Data Mining**: make use of our data base for better understanding of our systems on all levels
- **Predictive controls:** we like to achieve → prediction of accelerator behavior for finite time horizons
- **Goal of seminar:** bring to together the various domain experts across M-division

Some additional information

- **Machine Learning at DESY:** desy-ml@desy.de
  Indico page: https://indico.desy.de/indico/category/641/

- **OWLE:** Virtual seminar series on computational accelerator physics, machine learning and experimental demonstration of AI-ML; https://sites.google.com/view/owle/

- **IPC – mailing list:** ipc-info@desy.de
- **IPC – confluence:** https://confluence.desy.de/display/IPC
- **IPC – seminar:** https://indico.desy.de/indico/category/740/

October 2020

20 Oct Sergey Tomin, "OCELOT" New!

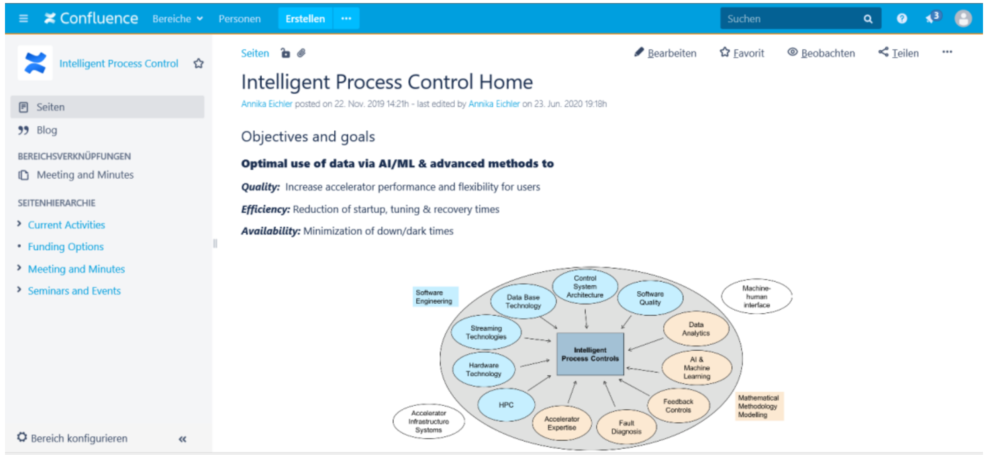
September 2020

22 Sep Gianluca Valentino, "Reinforcement Learning applications in particle accelerators"

08 Sep Maxence Thevenet, "Introduction to High Performance Computing"

August 2020

25 Aug Prof. Gerwald Lichtenberg, "Automated Fault Detection and Diagnosis - an Introduction"



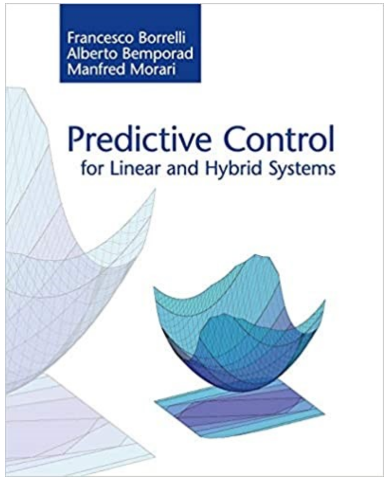
General Contact:

- Annika Eichler
- Raimund Kammering
- Ilya Agapov
- Holger Schlarb

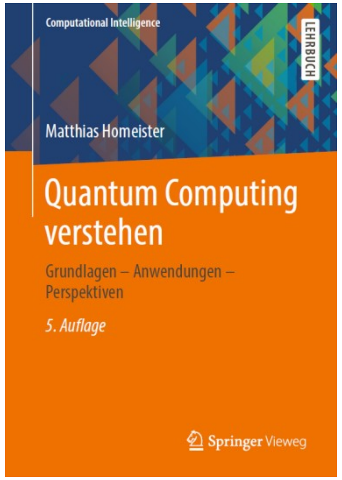
Thanks for attention

Literature:

