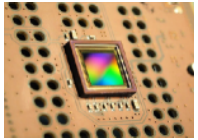
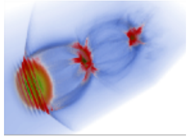
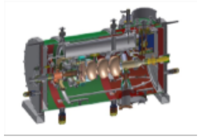
HELMHOLTZ

MATTER AND TECHNOLOGIES







RESEARCH FOR GRAND CHALLENGES

MACHINE LEARNING FOR DATA ANALYSIS, MODELLING AND CONTROLS

F.Gaede, DESY

Matter meets Information: Common Challenges and Perspectives

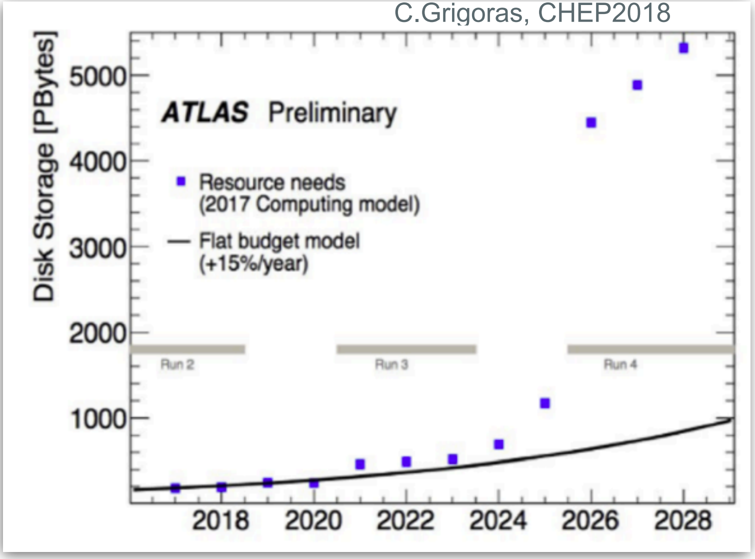
DESY, Jan. 14-15

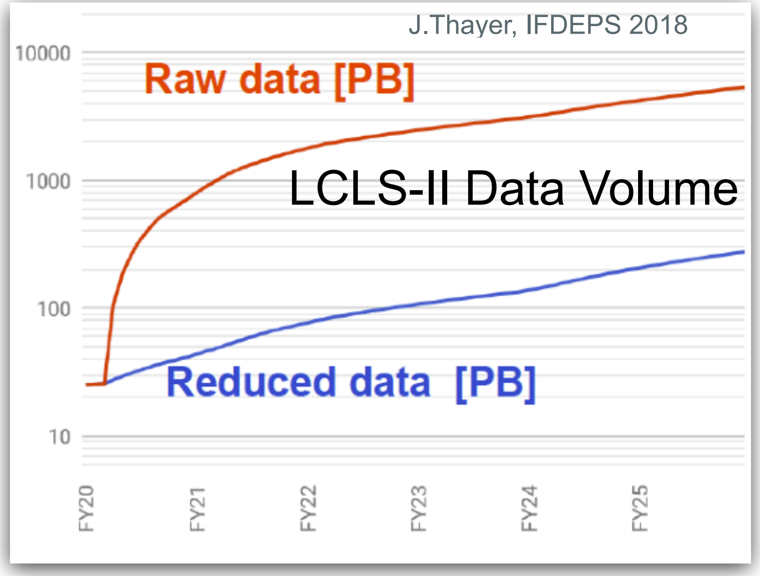
OUTLINE

- Introduction: The Data Challenge:
  - Volume, Rates and Complexity
- Machine Learning: The Answer ?
- some ML Examples from
  - HEP
  - Photon Science
  - Accelerators
- Future Activities
- Summary and Outlook

ML artist's view of the Elbphilharmonie

DATA CHALLENGE: VOLUME





- data volumes from new/upgraded accelerators for HEP and photon science will reach **Exa-Bytes/year** in the coming decade (prime examples: HL-LHC and LCLS-II)
- slightly smaller, yet still challenging: XFEL, FAIR,…

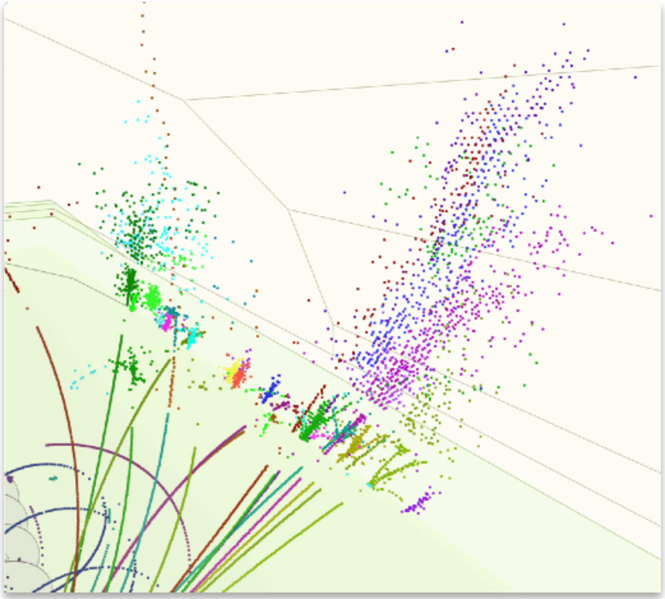
DATA CHALLENGE: RATE

R.Mount,SLAC	LCLS-II 2022	LCLS-II 2026+	ATLAS Today	ATLAS 2026+
Wanted fraction of collisions	0.01 to 1.0	0.01 to 1.0	$< 10^{-6}$	$< 10^{-5}$
Typical experiment duration (same data-taking conditions)	3 days	3 days	3 years	3 years
24x7 availability of offline computing	Essential	Essential	Desirable	Desirable
Required turnround for data-quality checks	Seconds to minutes	Seconds to minutes	Hours to days	Hours to days
Raw digital data rate	200 GB/s	300+ GB/s	160 GB/s	1,000 GB/s
Zero-and-Junk-suppressed rate	10 GB/s	30+ GB/s	1.5 GB/s	20 GB/s
Storage need dominated by	Mainly raw data		Mainly simulated and derived data	
Role of Simulation	Growing in science analysis Growing in experiment design		Vital in physics analysis Vital in experiment design	
Analysis, Simulation and Workflow Software development community	Individuals (in the past) → Organized effort		~100 organized collaborators (mainly research physicists)	

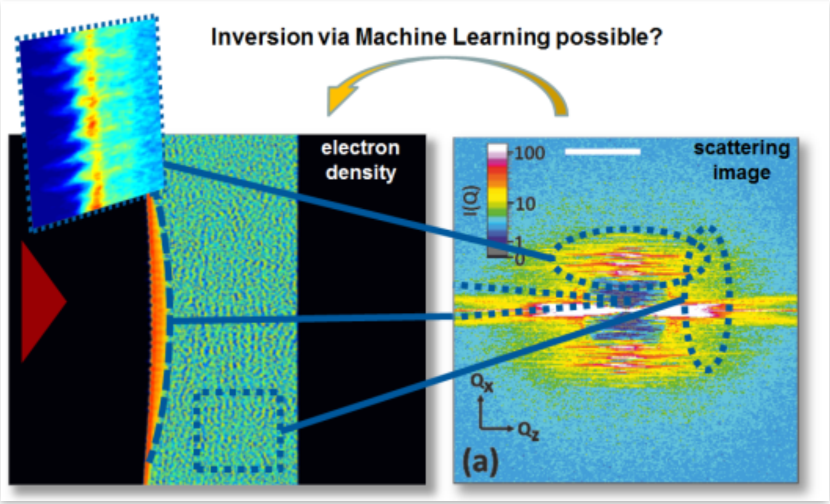
- correspondingly we expect dramatic increases in data rates
- requiring large (real time) **data reduction** and/or **compression** - (10^{-1} - 10^{-6})

