Optimization of Large Setups Using machine learning

Denis Derkach

LAMBDA, HSE University, Moscow, Russia



02 November 2022

Why Optimization?

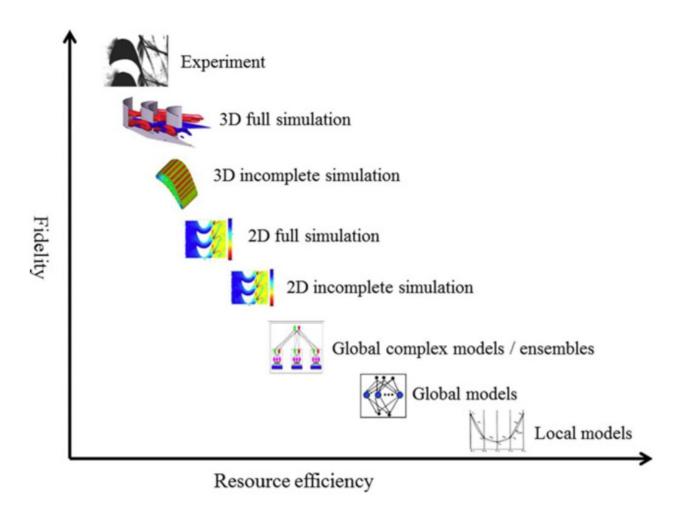
- Optimization is a question of survival.
- Industrial equipment can be optimized for:
 - Lower costs.
 - Less emission.
 - Faster production.
 - Etc.
- Need digital twin of the setup for better tuning.

Surrogate modeling in a wild



Surrogate Models

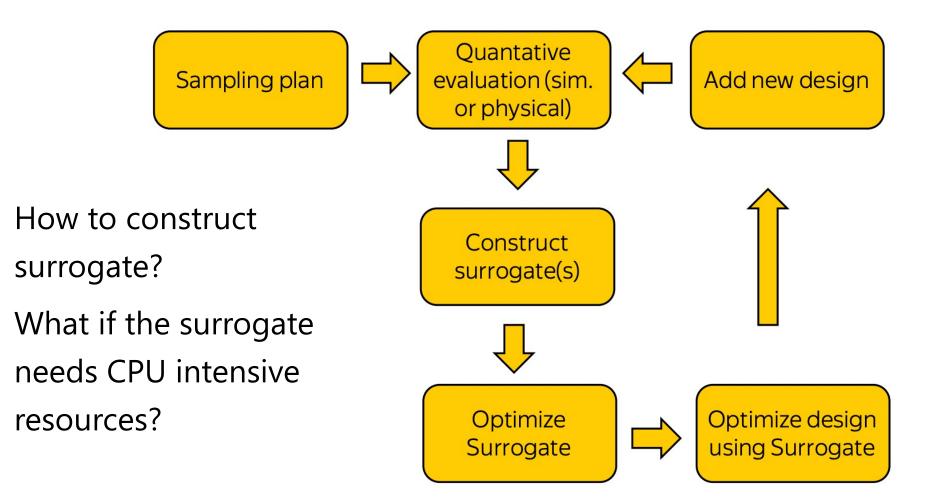
- Model
 - Computational (e.g.
 Computational Fluid
 Dynamics)
 - Data–driven
- Goals
 - Design optimization
 - Configuration space approximation
- Most critical task:
 - A strategy for using the surrogates is known as model management.



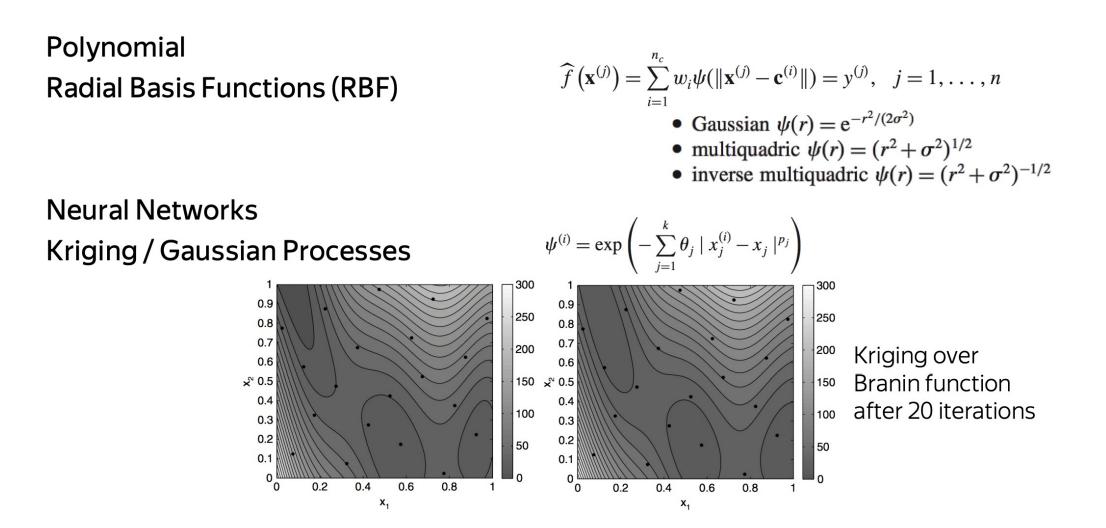
Why Bother

- Examples
 - Drug design
 - Aero engineering (aircraft wing profile, turbine)
 - Steel production quality control
 - Weather forecasting
- Regular physically-based computational models are fine but
 - Expensive computation
 - Large parameter space
 - Numerical noise / uncertainty

Surrogate Modelling. Generic Optimization



Surrogate Modelling. Generic Optimization



The Collider



Collider

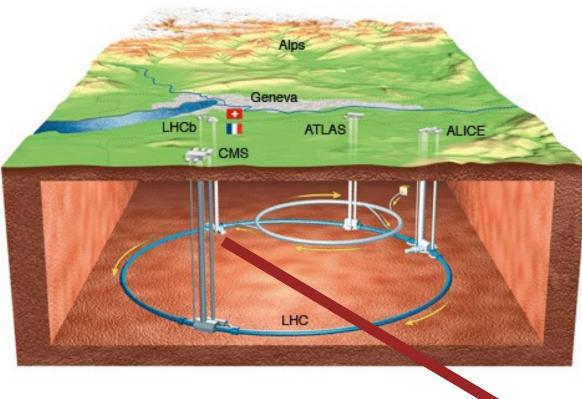


Large Hadron Collider



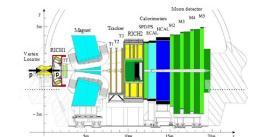
- Millions of events per second.
- Underlying models govern collision outcome.
- Hypotheses testing for theories that describe models.
- Higgs Boson.

