LOBACHEVSKY UNIVERSITY
UNIVERSITY OF GOTHENBURG
Institute of Applied Physics
Russian Academy of Sciences

Аспекты применения машинного обучения для задач физики лазерной плазмы

Юрий Родимков, Валентин Волокитин,
Елена Панова, Александр Панов,
Игорь Сурмин, Кирилл Тараканов,
Анастасия Аризова, Иосиф Мееров
ННГУ

Christoffer Olofsson,
Shikha Bhadoria,
Tom Blackburn,
Arkady Gonoskov
University of Gothenburg

Евгений Ефименко
ИПФ РАН

Алексей Половинкин
Adv Stat & Machine Learning, LTD, Intel
(USA)

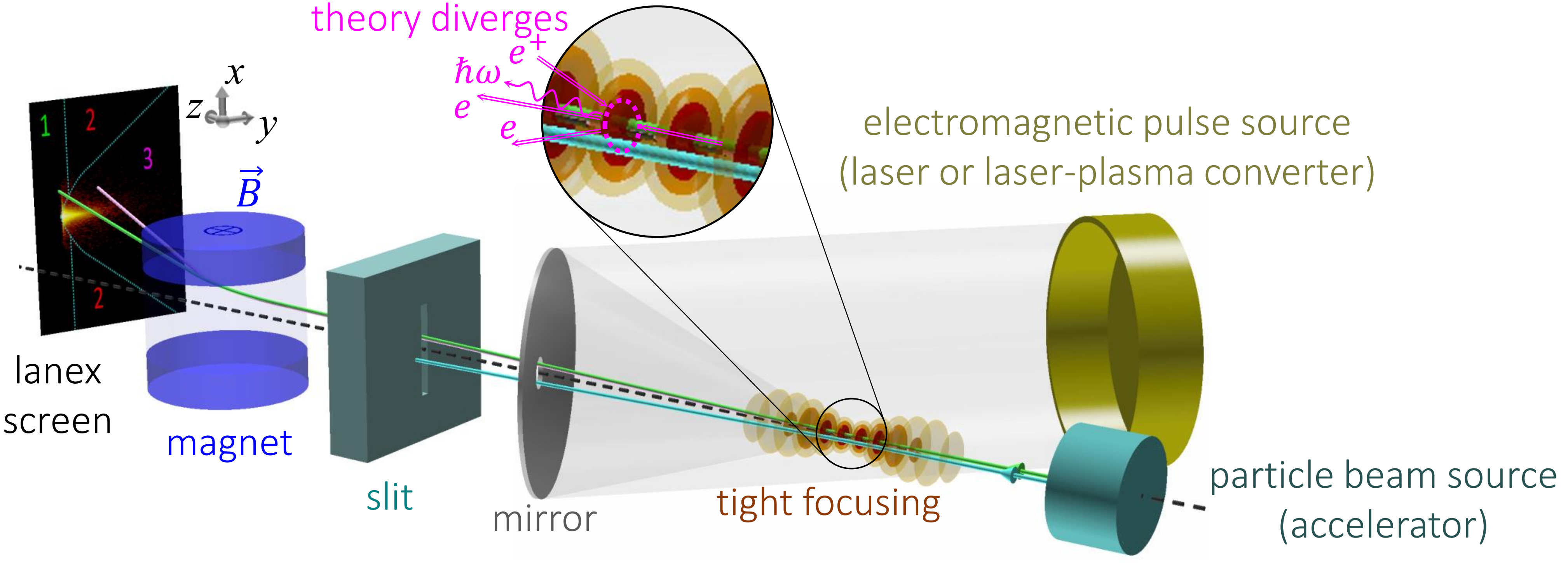
László Veisz
Umeå University
(Sweden)

2021-06-04

План семинара

- Контекст: от большой задачи физики к большой задаче ML (XAI)
- Модельные задачи и основные трудности:
  - Анализ вероятностных распределений (laser-particle collider)
  - Анализ данных стохастических процессов (laser-plasma converter)
  - Определение трудноизмеримых параметров (fine-tuning of tight focusing)
- Направления развития

Context: from grand challenge in physics to grand challenge in ML



Concept of fundamental experiments:

1. Reach high intensity

1^+. Use laser-plasma converter

2. Expand model: quantified deviations

3. Find favorable layout

4. Infer deviations from experiment

Olofsson & Gonoskov: (preprint in autumn 2021)

Role of ML:

a. Assisting 1, 1^+ and 4 (automation)

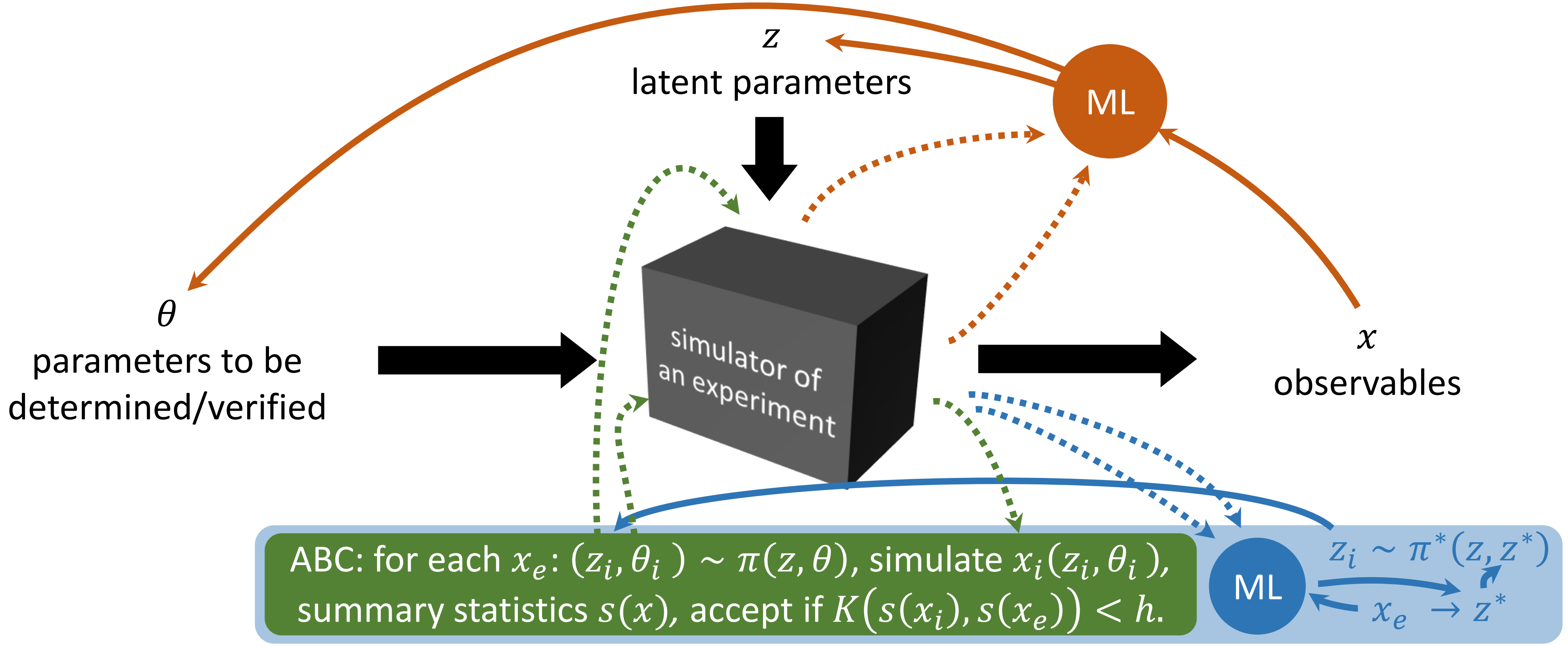
b. Elimination of latent variables

c. Accelerated convergence of ABC

Difficulties (role of XAI):