MT

DATA CHALLENGE: COMPLEXITY



example: calorimeter showers in **high granular calorimeters** for linear colliders O(10^7) cells



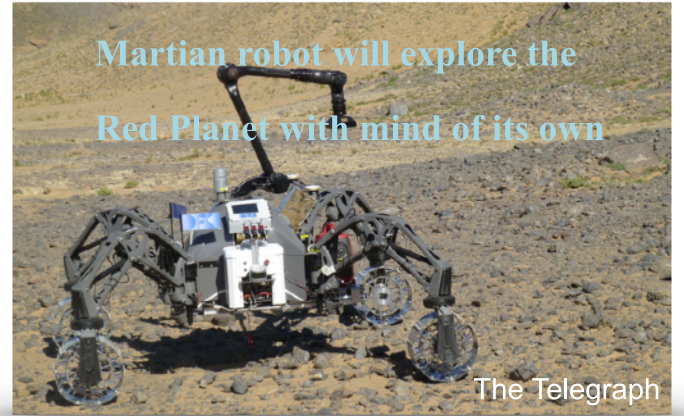
example: **high density** detectors for scattering images at FELs

- ever increasing density and granularity of modern detectors for HEP and PhotonScience
- considerable increase of complexity in the data to be analysed and modelled

MASSIVE (R)EVOLUTION IN SCALE NEEDED

… in storage, triggering, compression, filtering, analysis …

MACHINE LEARNING: THE ANSWER TO EVERYTHING ?

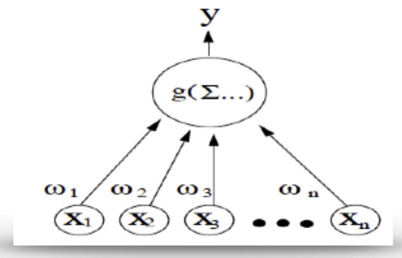


- dramatic and exiting progress in using ML real world applications: autonomous vehicles - GO computers …
- **can we use ML to meet the data challenges in our research field ?**

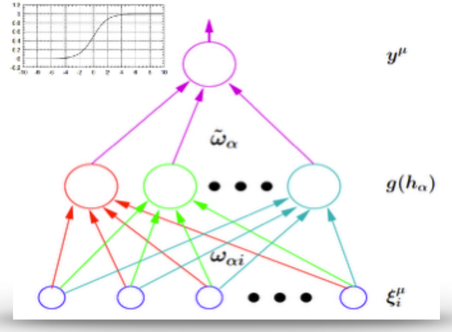
MACHINE LEARNING IN A NUTSHELL

from the Perceptron to CNNs,RNNs,GANs et al

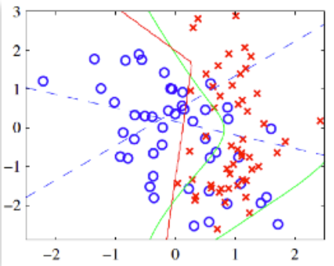
McCullogh-Pitts-Neuron, **1943**



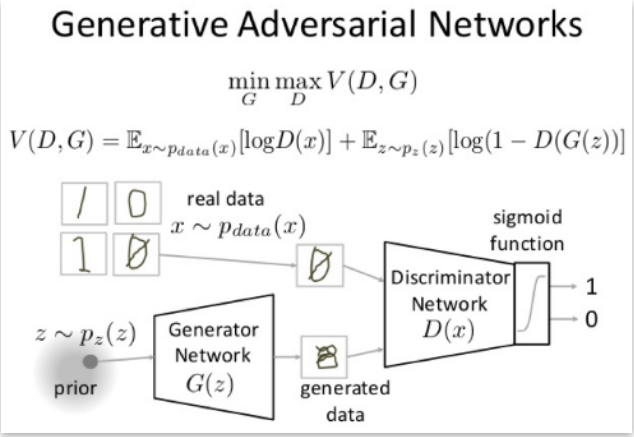
- basic **machine learning** idea:
- create **suitable** network structure
- minimise Error function
    - based on target (training) data
- use for
    - **classification**
    - **regression**
    - **generation** (modelling)