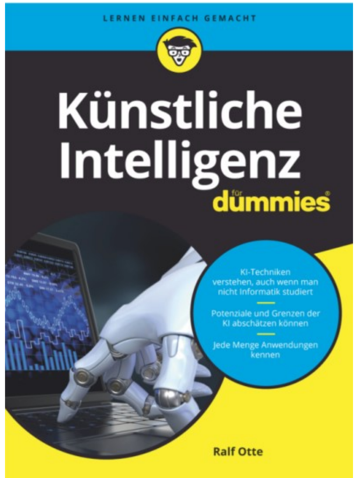
ISBN 978-1-108-48072-7



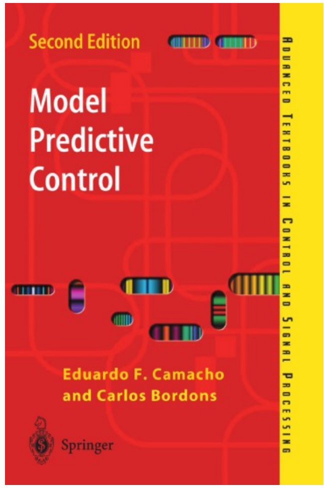
ISBN 978-1-107-01688-0



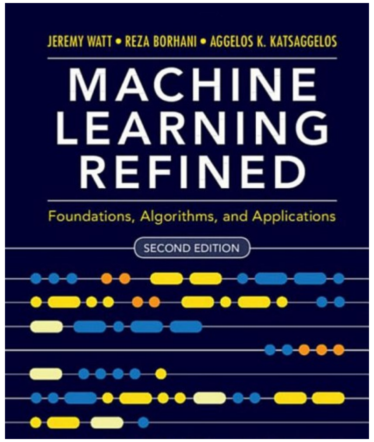
ISBN 978-3-658-22884-2



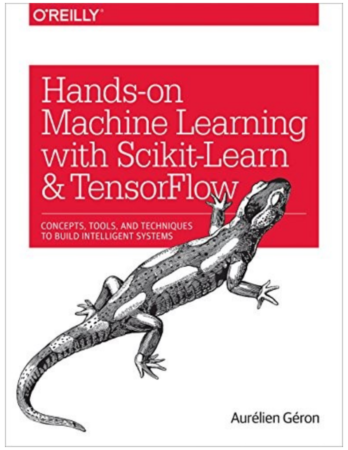
ISBN: 978-3-527-71494-0



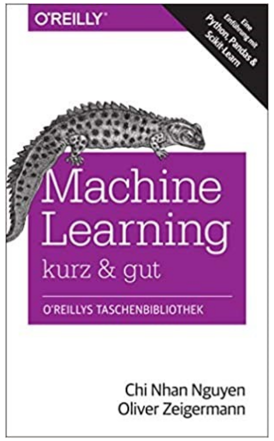
ISBN 978-1-85233-694-3



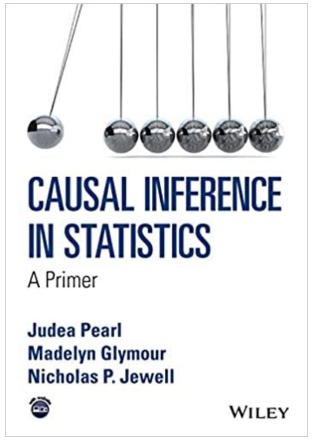
ISBN 978-1-108-48072-7



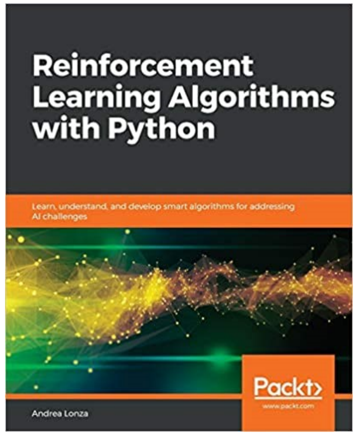
ISBN 978-1-491962299



ISBN 978-3-96009-052-6



ISBN 978-1-119-186847



ISBN 978-1-789-111-6

And many more….