- Quantify errors
- Reach and quantify generalization capabilities
- Combat differences between simulated and real data
- Learn from ML results

The layout of an inverse problem

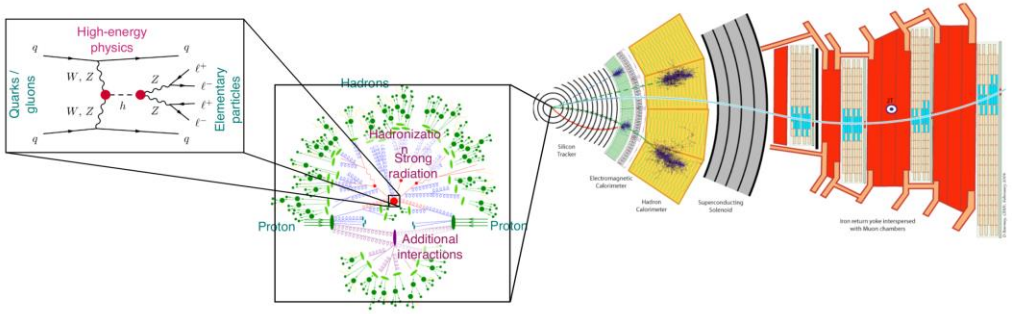


Problem: Infer θ from experimental data

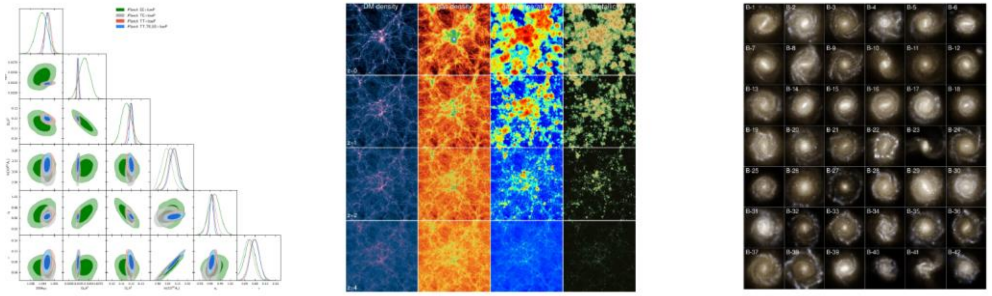
Approaches and difficulties:

1. ML: explanation and reliability (XAI is needed); irreversibility due to probabilistic or/and stochastic nature of the process; difference between experiment and simulation
2. Approximate Bayesian Computation (ABC): large dimensionality of x and z make the likelihood function intractable (requires integration over all possible outcomes)
3. ABC+ML (elimination of z): error quantification ($\pi^*(z, z^*)$)

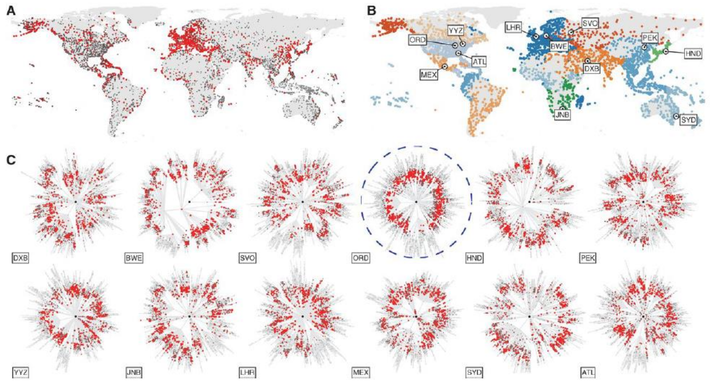
Similar problem statements appear in many areas