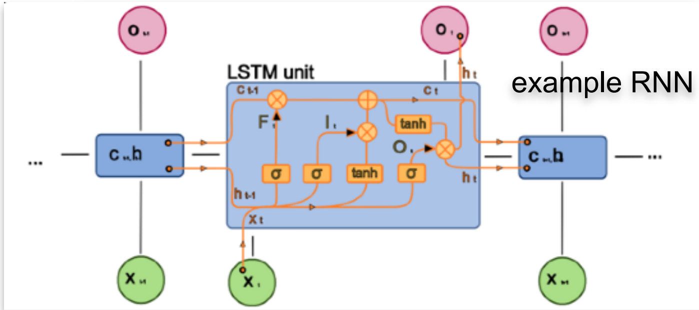


Multi-layer Perceptron
trained w/ **Backpropagation (1980s)**

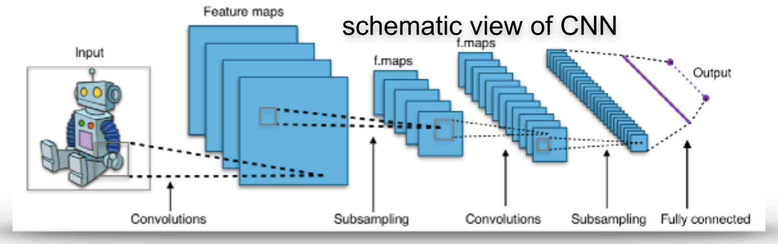


simple **classification** w/ 2x2 MLP



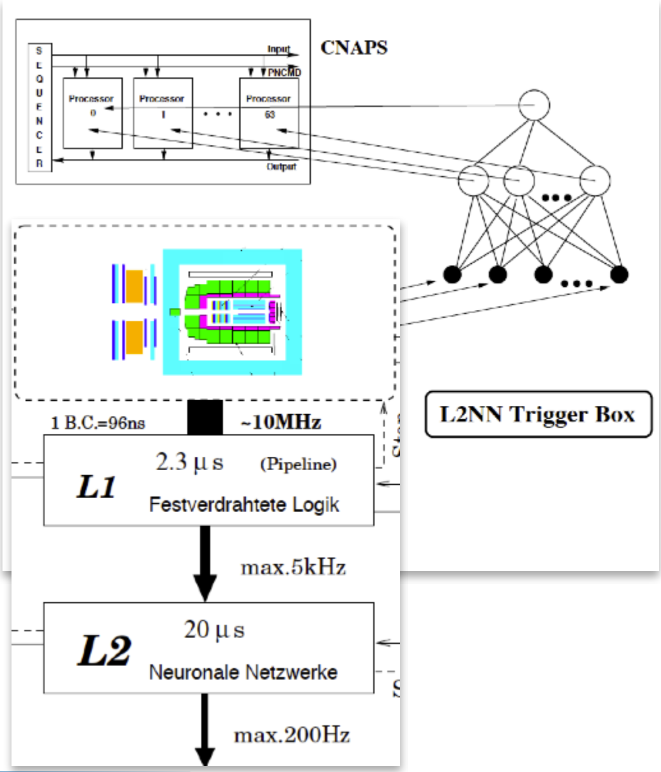


schematic view of CNN



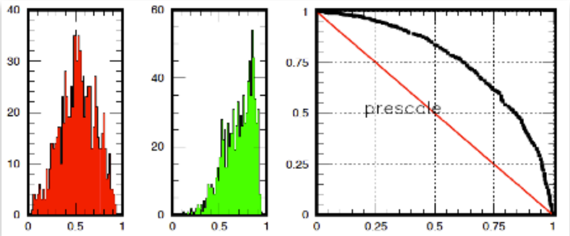
THE H1 LEVEL-2 NEURAL NETWORK TRIGGER

Machine Learning in HEP in the **late 90s**

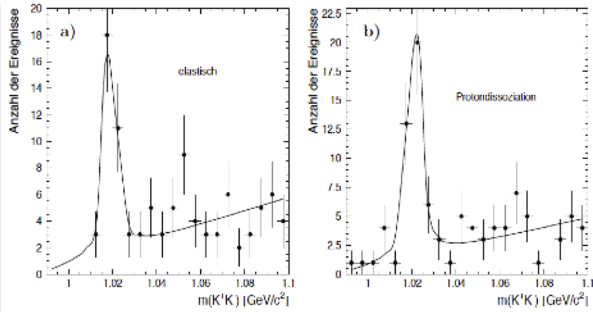


advantages of **ANN** for Triggering:

- exploit **high dimensional correlations**
  - -> better physics performance
- implement the ANN on a dedicated **parallel hardware** (CPU)
  - -> much faster trigger decision



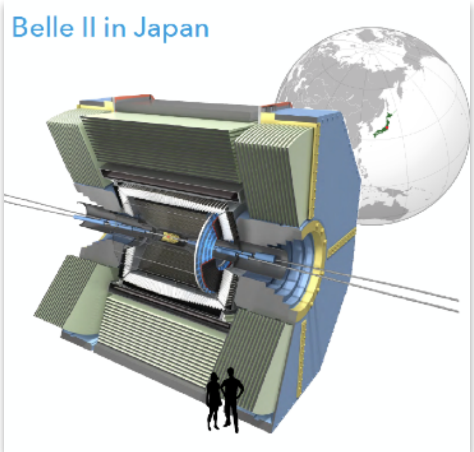
first signal of Φ photo-production in the H1 detector (PhD thesis FG)



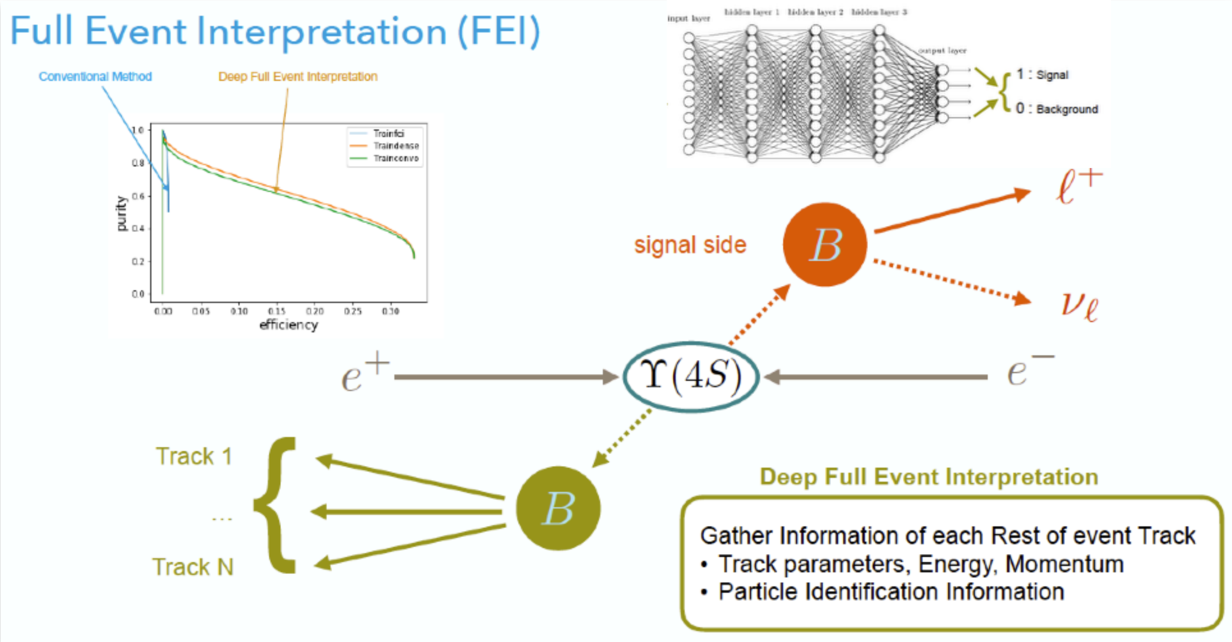
- use of ANNs already very common in HEP event classification at the time:
- HEP-inspire search finds ~500 papers with *Neural Network* written in the 1990s (not all HEP) !

MACHINE LEARNING AT BELLE-II

T.Ferber, S.Wehle DESY



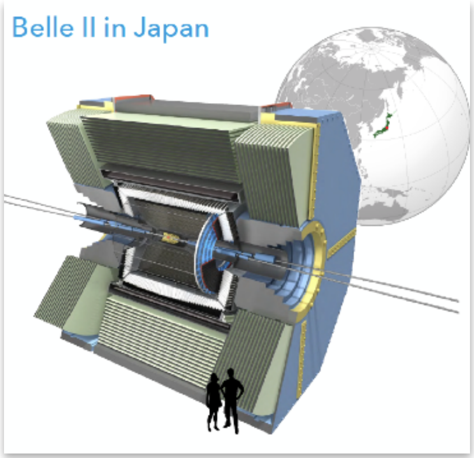
- **intensity frontier** flagship experiment at KEK
- precision measurement and (extremely) **rare B-decays**