Large Hadron Collider



"Cameras" under the surface:

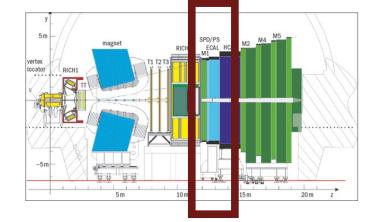
- many layers of sensors: ~200 sq.m. matrices
- resolution: ~100M pixels
- photo speed: 40 000 000 events per second
- record: 200-1000 photos per second
- work for many years



Detector Proper



Size: 7.8x6.3x0.5 m



Module size 12x12 cm²

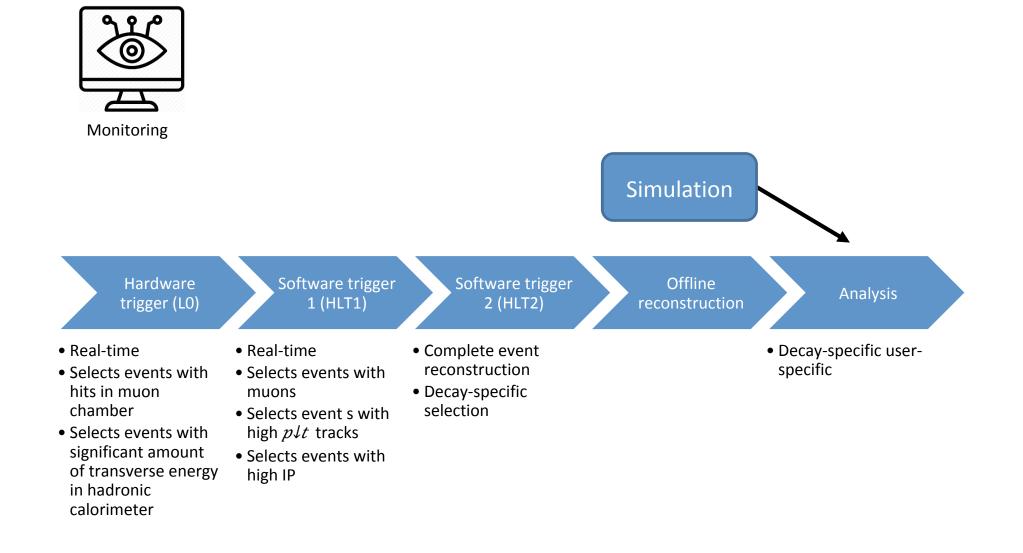
176 inner modules: 9 cells with size 4x4 cm²
448 middle modules: 4 cells with size 6x6 cm²
2688 outer modules: 1 cell with size 12x12 cm²



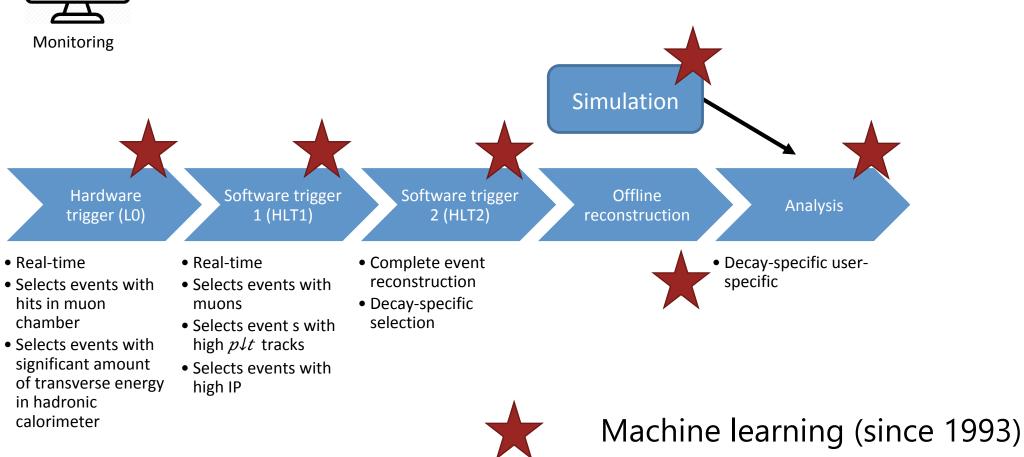
D. Derkach, NRU HSE

13

LHC: Machine Learning



LHC: Machine Learning





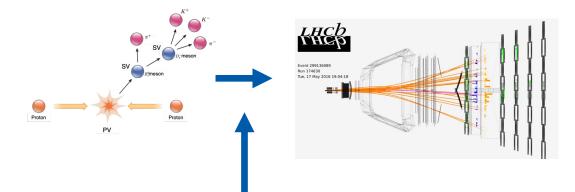


Theory Construction

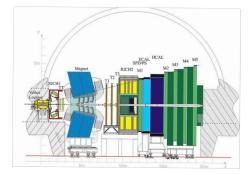


 $+rac{1}{2}|\partial_{\mu}\eta+iM_{Z}Z_{\mu}+rac{ig}{2c_{w}}\eta Z_{\mu}|^{2}-\sum_{f}rac{g}{2}rac{m_{f}}{M_{W}}ar{\Psi}_{f}\Psi_{f}\eta$