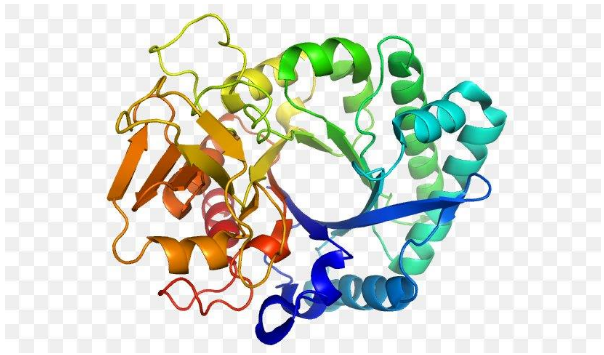
Particle physics



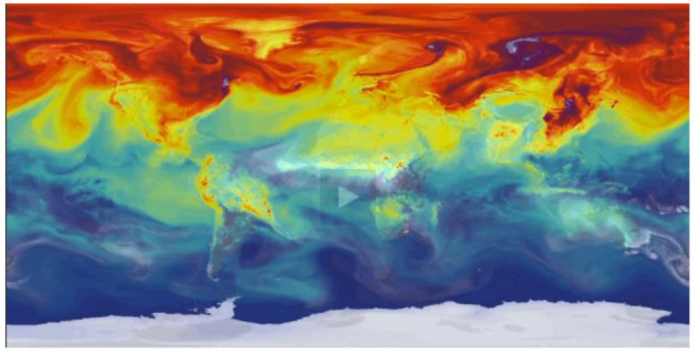
Astrophysics and cosmology



Epidemiology



Protein folding



Climate science

Credits: Gilles Louppe, Machine Learning and Physics Series, 2019

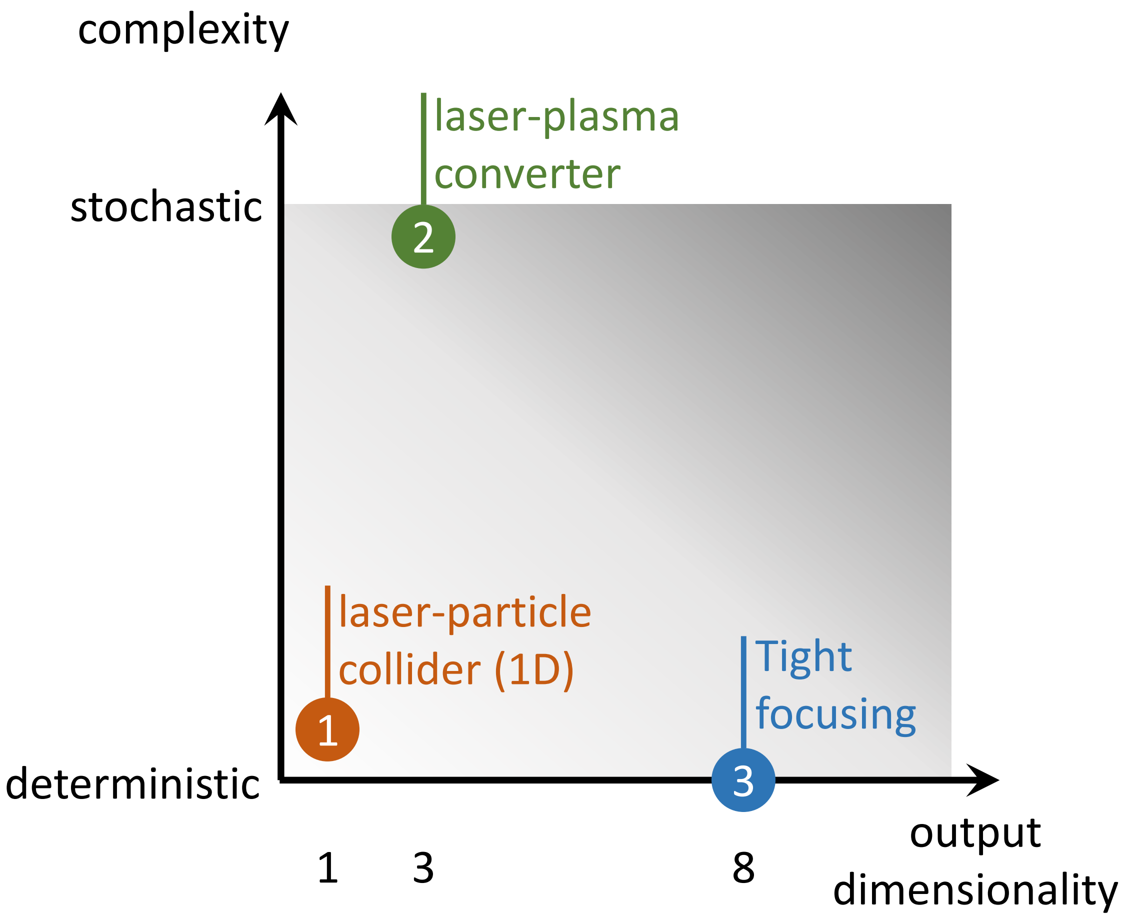
The role of XAI

Incremental improvements:

- Achieve narrower distribution of errors
- Quantify upper limits for error distribution
- Identify reliable cutoffs

Game-changing improvements:

- Generalizability (simulations → experiment)
- Overcoming irreversibility (detect, explain)
- Reliability (retrieve sufficient summary statistics; identify indicative features)

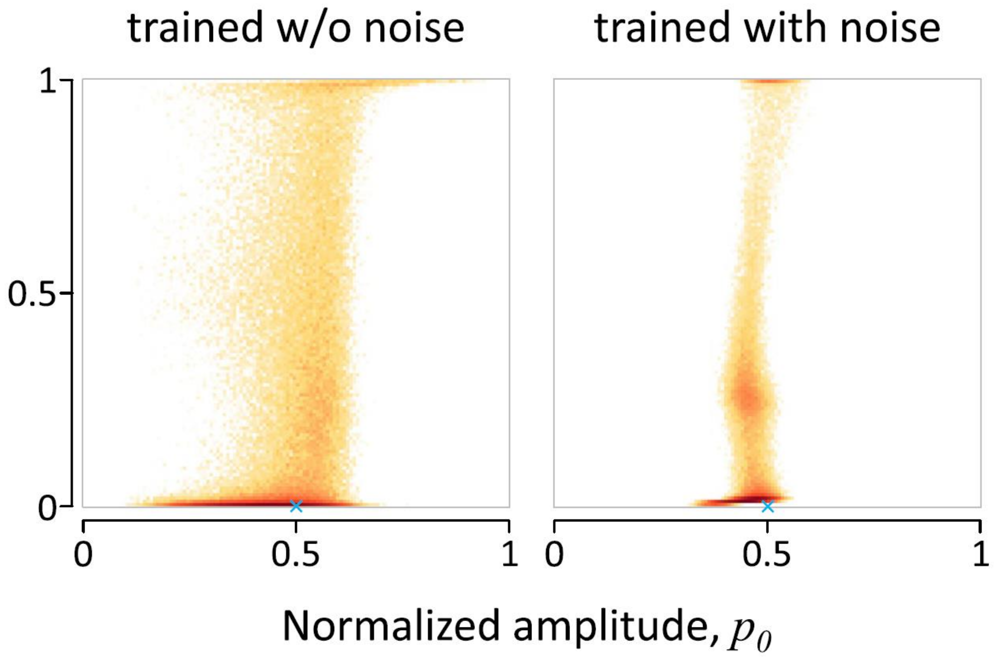


Methodology:

1. Identify ML models tolerant to noise (varied by binning strategy): use noise to enhance generalizability.
2. Transfer learning: (1) pre-train using simplified analytical models (uncostly data) to accentuate indicative features; (2) generalize using ab-initio simulations (cheap data); (3) fine-tune using actual experiment (expensive data).
3. Identify features by visualizing the direction of maximized change of an output parameter.

Problem 1: motivation

- Suppose we train an ML model to infer the amplitude and variance of a given Gaussian distribution.
- The are many solutions, e.g. the amplitude can be inferred directly from the value at center, and the variance via the value at a single off-center point.
- However this solution doesn't reflect the idea of amplitude and variance and may work poorly for experimental data.
- Hypothesis: by adding noise to the training data we can favor learning more "integral" properties, because they are more tolerant to noise.

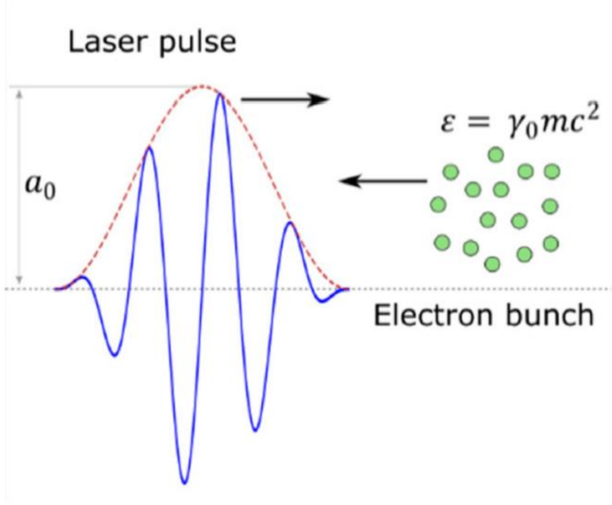


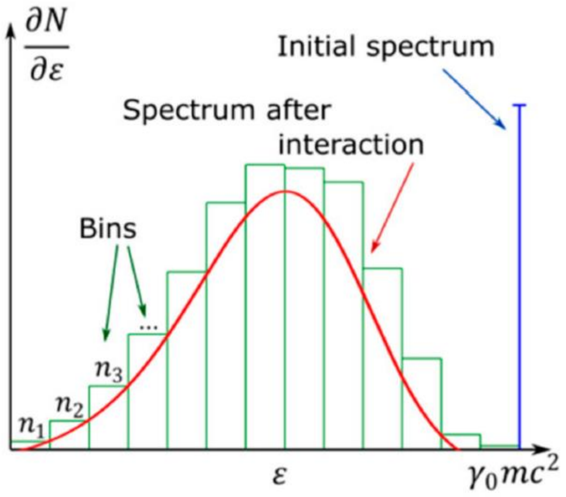
Gonoskov et al. SciRep **9**, 7043 (2019)

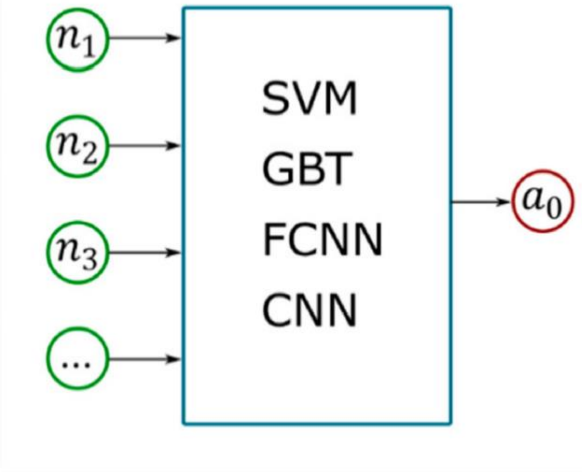
Concept:

- The level of noise can be naturally controlled by size of the bins (the size of the input vector)
- Inappropriate noise level (size of input vector) can cause deterioration of ML training process. Thus Ml models tolerant to noise are favorable.