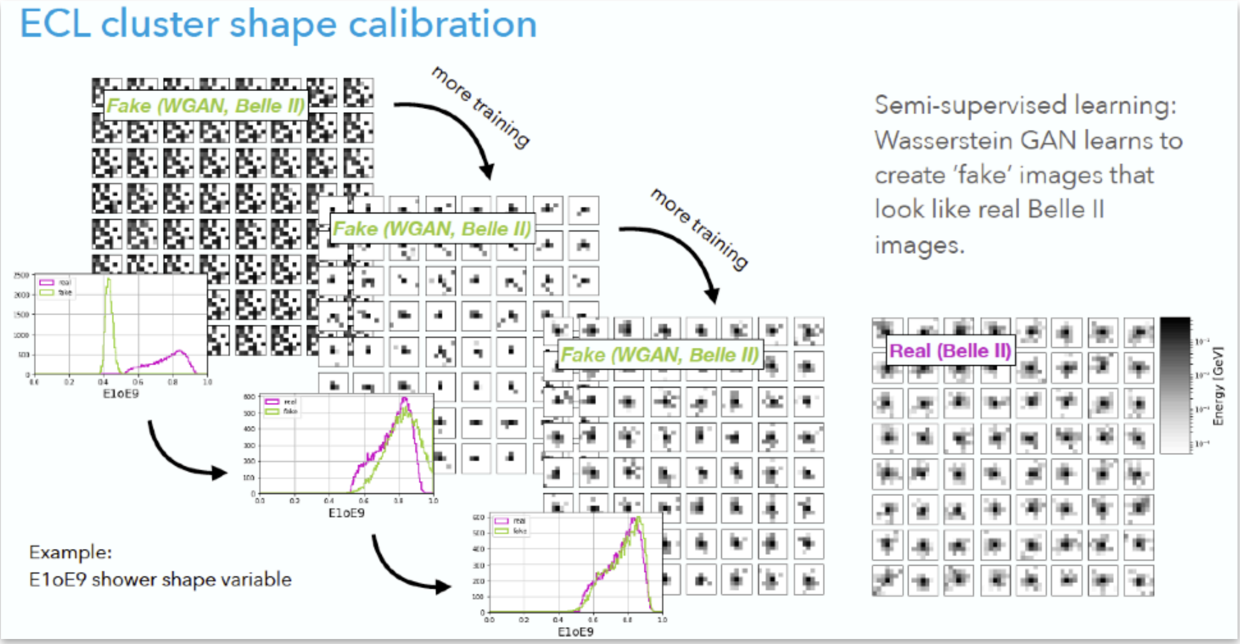
- DDNN and CNN used for this *classification* task
- performance considerably better than conventional method

MACHINE LEARNING AT BELLE-II

T.Ferber, S.Wehle DESY



- **intensity frontier** flagship experiment at KEK
- precision measurement and (extremely) **rare B-decays**

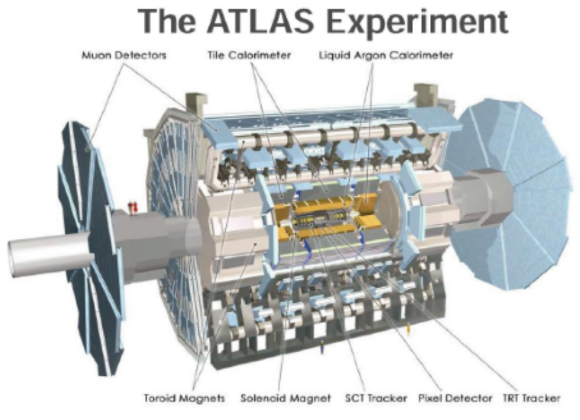


- use of a WGAN for generating *images of calorimeter showers*
- needed to correct imperfect '*conventional*' simulation

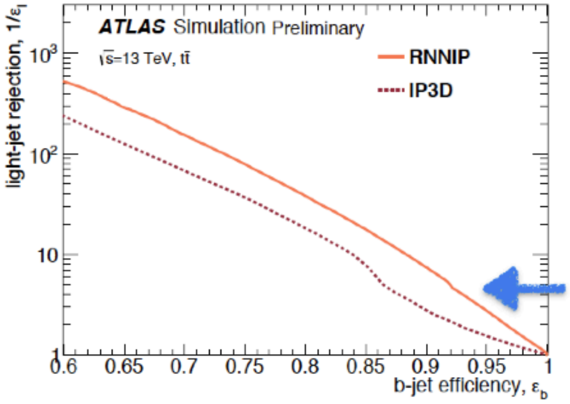
MACHINE LEARNING AT ATLAS

P.Glaysher, DESY

Identification of particle jets



- multi purpose detector at the **LHC** at CERN
- 13 TeV pp-collisions



➡ With deep learning, we achieve similar level of improvement as multi million € upgrade of the detector hardware.

--- IP3D: Probabilistic classifier, a log-likelihood ratio b-jet/light-jet

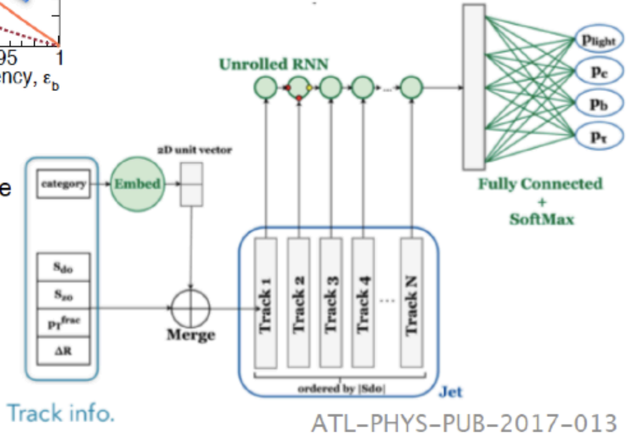
— RNNIP: recurrent NN, exploits sequence based correlations in jet tracks

Improved performance due to deep learning

Success of RNN also due to LSTM units
- establish long range correlations
- mechanism to regulate relative importance of short & long term information

- track info in 50 dimensional vectors
- final network has ~11k trainable free parameters.
- trained on 3.2 M jets.

5

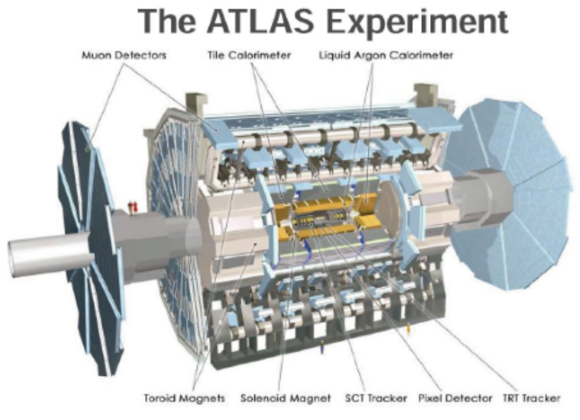


ATL–PHYS–PUB–2017–013

MACHINE LEARNING AT ATLAS

Fast shower simulation

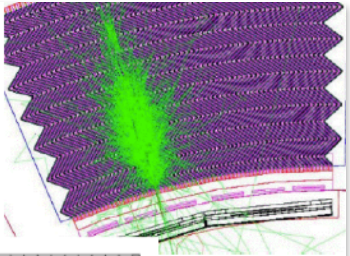
- large fraction of CPU needs due to Monte Carlo Simulation
- speed-up of x100 for generative models (VAE and GAN)

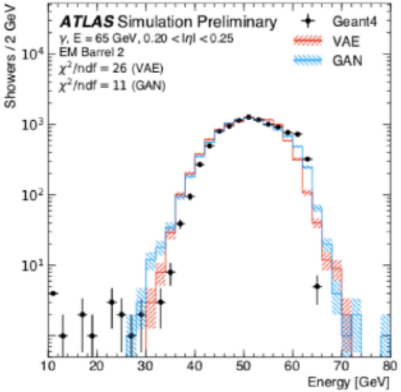


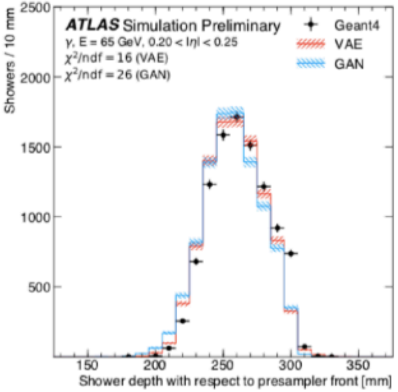
- multi purpose detector at the **LHC** at CERN
- 13 TeV pp-collisions