Data collection

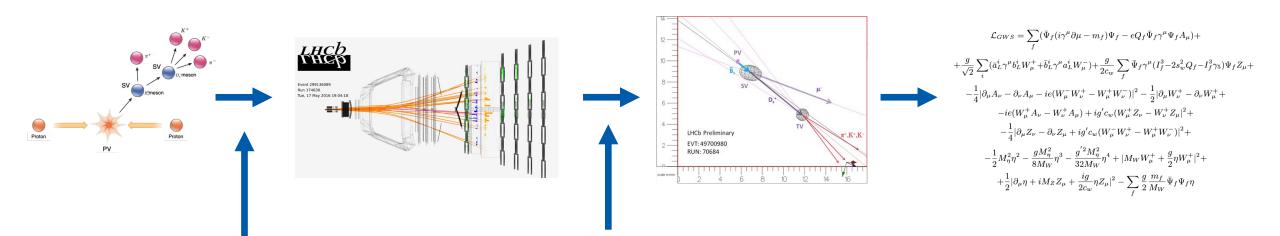


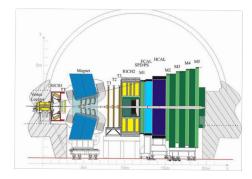
$$\begin{split} \mathcal{L}_{GWS} &= \sum_{f} (\bar{\Psi}_{f}(i\gamma^{\mu}\partial\mu - m_{f})\Psi_{f} - eQ_{f}\bar{\Psi}_{f}\gamma^{\mu}\Psi_{f}A_{\mu}) + \\ &+ \frac{g}{\sqrt{2}}\sum_{i} (\bar{a}_{L}^{i}\gamma^{\mu}b_{L}^{i}W_{\mu}^{+} + \bar{b}_{L}^{i}\gamma^{\mu}a_{L}^{i}W_{\mu}^{-}) + \frac{g}{2c_{w}}\sum_{f}\bar{\Psi}_{f}\gamma^{\mu}(I_{f}^{3} - 2s_{w}^{2}Q_{f} - I_{f}^{3}\gamma_{5})\Psi_{f}Z_{\mu} + \\ &- \frac{1}{4}|\partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu} - ie(W_{\mu}^{-}W_{\nu}^{+} - W_{\mu}^{+}W_{\nu}^{-})|^{2} - \frac{1}{2}|\partial_{\mu}W_{\nu}^{+} - \partial_{\nu}W_{\mu}^{+} + \\ &- ie(W_{\mu}^{+}A_{\nu} - W_{\nu}^{+}A_{\mu}) + ig'c_{w}(W_{\mu}^{+}Z_{\nu} - W_{\nu}^{+}Z_{\mu}|^{2} + \\ &- \frac{1}{4}|\partial_{\mu}Z_{\nu} - \partial_{\nu}Z_{\mu} + ig'c_{w}(W_{\mu}^{-}W_{\nu}^{+} - W_{\mu}^{+}W_{\nu}^{-})|^{2} + \\ &- \frac{1}{2}M_{\eta}^{2}\eta^{2} - \frac{gM_{\eta}^{2}}{8M_{W}}\eta^{3} - \frac{g'^{2}M_{\eta}^{2}}{32M_{W}}\eta^{4} + |M_{W}W_{\mu}^{+} + \frac{g}{2}\eta W_{\mu}^{+}|^{2} + \\ &+ \frac{1}{2}|\partial_{\mu}\eta + iM_{Z}Z_{\mu} + \frac{ig}{2c_{w}}\eta Z_{\mu}|^{2} - \sum_{f} \frac{g}{2}\frac{m_{f}}{M_{W}}\bar{\Psi}_{f}\Psi_{f}\eta \end{split}$$





Data reconstruction





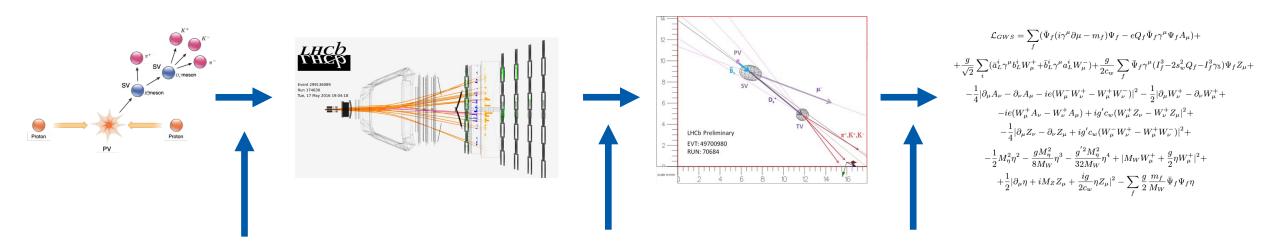
Data Collection

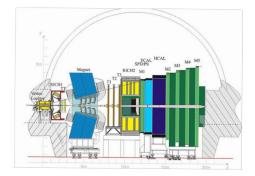


Data Reconstruction

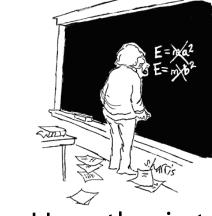
D. Derkach, NRU HSE

Hypothesis testing









Hypothesis testing

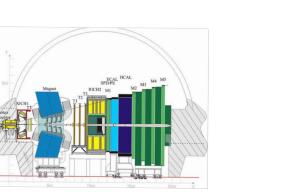


Tuning the process

Protor

 $\mathcal{L}_{GWS} = \sum_{i} (\bar{\Psi}_f (i\gamma^\mu \partial \mu - m_f) \Psi_f - eQ_f \bar{\Psi}_f \gamma^\mu \Psi_f A_\mu) +$ $+\frac{g}{\sqrt{2}}\sum_i (\bar{a}_L^i\gamma^\mu b_L^iW_\mu^+ + \bar{b}_L^i\gamma^\mu a_L^iW_\mu^-) + \frac{g}{2c_w}\sum_\ell \bar{\Psi}_f\gamma^\mu (I_f^3 - 2s_w^2Q_f - I_f^3\gamma_5)\Psi_fZ_\mu +$ $-\frac{1}{4}|\partial_{\mu}A_{\nu}-\partial_{\nu}A_{\mu}-ie(W_{\mu}^{-}W_{\nu}^{+}-W_{\mu}^{+}W_{\nu}^{-})|^{2}-\frac{1}{2}|\partial_{\mu}W_{\nu}^{+}-\partial_{\nu}W_{\mu}^{+}+$ $-ie(W^+_{\mu}A_{\nu} - W^+_{\nu}A_{\mu}) + ig'c_w(W^+_{\mu}Z_{\nu} - W^+_{\nu}Z_{\mu})^2 +$ $-rac{1}{4}|\partial_{\mu}Z_{
u}-\partial_{
u}Z_{\mu}+ig'c_w(W^-_{\mu}W^+_{
u}-W^+_{\mu}W^-_{
u})|^2+$ $-\frac{1}{2}M_{\eta}^{2}\eta^{2}-\frac{gM_{\eta}^{2}}{8M_{W}}\eta^{3}-\frac{g^{'2}M_{\eta}^{2}}{32M_{W}}\eta^{4}+|M_{W}W_{\mu}^{+}+\frac{g}{2}\eta W_{\mu}^{+}|^{2}+$ $+rac{1}{2}|\partial_{\mu}\eta+iM_{Z}Z_{\mu}+rac{ig}{2c_{w}}\eta Z_{\mu}|^{2}-\sum_{f}rac{g}{2}rac{m_{f}}{M_{W}}ar{\Psi}_{f}\Psi_{f}\eta$