Problem 1: problem statement





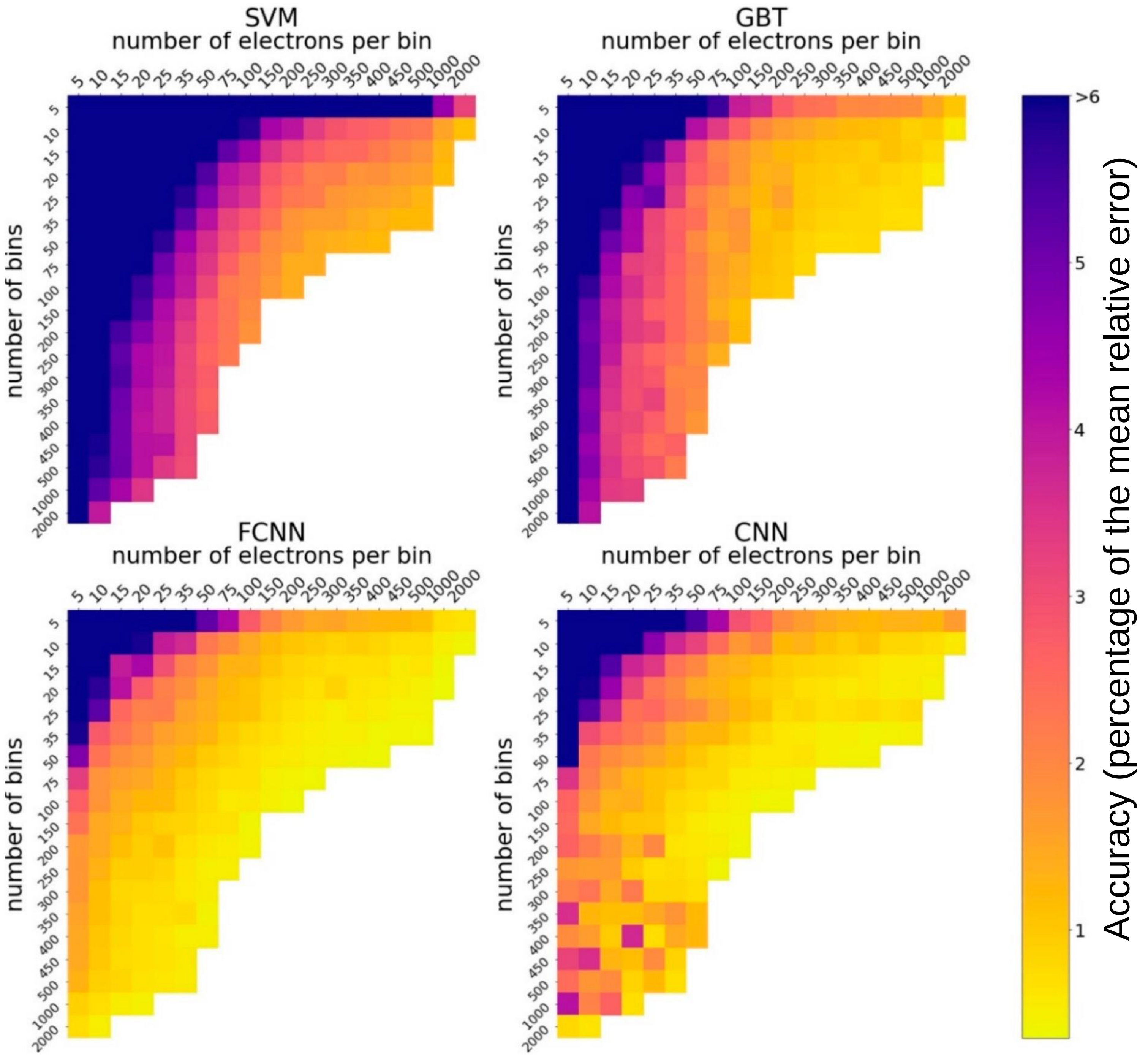


- We simulate head-on collision of an ultra-intense laser pulse with an electron bunch (hi-χ framework).
- We obtain electron energy distribution.
- Obtained distributions are used as an input for different ML methods that are trained to determine dimensionless amplitude of the laser pulse a_0.
- We varied the bin size to determine tolerance of ML models to noise.

Y. Rodimkov et al. ML-Based Analysis of Particle Distributions in High-Intensity Laser Experiments: Role of Binning Strategy, Entropy, 23 (1), 21 (2020)

Problem 1: results

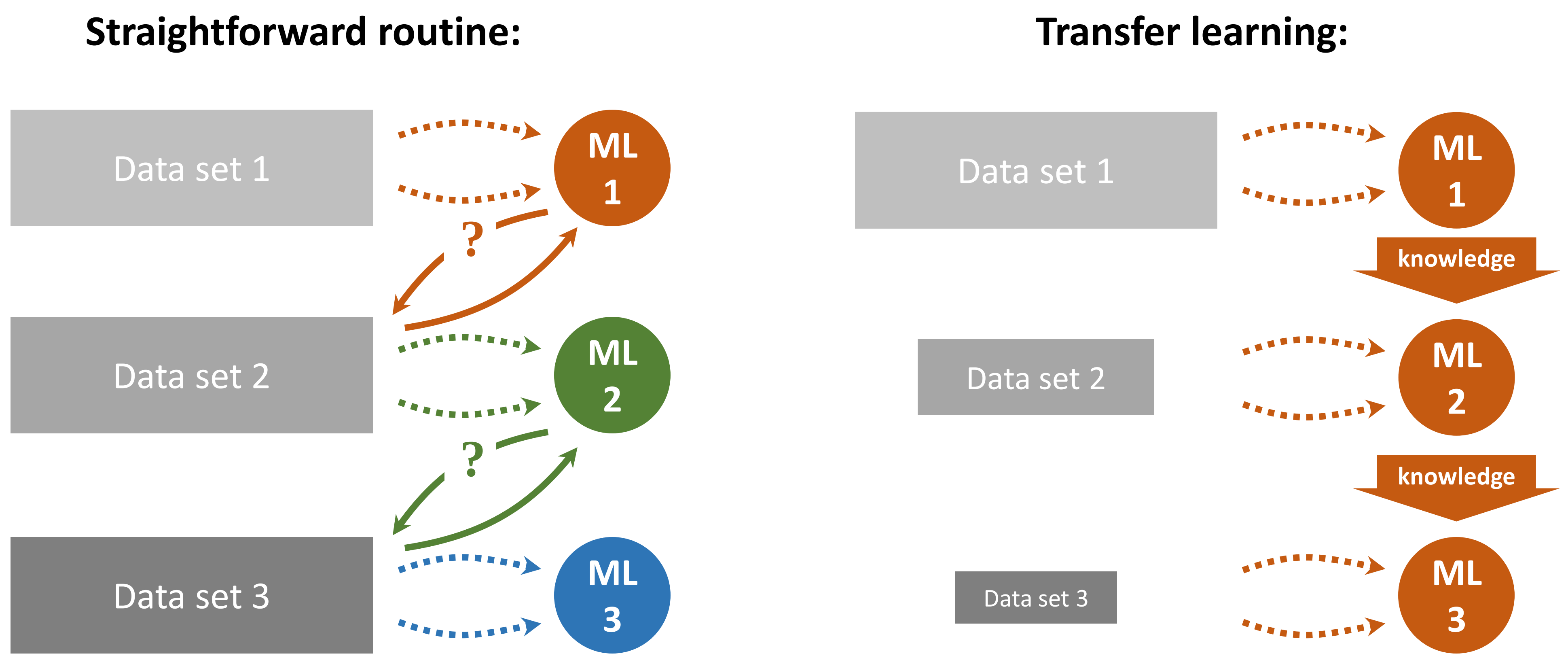
Measure	FCNN	CNN	PCA+FCNN
Mean absolute error	1.784	1.827	2.000
Mean relative percentage error	0.512	0.496	0.709
Coefficient of determination	0.99993	0.99992	0.99991