> Very first results for single photon shower
  > VAE & GAN models compared to Geant4 simulation

> Promising feasibility study, level of accuracy not yet high enough for physics analyses



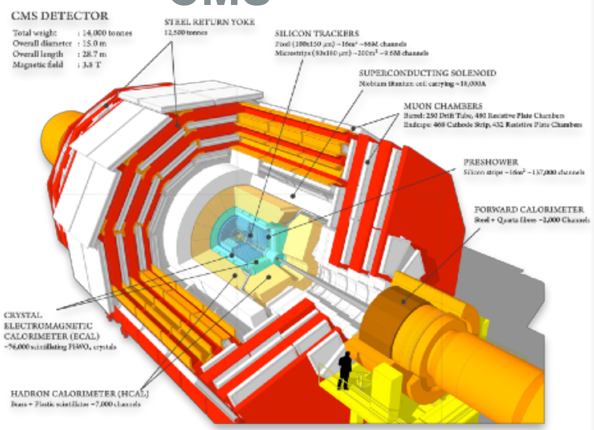




MACHINE LEARNING AT CMS

G.Kasieczka, UHH

CMS

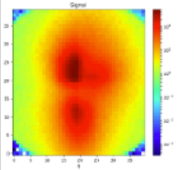


- multi purpose detector at the **LHC** at CERN
- 13 TeV pp-collisions

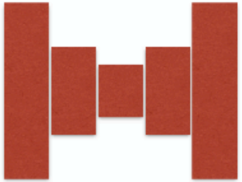
Autoencoder for New Physics

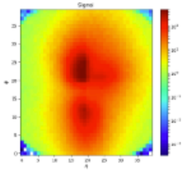
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Input | Autoencoder | Output



X_{ij}

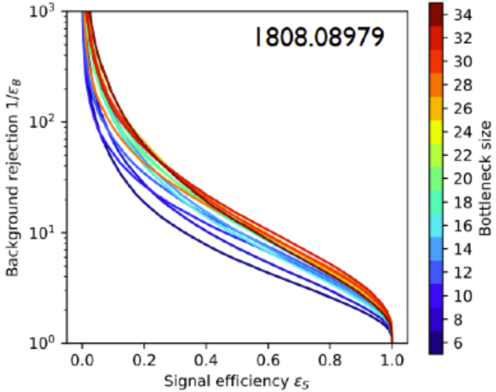




$\widetilde{X}_{ij}$

$$L_{\text{Auto}} = \sum_{\text{Pixels } ij} \left(X_{ij} - \widetilde{X}_{ij} \right)$$

- Can we find new physics without knowing what to look for?
- Train on QCD light quark/gluon jets (can be done in data)
- Top quarks (or new physics) identified as anomaly (high loss)
- Adversarial training against mass-mass-sculpting
- Use for model independent new-physics searches trained on data



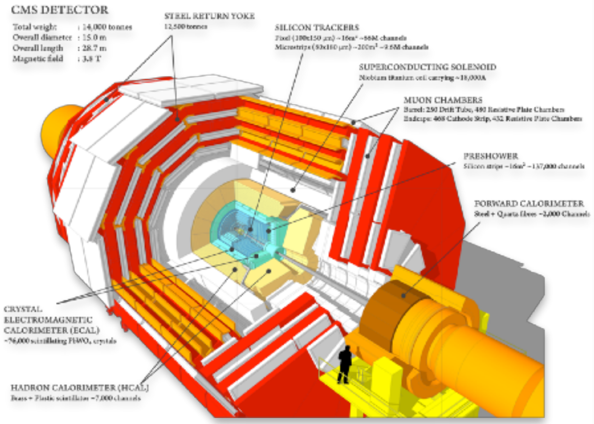
QCD or What?
T Heimel, GK, T Plehn, JM Thompson, 1808.08979

17

MACHINE LEARNING AT CMS

G.Kasieczka, UHH

CMS

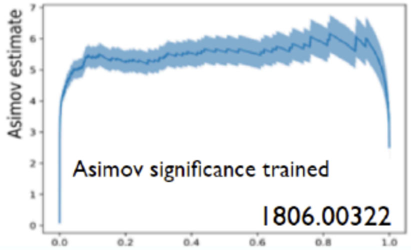


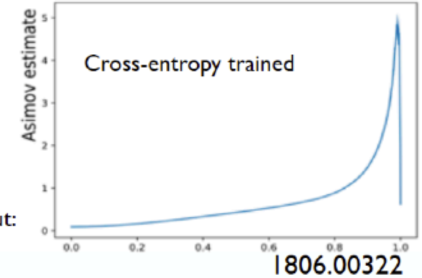
- multi purpose detector at the **LHC** at CERN
- 13 TeV pp-collisions

Asimov significance

$$Z_A = \left[2\left((s+b)\ln\left[\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right] - \frac{b^2}{\sigma_b^2}\ln\left[1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right]\right)\right]^{1/2}$$

- Is there a better optimisation target than cross-entropy for classifers?
  - Try to optimize directly for the Asimov significance, i.e. use it as a loss function
- Caveat: To define the number of signal and background events we need to cut on the discriminator output
  - Makes it non-differentiable





Asimov loss training:
best Z_A = **6.2 ± 0.6**
Cross-entropy training +purity cut:
best Z_A = **4.8 ± 0.3**
(Systematic uncertainty 50%)

$$s = W_s \sum_i^{N_{batch}} y_i^{pred} \times y_i^{true}$$

$$b = W_b \sum_i^{N_{batch}} y_i^{pred} \times (1 - y_i^{true})$$

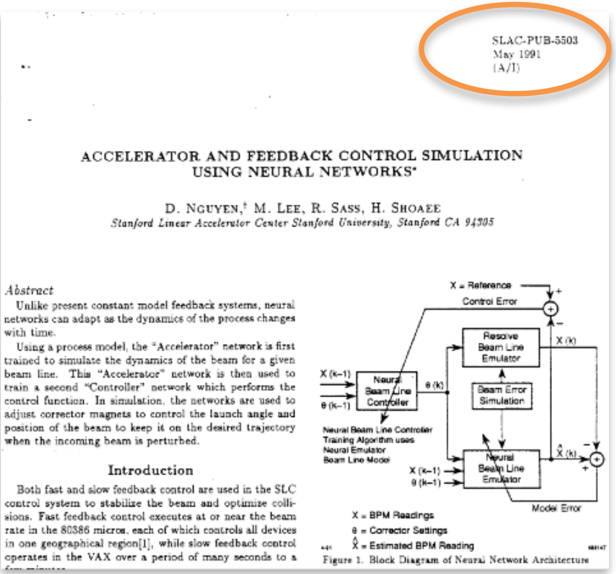
$1/Z_A(s,b)$ becomes a smooth function of y_i^{pred}

- A single sigmoid output neuron
- Replace the discrete number of signal and background events by a smooth function of the predicted label

Direct optimisation of the discovery significance when training neural networks to search for new physics in particle colliders,
A Elwood, D Krücker, 1806.00322

18

MACHINE LEARNING FOR ACCELERATORS



SLAC-PUB-5503
May 1991
(A/I)

ACCELERATOR AND FEEDBACK CONTROL SIMULATION USING NEURAL NETWORKS*

D. Nguyen,† M. Lee, R. Sass, H. Shoaee
Stanford Linear Accelerator Center Stanford University, Stanford CA 94305

Abstract

Unlike present constant model feedback systems, neural networks can adapt as the dynamics of the process changes with time.