Data Reconstruction



Data Collection

D. Derkach, NRU HSE

LHCb ГНСр

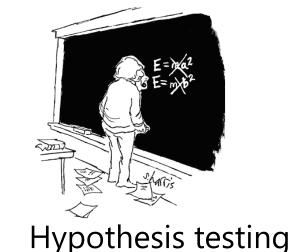
174630



LHCb Preliminary

EVT: 49700980

RUN: 70684

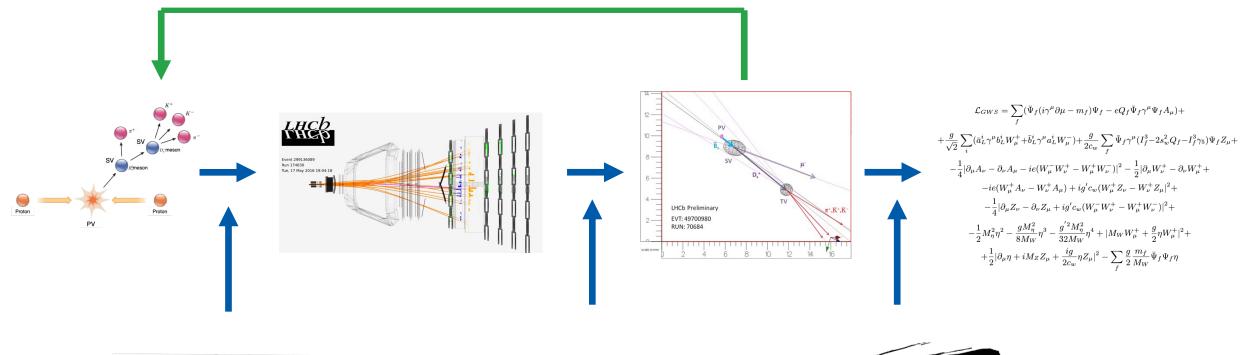


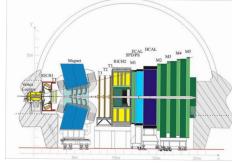




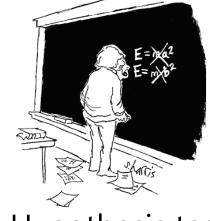
Tunable Data Analysis









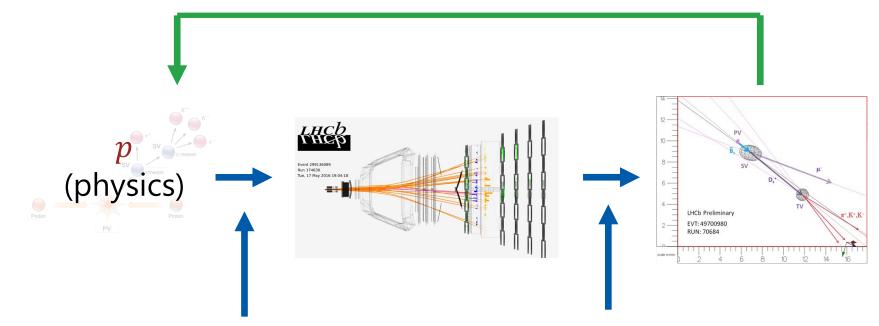


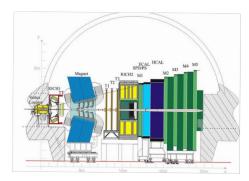
Hypothesis testing



Analyzing the Inputs





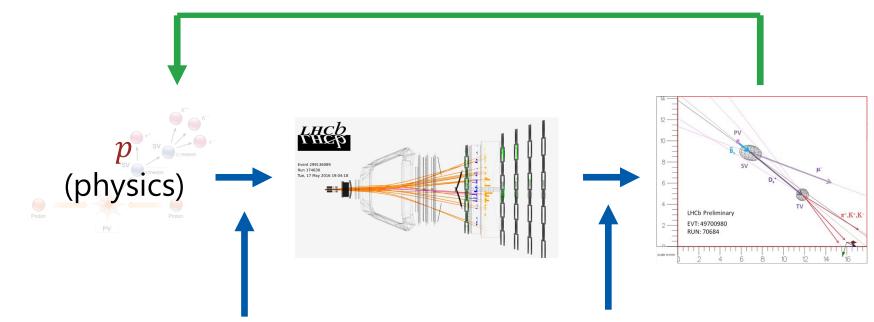


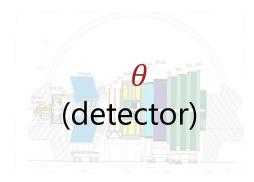
Data Collection



Building the Detector





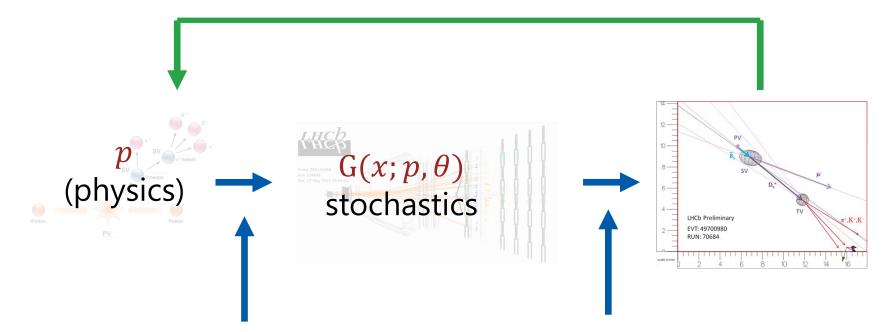


Data Collection



Obtaining the data





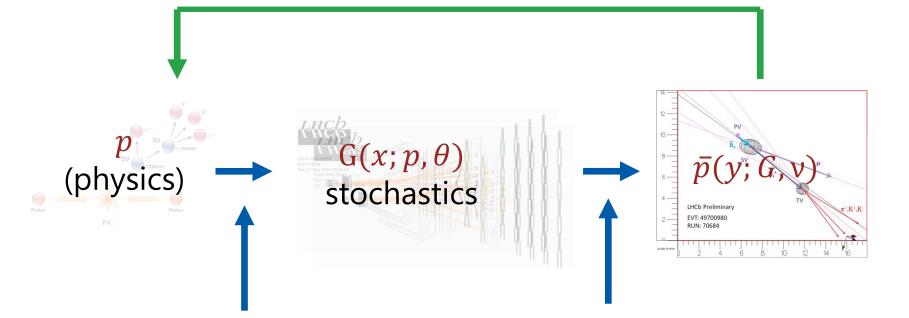




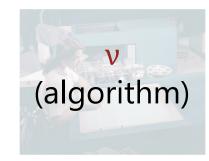
Data Collection D. Derkach, NRU HSE

Understanding the data





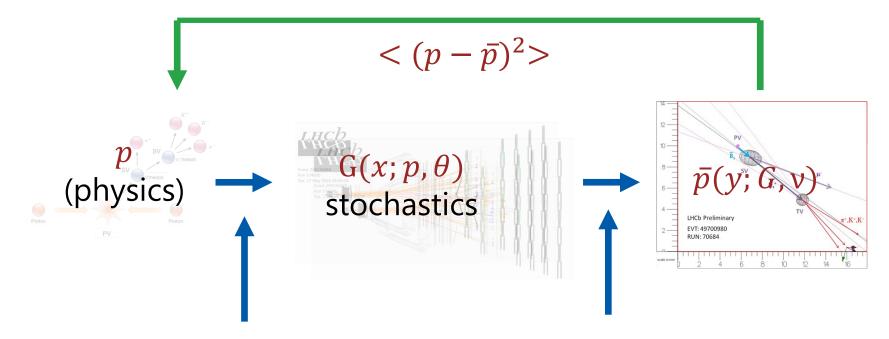


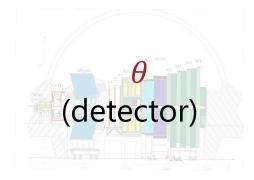