Conclusions:

- Non-optimal binning can crucially deteriorate the performance of SVM and GBT, and, to a less extent, FCNN and CNN.
- PCA (linear) can reduce training time at the cost of minor accuracy deterioration, but doesn't provide higher accuracy overall.

Problem 2: motivation

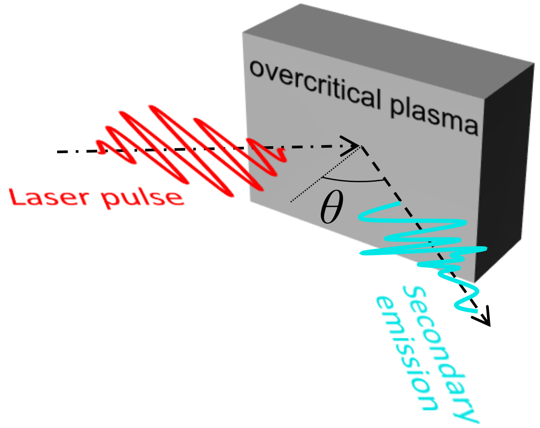


Concept:

- pre-train using simplified analytical models (uncostly data) to accentuate indicative features
- generalize using ab-initio simulations (cheap data)
- fine-tune using actual experiment (expensive data).

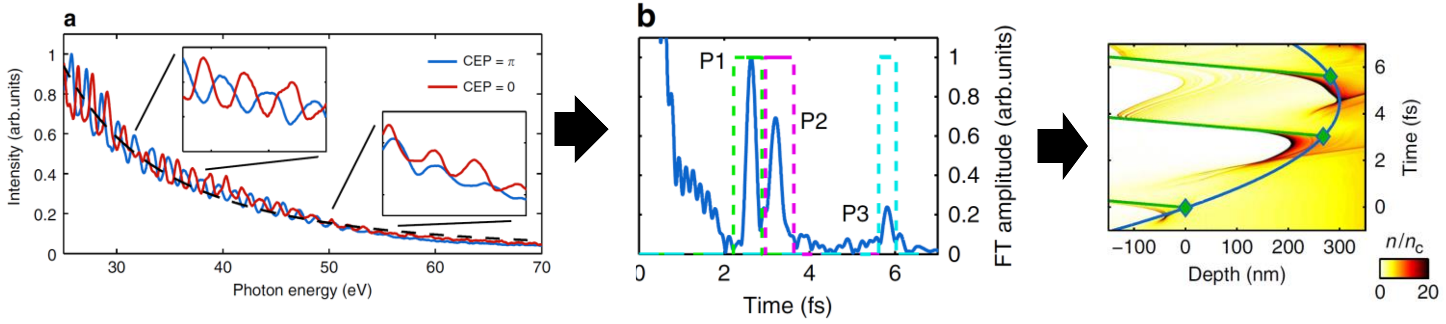
Problem 2: problem statement

Process: An intense few-cycle laser pulse with some carrier envelope phase (CEP) impinges on an overdense plasma target at some incidence angle and causes the generation of secondary radiation.



Problem: infer CEP, pre-plasma scale length and angle of incidence from the spectrum of secondary emission (the only routinely measurable data).

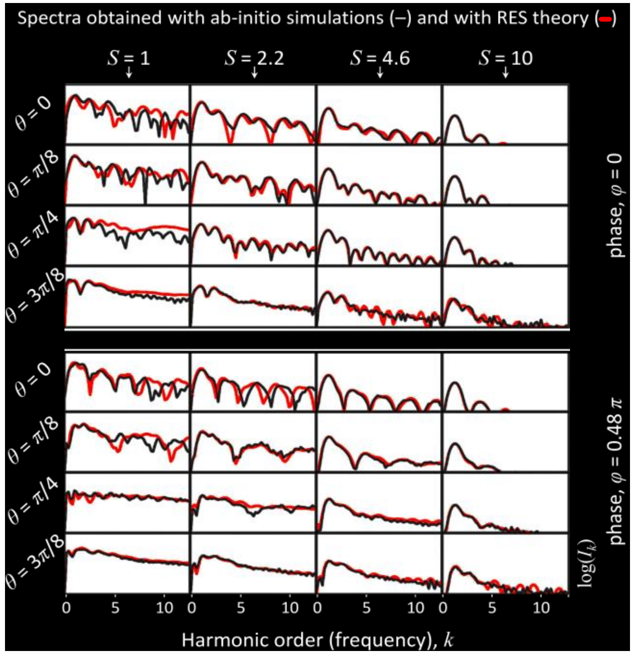
Spectral interferometry (works only for 2-3 cycle pulses, known parameters):



D. Kormin et al. Nat. Comm. 9, 4992 (2018)

Goals:

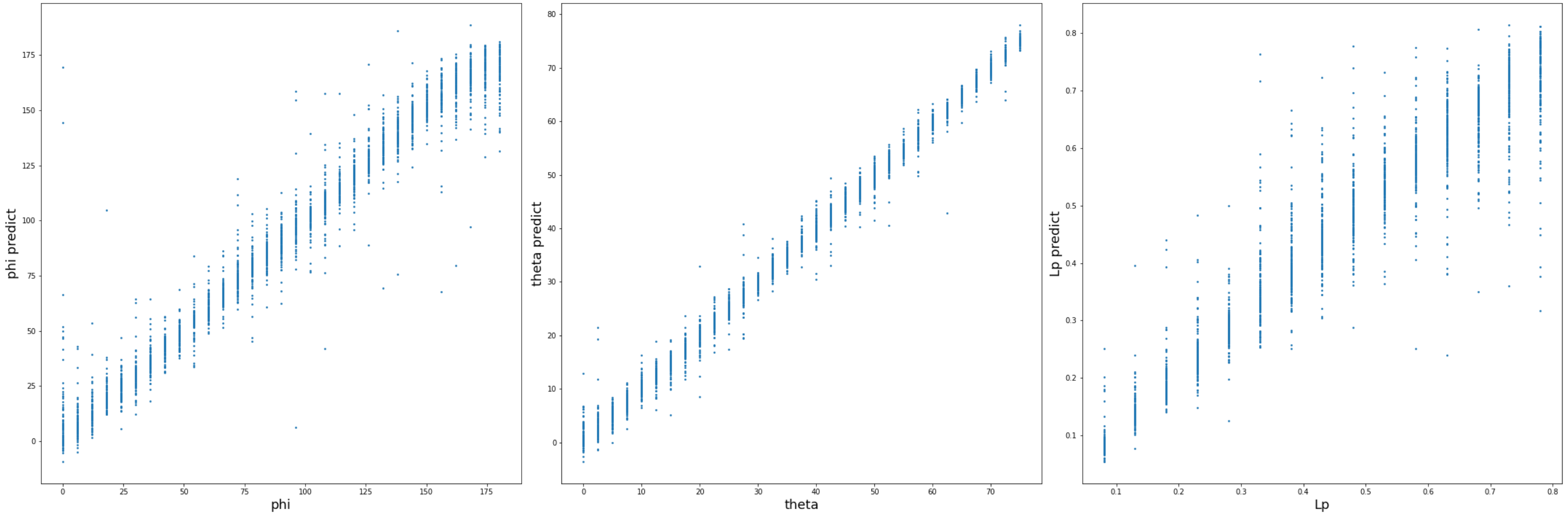
- use ML to learn more general features (reconstruct more parameters)
- apply transfer learning to reach applicability for experimental data
  RES (~1 ms per simulation), PIC ($\sim$ 1 min, $\sim 10^5$ cases), experiment ($\sim 10^3$ cases)
- determine (highlight?) indicative features