Using a process model, the "Accelerator" network is first trained to simulate the dynamics of the beam for a given beam line. This "Accelerator" network is then used to train a second "Controller" network which performs the control function. In simulation, the networks are used to adjust corrector magnets to control the launch angle and position of the beam to keep it on the desired trajectory when the incoming beam is perturbed.

Introduction

Both fast and slow feedback control are used in the SLC control system to stabilize the beam and optimize collisions. Fast feedback control executes at or near the beam rate in the 80386 micros, each of which controls all devices in one geographical region[1], while slow feedback control operates in the VAX over a period of many seconds to a

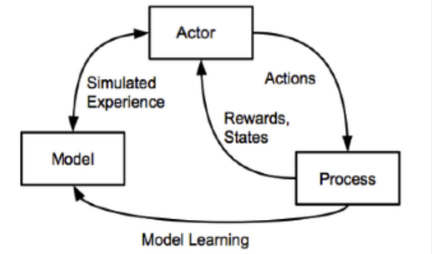
Figure 1. Block Diagram of Neural Network Architecture

- applying Machine Learning (ANN) to accelerators also goes back 30 years !
- more recent activity with tremendous progress in ML activity and increasingly complex accelerators
- mostly applied to: Simulation/Modelling and Control (Fast Feed-back)
- material taken from recent workshop at SLAC:

MACHINE LEARNING FOR ACCELERATORS

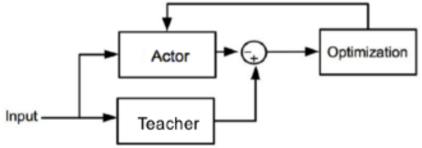
Application of Reinforcement Learning

A.Edelen, CSU



Can train on models first to get a good initial solution before deployment

RL also used in alphaGo



Can use supervised learning to first approximate the behavior of a different control policy

FAST

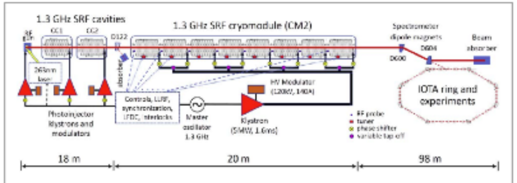
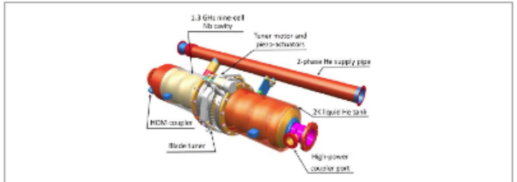


Figure 2. Schematic layout of the FAST/IOTA facility (not to scale).



- **F**ermilab **A**ccelerator **S**cience and **T**echnology Facility
- accelerator R&D, e.g for the ILC

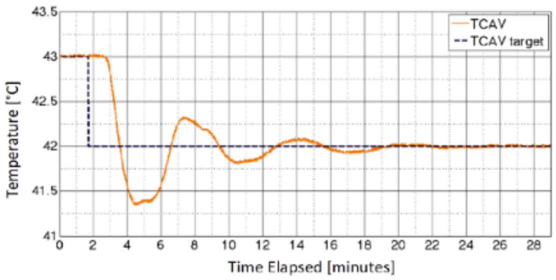
Temperature Control for the RF Photoinjector at FAST

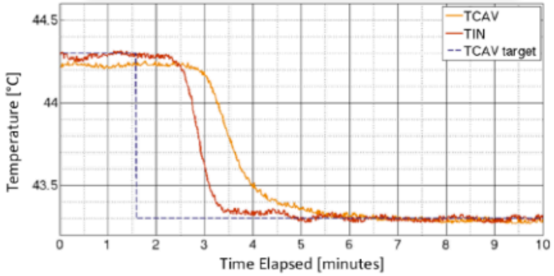
Resonant frequency controlled via temperature

PID control is undesirable in this case:

- Long transport delays and thermal responses
- Recirculation leads to secondary impact of disturbances
- Two controllable variables: heater power + valve aperture

Applied model predictive control (MPC) with a neural network model trained on measured data: ~ 5x faster settling time + no large overshoot





MACHINE LEARNING FOR ACCELERATORS

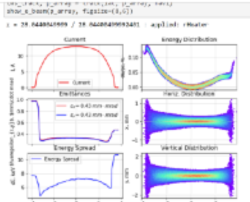
I.Agapov, DESY

Modelling and Control of FELs I (Flash, EU-XFEL, LCLS)

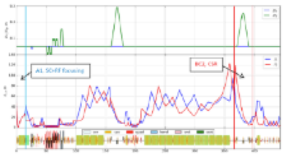
OCELOT

- Software tools used in optics design: *madx, ptc, opa, elegant, at, lego, ocelot*
- OCELOT started in 2011 as a C++ beam dynamics framework with a wakefield focus *(Agapov and Zagorodnov KWT Seminar 2011)*, completely rewritten in python for XFEL.EU SR/FEL simulation purposes *(Agapov et al. Proc IPAC13 Shanghai)*, implemented comprehensive beam dynamics capabilities (*I. Agapov et al., NIM A 768 2014*)
- Mostly used for linac and FEL simulations
- Physics: beam transport up to 2nd order or drift-kick-drift, CSR, wakefields, space charge, Ming Xie FEL estimator, genesis interface, x-ray optics, SR radiation
- Includes advanced on-line control and optimization module
- Open source "community project" https://github.com/ocelot-collab current lead developer S. Tomin (since 2016). Web-based tutorials to get started

From XFEL laser heater tutorial

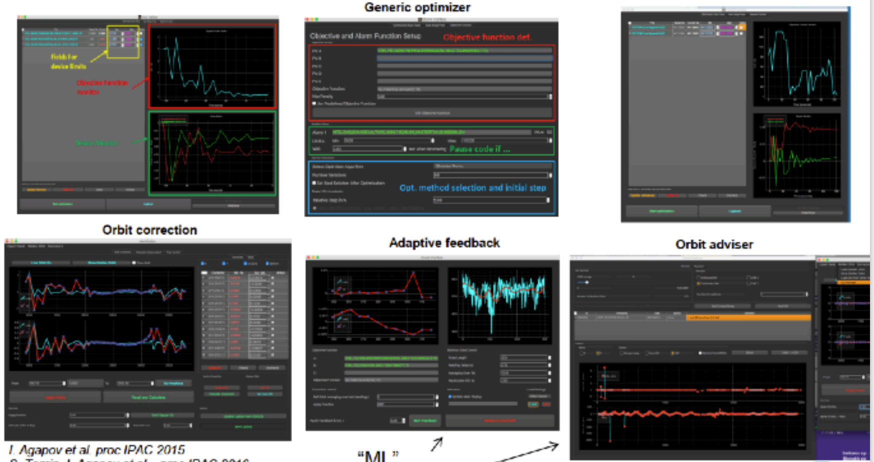


European XFEL linac



OCELOT optimization

- Started as FLASH accelerator R&D in 2015. Now a suite of tools widely used at XFEL.EU, LCLS and FLASH. Partially ported to PETRA (generic optimizer: configured for linac transmission efficiency). Collaboration with SLAC/LCLS.