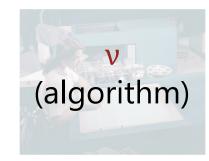


Best performance metrics





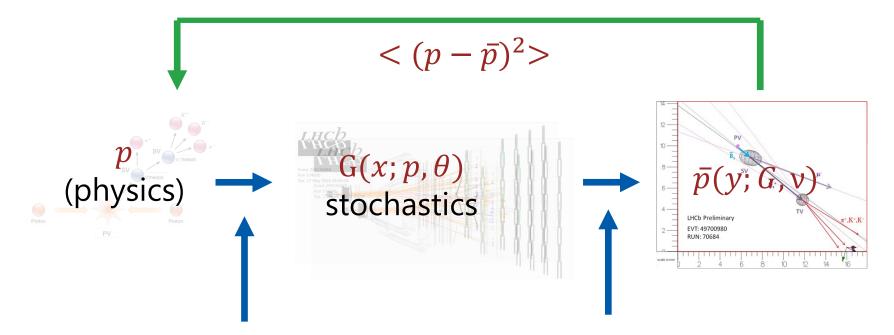




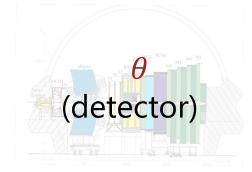


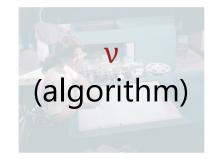
Minimizing the cost





 $L(\mathbf{P}) \sim exp(\mathbf{P}_{fix} - \mathbf{P})$

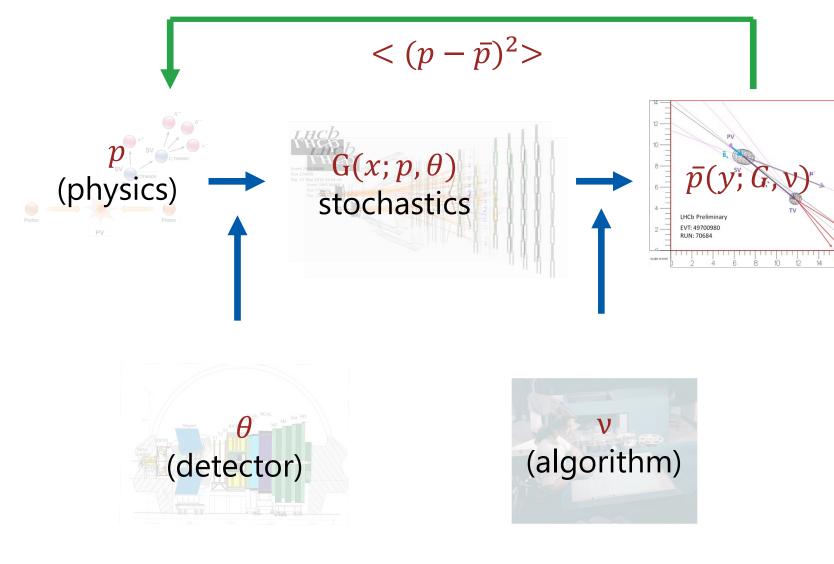






Minimizing the Time

Automate



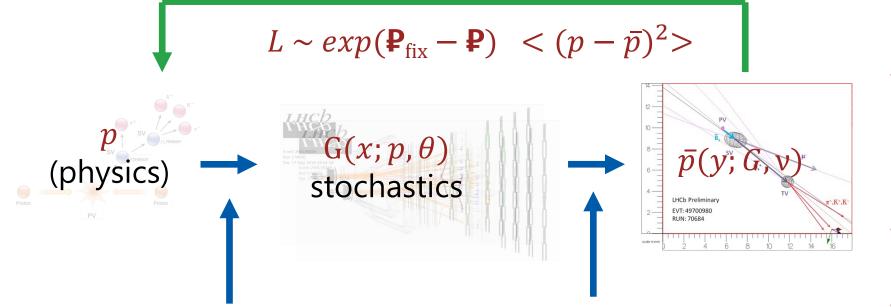


Data Reconstruction

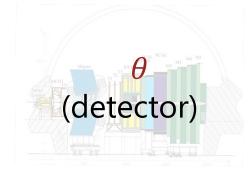
 $L \sim exp(\mathbf{P}_{fix} - \mathbf{P})$ $\mathbf{P} = \mathbf{P}(\theta)$

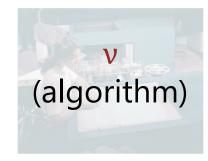
Difficulties





- $G(x; p, \theta)$ physics simulator
 - ~1event/minute
- ν chosen manually
- *L* not differentiable





SIMULATION D. Derkach, NRU HSE

Forward Problem



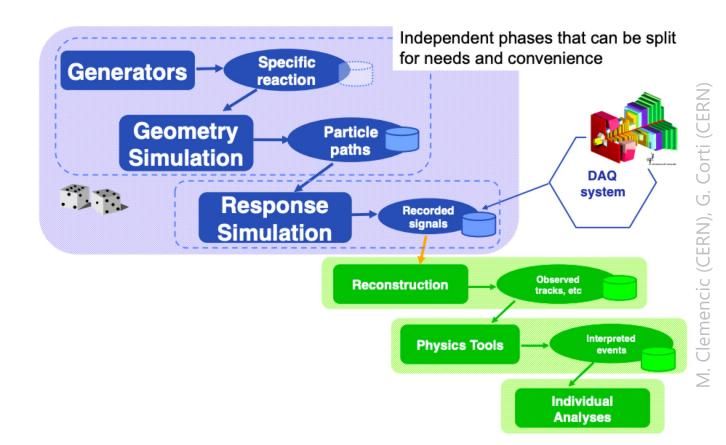
Simulation $G(x; p, \theta)$

Several model-motivated transitions.