Gonoskov et al., Sci. Rep. 9 (1), 1-15 (2019)
RES model: Gonoskov, Phys. Plasmas (2018)

Problem 2: results, FCNN

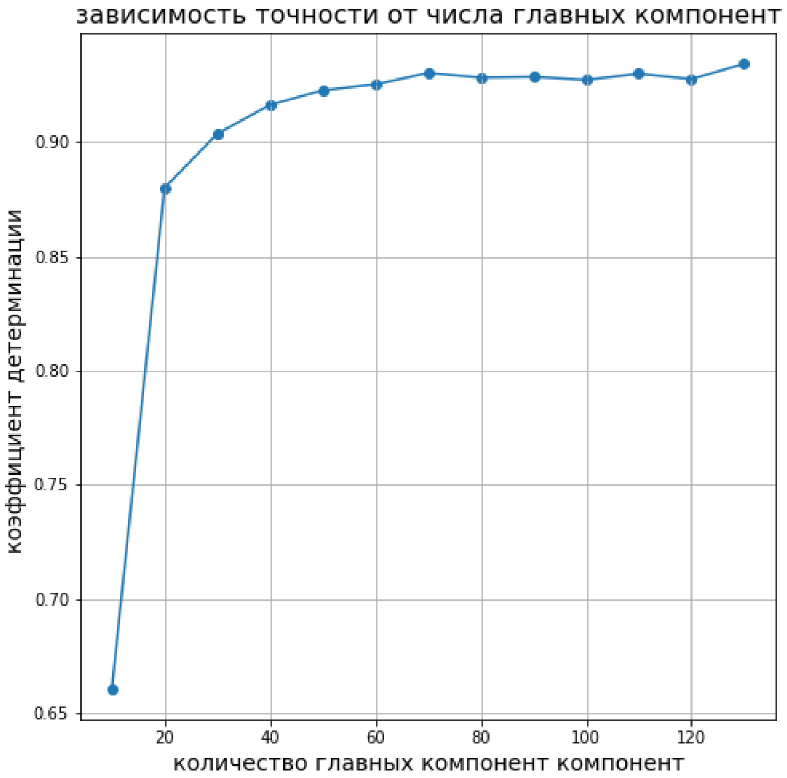
Обучение полносвязной нейронной сети
Восстановление параметров на основе спектра проводилось с помощью полносвязной архитектуры нейронной сети с количеством нейронов на каждом слое: 1024, 512, 256, 256, 128, 32, 16 с функциями активации ReLU. Последний слой содержал три нейрона с линейной функцией активации. Нейронная сеть обучалась 400 эпох. Использовался оптимизатор Adam. Параметр скорости обучения изначально был равен 0,005, а после 250 эпохи уменьшался умножением на коэффициент 0,99 после каждой эпохи.



Метрика	ϕ_{cep}	θ_{inc}	L_p
Относительная ошибка в процентах	4.076	1.968	5.851
Коэффициент детерминации	0.947	0.990	0.894

Problem 2: results, PCA + FCNN

В качестве оптимального результата был рассмотрен вариант использования 10 главных компонент. Для данного числа новых признаков была произведена тонкая настройка нейронной сети, в результате которой была выбрана полносвязная модель сети с числом нейронов на каждом слое: 512, 256, 256, 256,128, 32, 16 с функциями активации ReLU. Нейронная сеть обучалась 300 эпох. Параметр скорости обучения изначально был равен 0,005.



Метрика	ϕ_{cep}	θ_{inc}	L_p
Относительная ошибка в процентах	3.807	1.847	5.364
Коэффициент детерминации	0.957	0.989	0.916

Problem 2: results, pruning, pre-processing (filtering)

Pruning

Использовался весь спектр (не только главные компоненты). Сеть обучалась с такими же параметрами, после чего проходило дополнительное обучение сети. В ходе дополнительных 200 эпох малые значения обращались в ноль, а другие донастраивались для решения поставленной задачи. Таким образом, 80 процентов весов на каждом слое было обращено в ноль.

Метрика	ϕ_{cep}	θ_{inc}	L_p
Относительная ошибка в процентах	3,663	1,938	5,622
Коэффициент детерминации	0.959	0.990	0.902

Filtering of input data

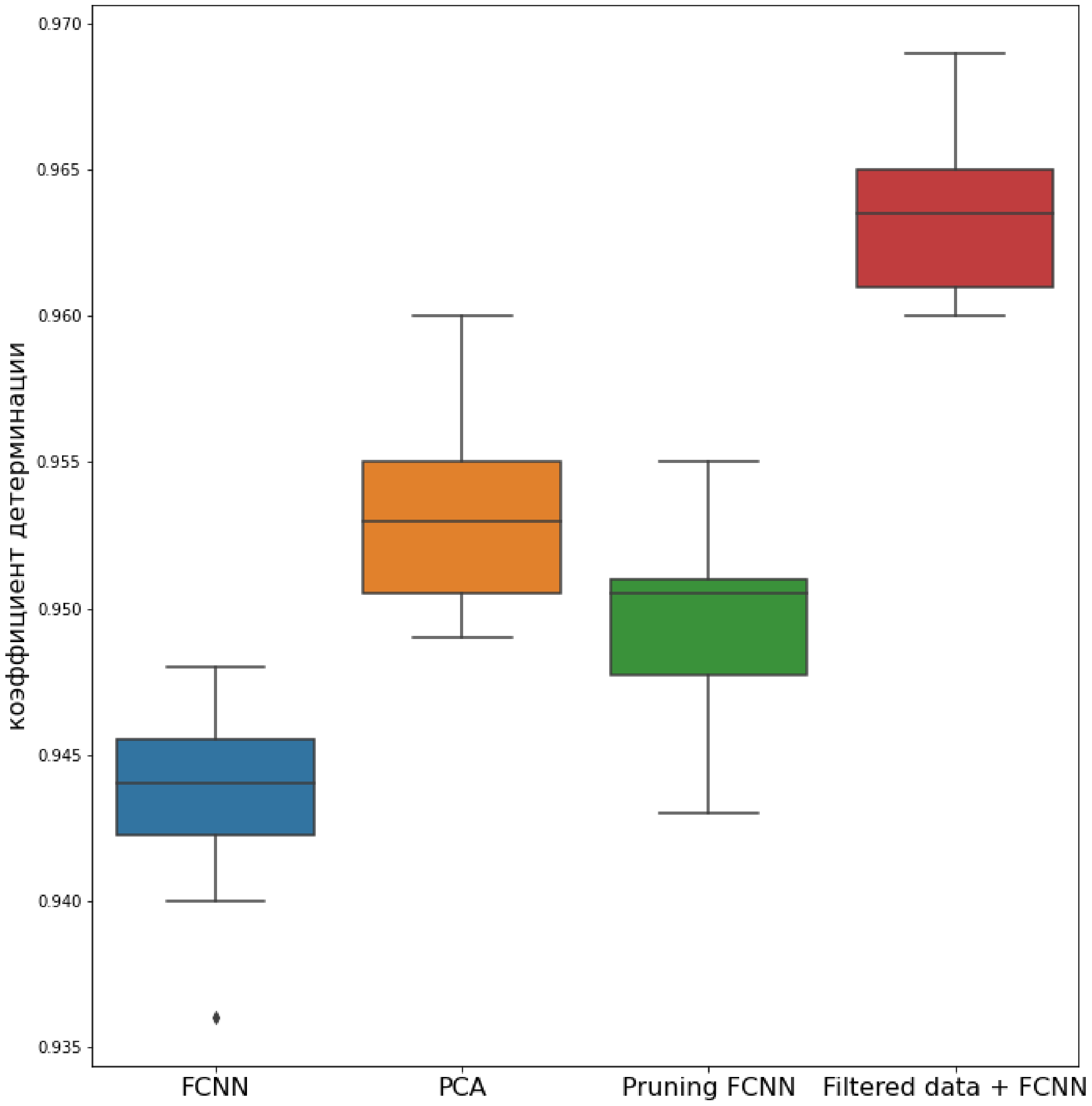
Сложность структуры исследуемых спектров подталкивает на идею использования фильтров цифровой обработки сигналов. Для этого использовался фильтр с бесконечной импульсной характеристикой Баттерворта.