I. Agapov et al. proc IPAC 2015
S. Tomin, I. Agapov et al., proc IPAC 2016
I. Agapov et al., proc. ICALEPCS 2017
I. Agapov et al., DESY-17-054
I. Agapov et al. proc IPAC 2018

"ML"

Orbit correction based on SASE history

large toolkit with some ML components - potentially to be extended

MACHINE LEARNING FOR ACCELERATORS

Application of OCELOT to BESSY-II

Optimization - OCELOT (DESY) @ HZB - BESSYII

Accelerator Controls optimization

- Developed @ DESY (Python)
- OCELOT optimizer test: optimize 4 skew quads
- Randomized 4 skew quads
- Run OCELOT optimizer
- Convergence achieved
- Comparable lifetime, loss-rate and beam size
- Next test planned for February

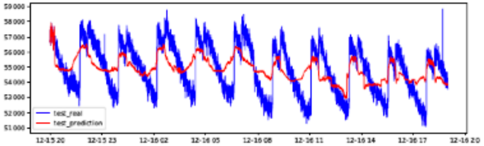
Agapov, Tomin, Birke, Ries, Li, Mertens

1

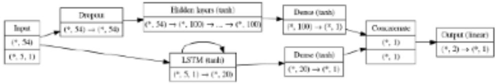
ML Prediction Models @ HZB - BESSYII

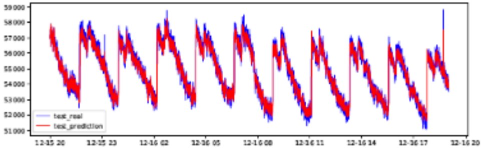
- First application: energy measurement (SCALER9ZR)
- (1) "Blind" models (RF, SVR, DNN...): use only descriptive features (readbacks) - some of them are self-explaining
- (2) Temporal models (DNN+RNN): incorporate time series
- Next steps: lifetime, purity, optimization...

(1)



(2)





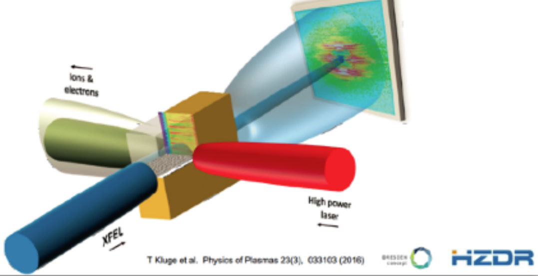
2

MACHINE LEARNING FOR PHOTON SCIENCE

M.Bussmann, HZDR

X-Ray FELs for probing laser-driven high energy density plasmas

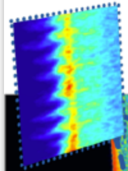
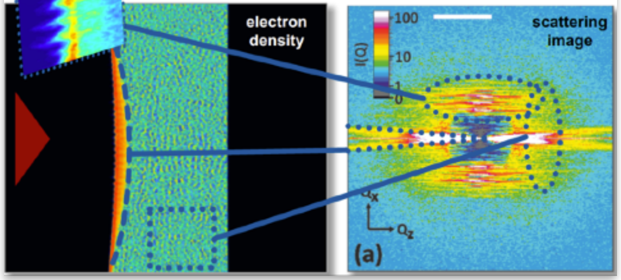
$$\Phi = \underbrace{\Phi_0}_{\text{Flux}} \cdot \underbrace{\Delta\Omega}_{\text{Solid angle}} \cdot \underbrace{T}_{\text{Trans-mission}} \cdot \underbrace{\epsilon}_{\text{Detection efficiency}} \cdot \left| \underbrace{r_0}_{\text{Thomson scattering cross section}} \cdot \underbrace{\int n_e(\vec{r}) \cdot e^{i\vec{q}\vec{r}} d\vec{r}}_{\text{Fourier transform of } e^- \text{ density}} \right|^2$$



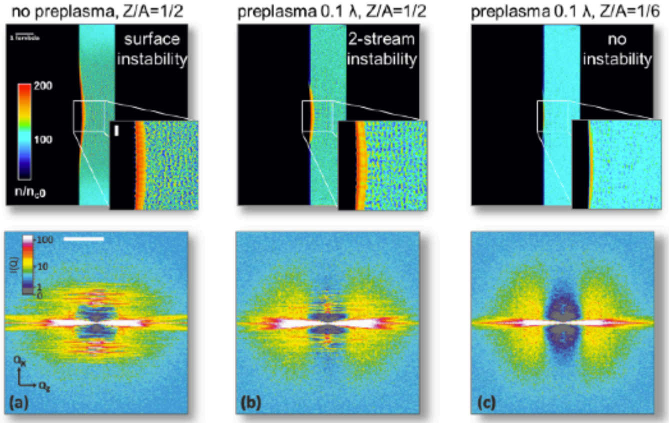
T Kluge et al. Physics of Plasmas 23(3), 033103 (2016)

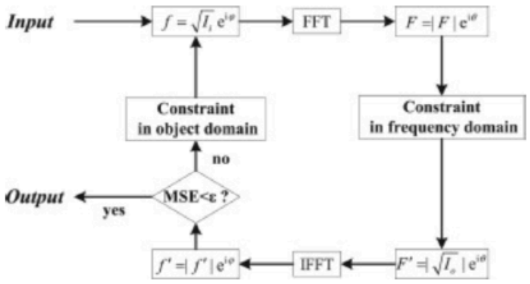
Small angle X-ray scattering (SAXS) on laser-driven HED plasmas

- Scattering image reveals density features
- Processes depend on density, temperature, fields, atomic physics, …
- Modells can parametrize scattering image in terms of physical parameters

Idea: Replace iterative phase retrieval by one-step NN inference





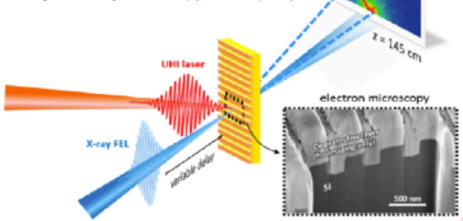
- Parameterize scattering image generation via synthetic (surrogate) model parameters from which the physical properties can be inferred
- Instead of using iterative phase retrieval and obtain density, directly infer model parameters from scattering image
- This removes the often unnecessary step in obtaing a real space density image and is better to account for in high-throughput scenarios
- Only works reliably for well defined, specific setups, as in general there‘s no bijection between the generating parameters and the scattering image

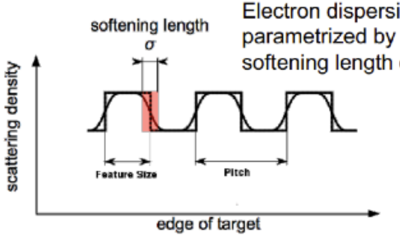
MACHINE LEARNING FOR PHOTON SCIENCE

M.Bussmann, HZDR

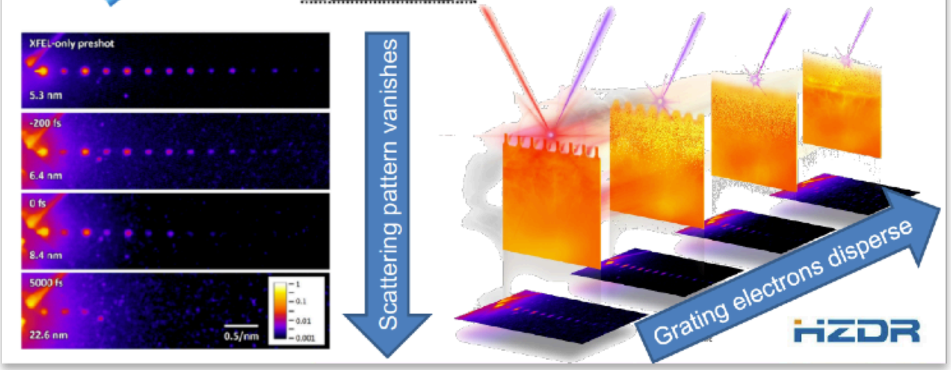
Example: A simple, 1D surface grating „melted“ by a laser