Sequence:

- collision;
- decay;
- matter interaction;
- digitisation;
- reconstruction.

Each event takes 1 minute to generate (real world data is "generated" at several MHz).



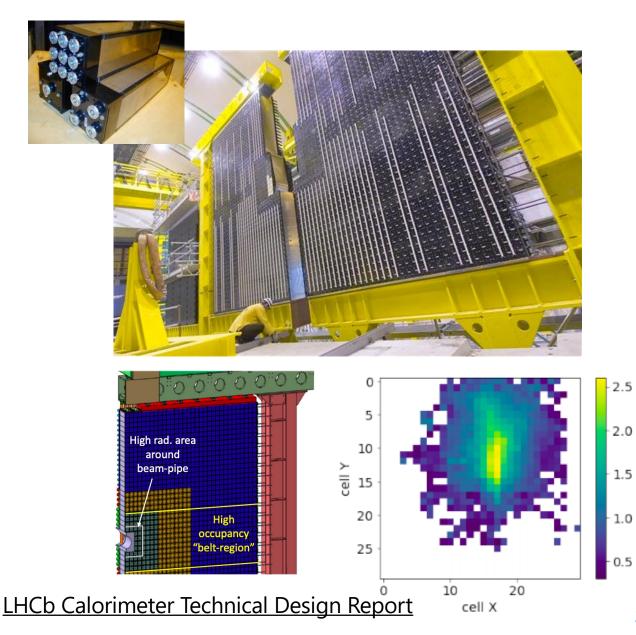
What is the event.

The calorimeter consists of many cells that reads out the energy deposit of a single particle.

The LHCb calorimeter construction motivated by the need to have better performance in the most populated regions.

A single particle deposits energy to several cells. An event is a sum of all particles and some noise.

We are normally in some reconstructed parameters of the event.



Ideas for Simulation

Since we know all processes in the subdetector, we can fully simulate an event using precise physics-motivated rules.

For calorimeters this means taking into account the structure of response that consists of many secondary particles.

This is done using Geant toolkit.

Pro: physics behind the simulation is controlled Cons: slow, needs fine tuning.

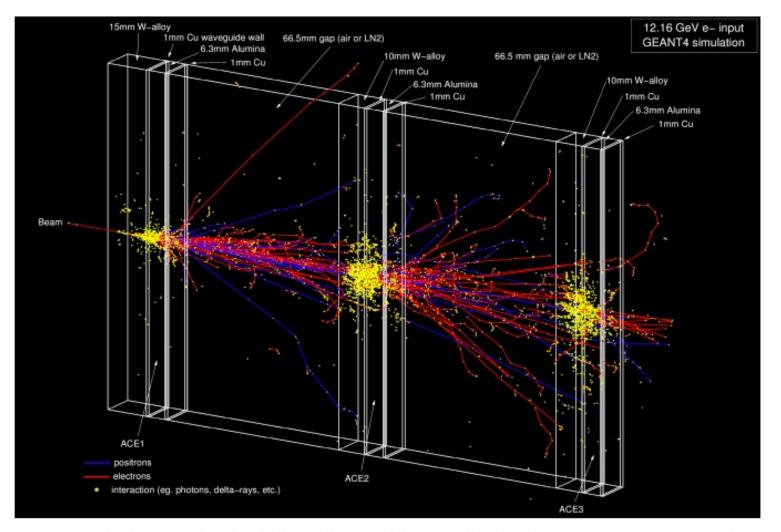
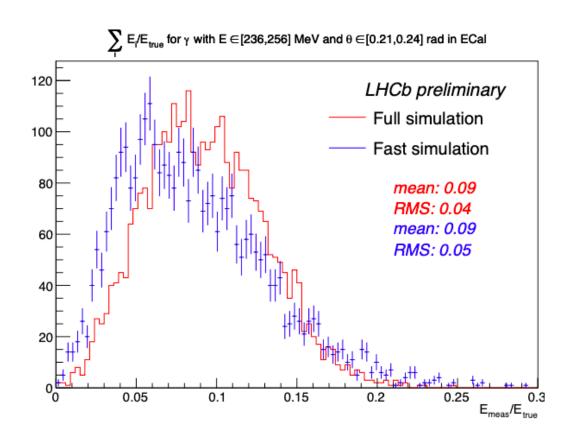


FIG. 2: Layout diagram, and GEANT4 simulation of a single 12.16 GeV electron event in our ACE detector system; in this case liquid nitrogen occupies the interelement spaces.

Ideas for Tabular Methods



Build a library of calorimeter responses to impact particle in corresponding 5D phase space using GEANT4 («frozen showers»).

5D = 3D momentum + 2D coordinate for every particle type.

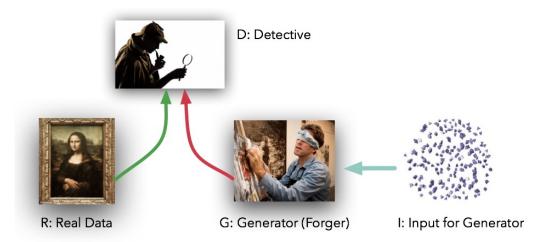
The whole phase space is split into bins, the exact observable is obtained interpolating between the bins.

One can also construct full interpolation (without using bins).

Pros: easy to interpret, quality is controlled by the number of samples.

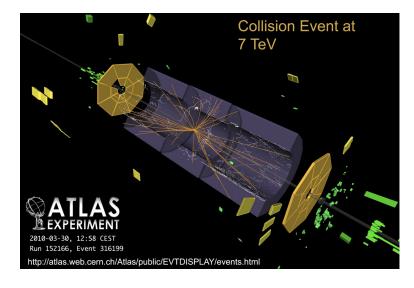
Cons: curse of dimensionality, memory consumption, full interpolation takes huge efforts.