Метрика	ϕ_{cep}	θ_{inc}	L_p
Относительная ошибка в процентах	3,186	1,544	4,848
Коэффициент детерминации	0.963	0.993	0.924

Problem 2: summary of results

Intermediate conclusion:
- Use of PCA, pruning and data pre-processing can increase accuracy, but not dramatically.

Plans:
- Identify whether the FCNN pre-trained by analytical model can achieve better performance on the data of numerical experiments.
- Consider ML models trained either only by simulations or by analytical model and then simulations. Measure performance of these ML models on real experimental data.



Problem 3: motivation

Dealing with high-dimensional output vectors is practically valuable because:

- one can mitigate difference between simulated and real data by augmenting numerical model with extra parameters that mimic/model variety or secondary factors.
- complex problems and experiments naturally include many parameters.

Difficulties:

- The multidimensional space of output parameters is difficult to inspect: it can be difficult to guarantee absence of sub-manifolds where the inverse problem is ill-posed or unstable (irreversibility, stochasticity, saddle points, etc.).
- If a good result of ML model is achieved, one can expect that the problem is “separable”, but it is unclear how to get related insight from the ML model.

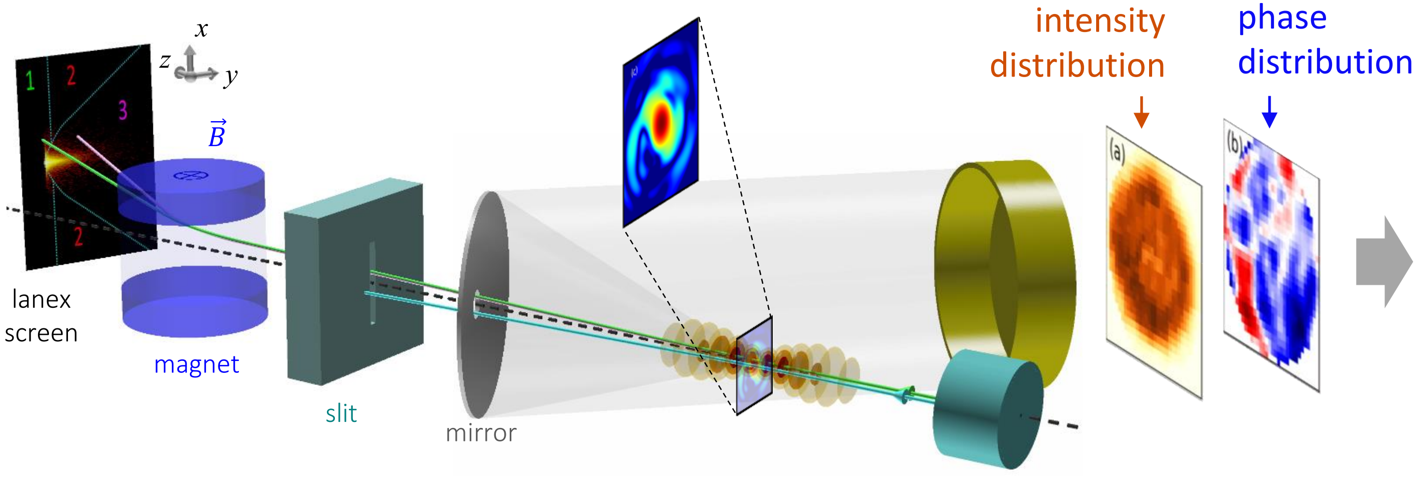
Traditional solution:

- Understand what cases are indistinguishable or difficult for distinguishing and redefine/restrict output vectors or extend the input vectors accordingly.

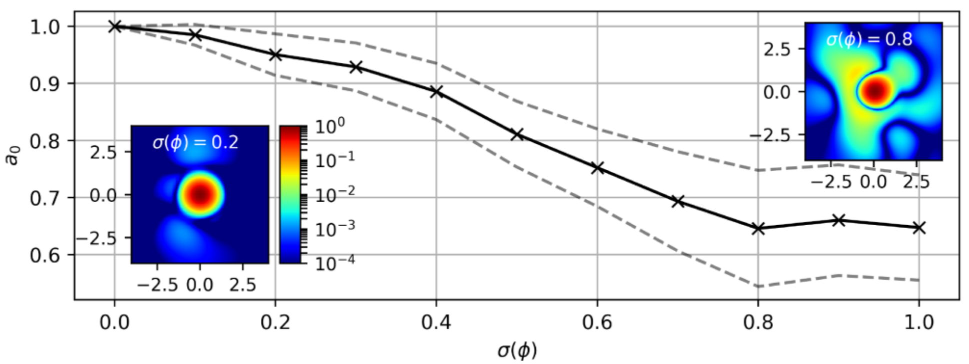
Potential solution by XAI:

- For input vectors highlight the gradients that are optimal for changing output parameters. This can help to see the indicative features and also understand where and why these features becomes inefficient in certain cases (sub-manifolds where the inverse problem is ill-posed).