T Kluge et al. Phys. Rev. X 8(3), 031068 (2018)



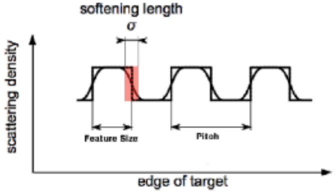


Electron dispersion parametrized by softening length σ



Example: Training for parameter fitting with synthetic data

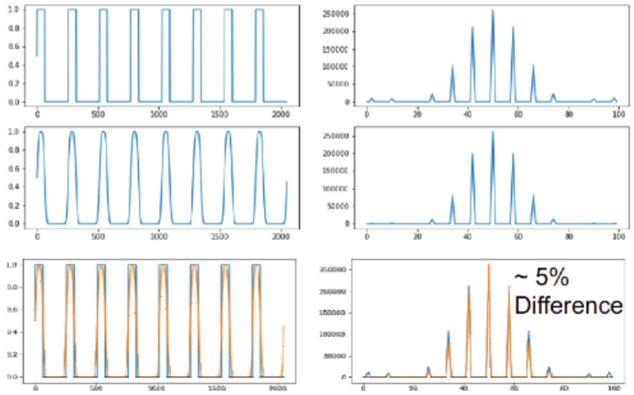
- Highly nonlinear dependence of scattering picture on softening lenght σ
- All other model parameters extracted with very high accuracy (> 99,9%)
- Nonlinearity results in a ~10% error in Sigma



σ =0.0001

σ =10

Comparison

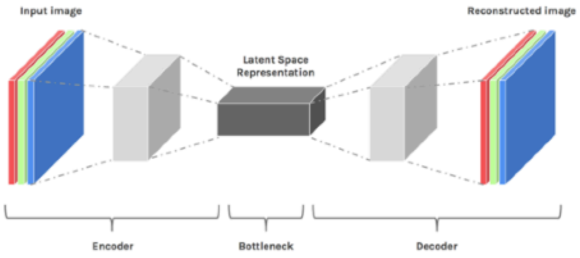


J.Kelling, HZDR

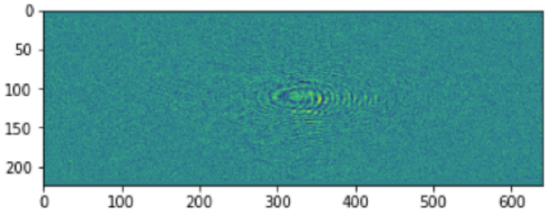
Compression of MRI Raw Data

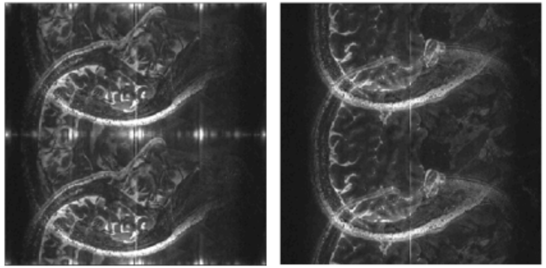
- model learns to encode patterns efficiently
- avoid uncontrolled information loss by retaining encoding error

- Autoencoders for domain-specific compression of raw data



- Use case: MRI data from Rheinland study*
  - Enable retention of raw data for future improved reconstruction





*long term health study investigating neurodegenerative diseases

AMALEA

Helmholtz Innovation Pool Project (2019/20)

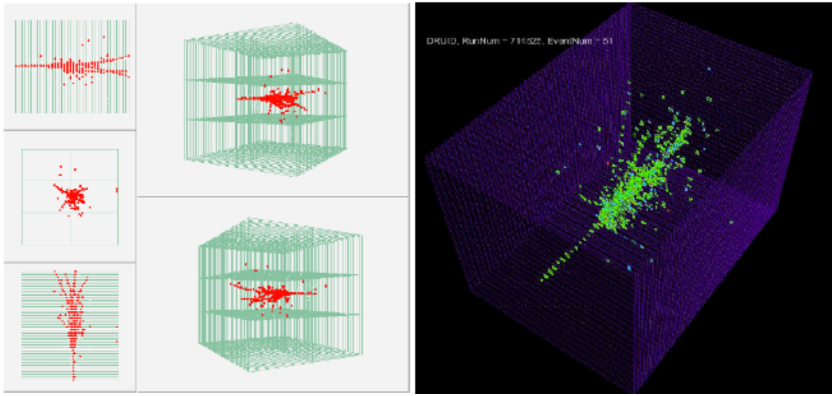
- application of Machine Learning to **challenging projects** in HEP, Photon-Science and Accelerators
- creation of a **sustainable infrastructure** (Distributed ML-Lab) with dedicated hardware (multi-core, GPU, FPGA) und software tools
- development of
  - **fast simulation and reconstruction for 3D images** of novel detectors, e.g. imaging calorimeters and *cameras* in photon science
  - **ultra-fast feedback algorithm for data reduction, compression and classification** of data from current and future light sources
  - **fast diagnosis and control systems** for the optimisation of optics, emittance and beam dynamics of accelerators, such as Petra-IV or BESSY III

DESY. HZB Helmholtz Zentrum Berlin HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF KIT Karlsruher Institut für Technologie

AMALEA EXAMPLE

fast - but detailed - simulation of hadronic showers

- already encouraging attempts to use W-GANs for fast modelling of calorimeter showers (ATLAS, HGCal)
  - so far not quite fit for physics use
- new imaging calorimeters for **particle flow*** will have much **increased granularity** that reveal the details of the hadronic shower substructure
- can we generate hadronic showers that are detailed enough and physically correct to replace the CPU intensive Geant4 simulations ?
  - longitudinal & transverse shapes
  - energy-profile
  - em/had - ratio



hadronic shower in SDHcal prototype (Calice)

*particle flow:
- reconstruct every particle shower individually
- main source of error:
  - confusion due to overlaps
  - needs to be modelled correctly

SUMMARY AND OUTLOOK

- **Machine Learning** has made **tremendous progress** in recent years
  - new algorithms, network architectures and training methods
  - combined with dramatic **progress in hardware** (multi-core, GPU, FPGA)
- ML could be a major ingredient to **match the data challenges** in our field
- HEP, Photon Science and Accelerator physics are (and have) actively used ML techniques
  - mostly as applied users - but also contributing to improving the methods, if needed
  - only tiny selection of activities shown in this talk
- **many exiting projects and applications of ML in matter ahead**:
  - need **dedicated domain knowledge** from the scientists in Matter
  - ideally complemented with cutting edge **expertise from information** scientists