Making it continuous



We can try to use the recent concept of Generative adversarial network, that can be characterised by the interplay between a forger and a detective.

At the Large Hadron Collider, nature is an artist. We need to learn how to fake these paintings.



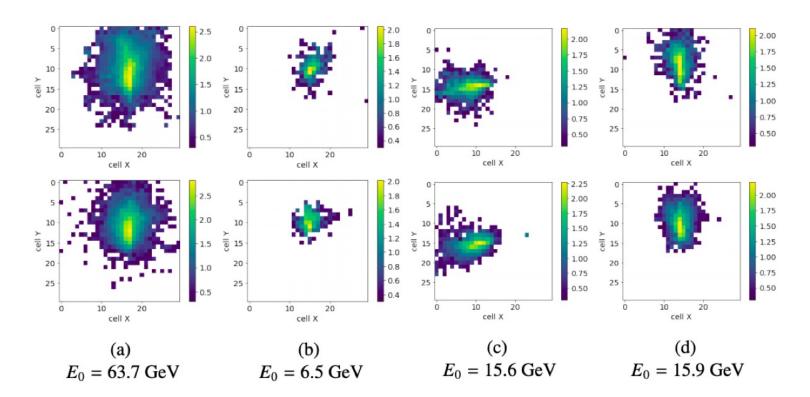
Generative modeling $G(x; p, \theta)$

Conditional

dependence on detector parameters θ and incident particle information p.

Need to be

- Tunable.
- Robust.
- Fast for sampling.



V. Chekalina et al. EPJ Web Conf. 214 (2019) 02034

Challenges: Implementations

More challenges:

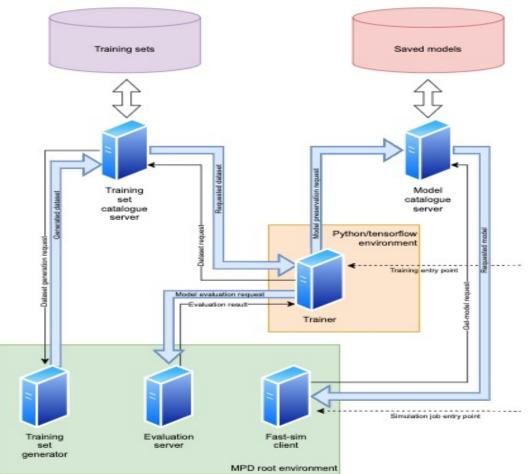
- Distilling the generators. Aim: beyond 100ms/event.
- Testing the generator quality in the limit of small data samples.

Aim: on-the-fly algorithms.

 Implementing pipeline in the online environment (200xNVidia RTX A5000 from LHCb).

Aim: Efficient architecture and Scheduling given resources.

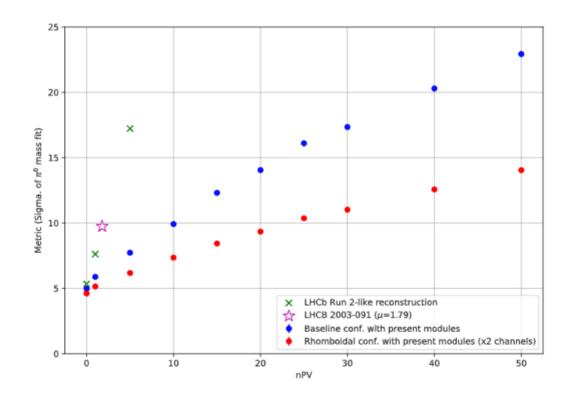
Expected pipeline (preliminary)



D. Zolotukhin (BSc Diploma) R. Drynkin (BSc Diploma) Inverse Problem



Reconstruction $v = v(\theta)$



Need to "get rid" of experts help in inverse problem

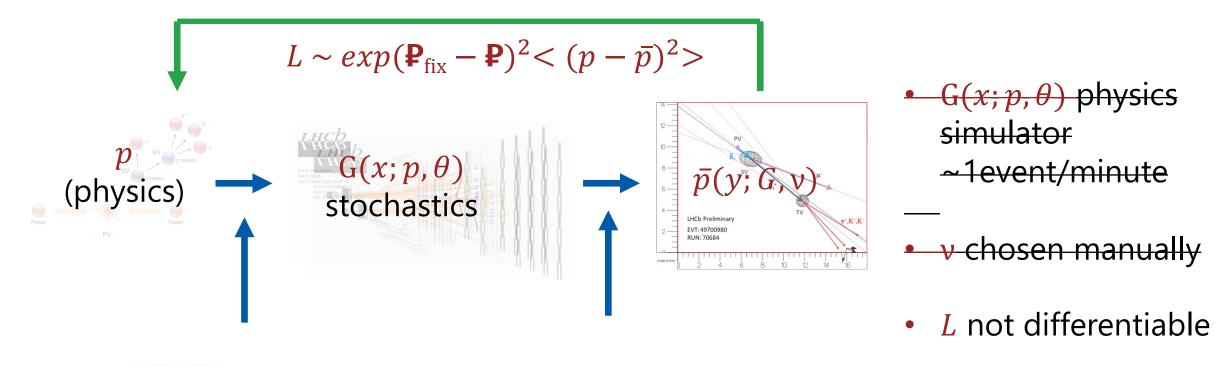
Current solution:

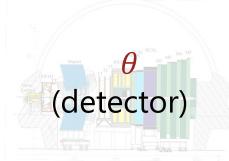
- Self-trained solution for every point from θ;
- Challenge: universal training for many points of θ .

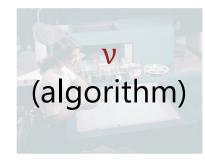
A Boldyrev et al. 2020 JINST 15 C09030

Difficulties











Data Reconstruction





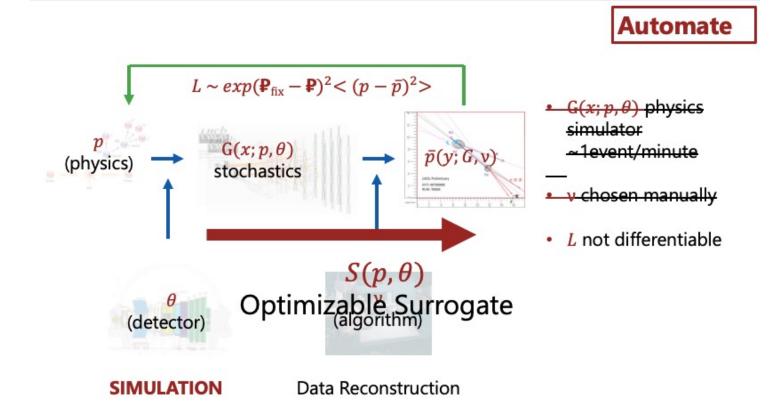
Black box optimization

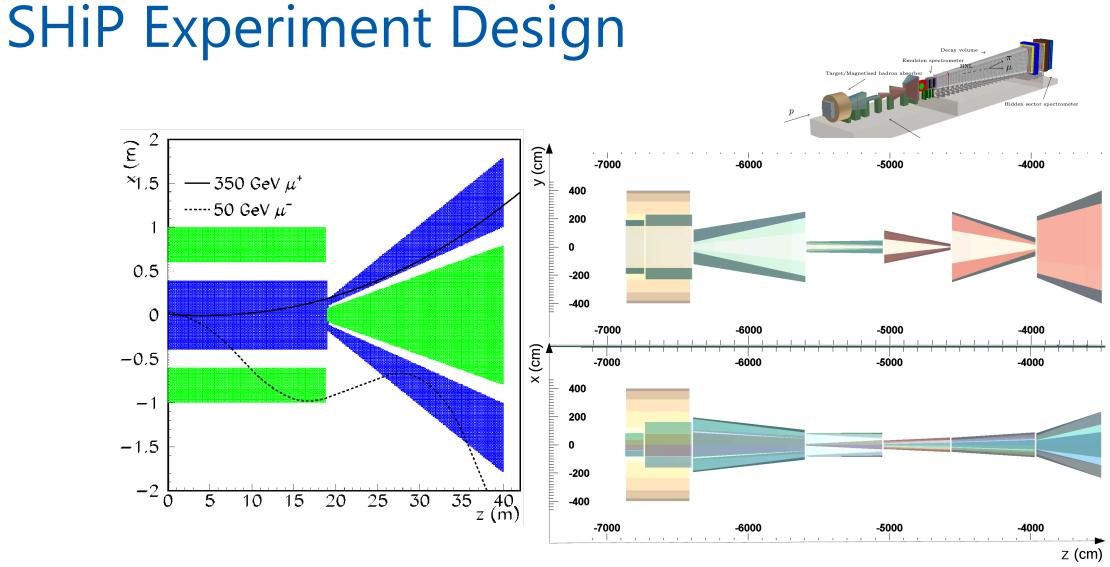
$$L \sim exp(\mathbf{P}_{\text{fix}} - \mathbf{P})^2 < (p - \bar{p})^2 >$$

- The value at any point is known.
- The analytical formula is unknown.
- The time to compute the value is several tens of hours.
- Classical black-box optimization problem:
 - Optimizing someone else's code (there is only a compiled library).
 - Systems described by differential equations (airplane wing shape).

Black Box Solution

- Expert Method
 - Might be wrong
- Random search.
 - Works sometimes.
- Surrogate modeling





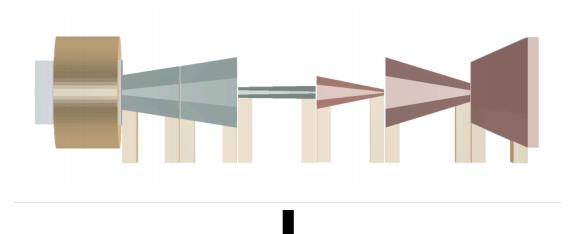
 Absorber shape optimization: background suppression at reasonable cost

Final Optimization

Surrogate modeling using Bayesian Optimization of Gaussian Processes.

Optimization brought 25% cheaper solution.



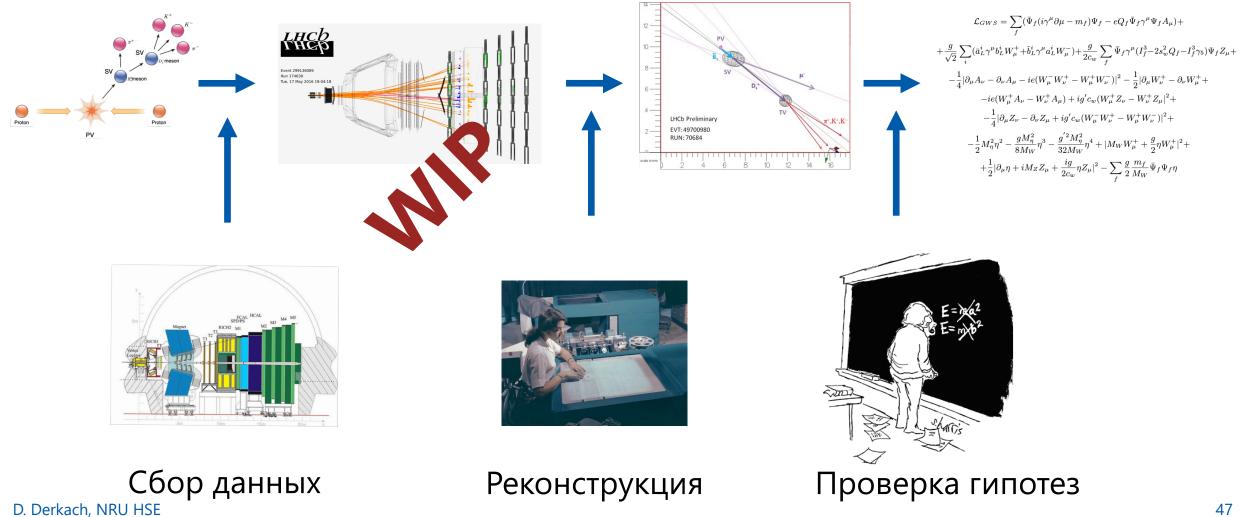


A. Filatov et al. Journal of Physics: Conference Series. 2017. Vol. 934. P. 1-5

Conclusions

- Optimization of large setups requires simultaneous approximate fast solutions of forward and inverse problems.
- Final stage of the optimization needs precise surrogate modeling.
- A complete optimization cycle brings in significant reduction in costs (and thus efficiencies) but requires several fundamental questions to be solved.
- Future outlook:
 - Many areas to be re-explored using new (data-driven) machine-learning approaches.
 - Many areas are already re-explored (<u>https://mode-collaboration.github.io/</u>)

Conclusions



Conclusions