Problem 3: problem statement

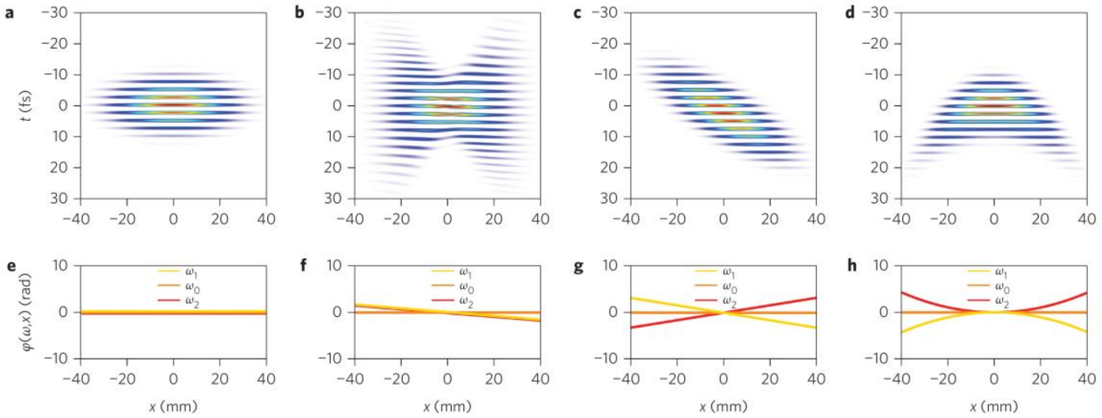


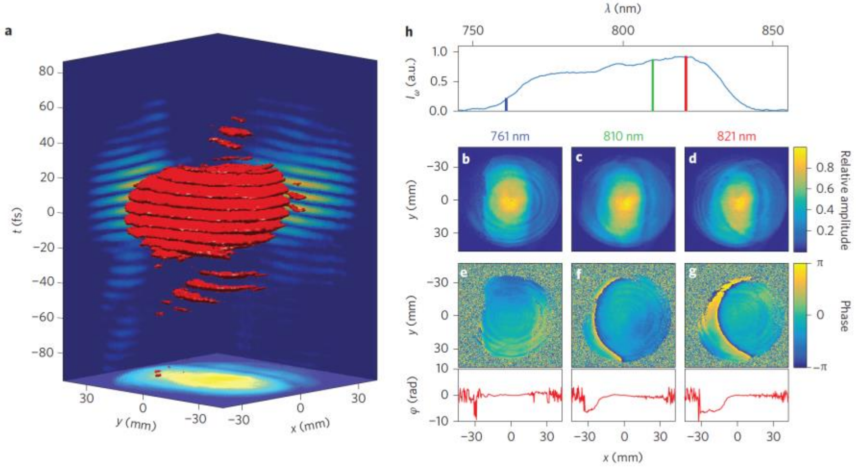
Phase deviations decrease peak field amplitude and affect its structure:



E. Panova *et al.* Appl. Sci., 11, 956 (2021)

Broad spectrum (spans over $\omega_1 < \omega_0 < \omega_2$) causes further complications:





G. Pariente *et al.* Nature Photonics 10, 547–553 (2016)

To reduce the costs and overcome limitations (tight focusing!) we intend to employ ML:

- infer angles of phase tills (for three frequencies) and their orientations from intensity distribution at the focus
- suggest optimal/automated tuning of adaptive optics

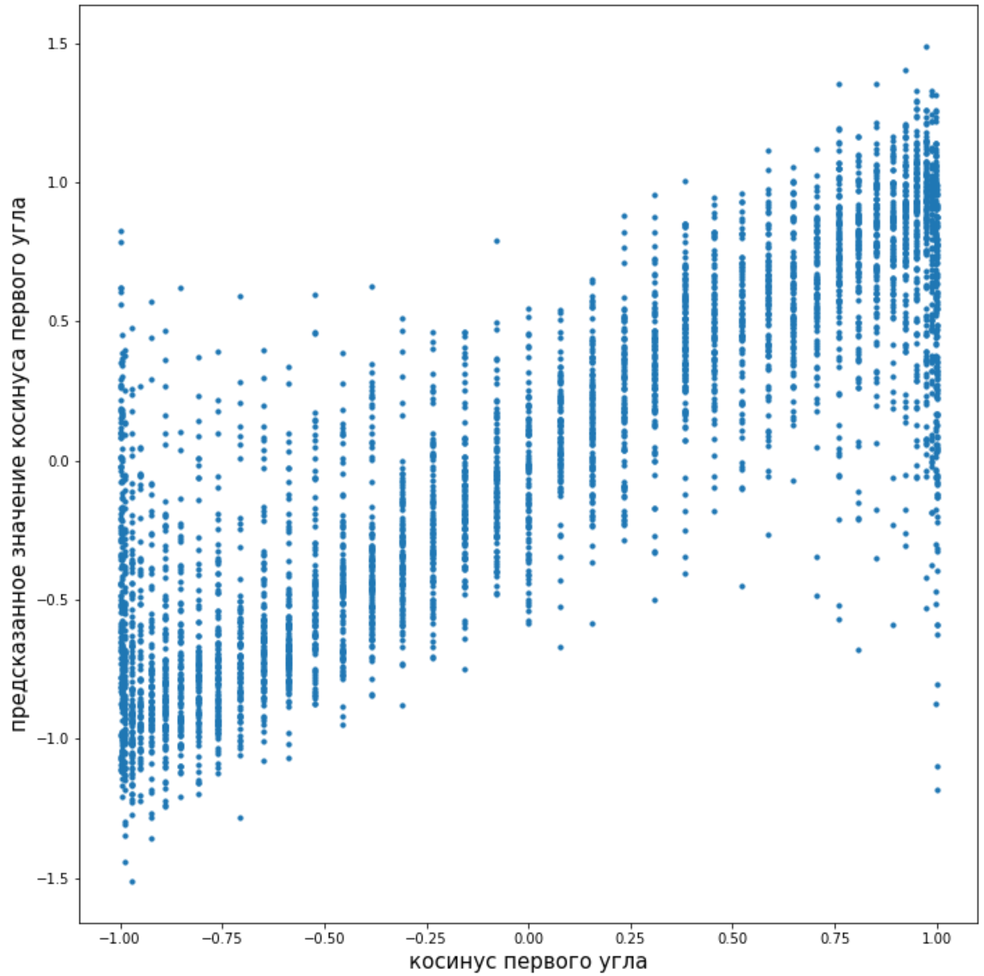
Problem 3: results

Для решения задачи была выбрана классическая модель компьютерного зрения VGG-подобная. Особенностью этой архитектуры является использование блоков из двух подряд сверточных слоев и последующей операции пулинга, после нескольких таких блоков используется несколько полносвязных слоев. В данной работе архитектура состояла из трех блоков со сверточными слоями и пулингом. Размер сверток был равен трем. Количество сверток сверток на первом блоке было равно 64, на втором 128, на последнем 256. После, идут три полносвязных слоя с количеством нейронов 128, 64, 32, функция активации ReLU.

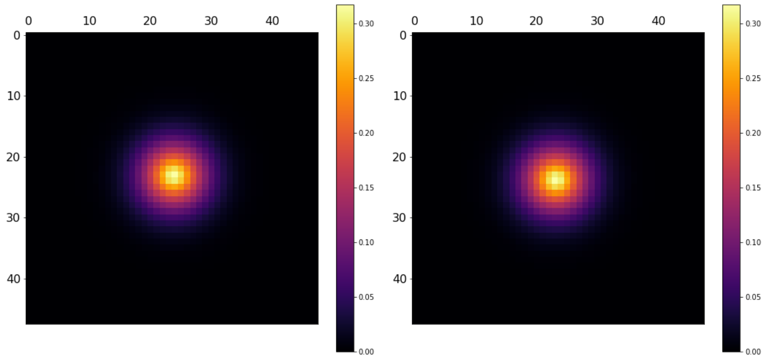
Метрика	наклон 1	наклон 2	наклон 3	Косинус угла 1	Косинус угла 3	Косинус угла 3	Интенсивность 2	Интенсивность 3
Средняя относительная ошибка	11.316	13.699	10.203	12.464	12.518	12.072	11.441	10.411
Коэффициент детерминации	0.919	0.861	0.921	0.736	0.677	0.715	0.946	0.955

Problem 3: results

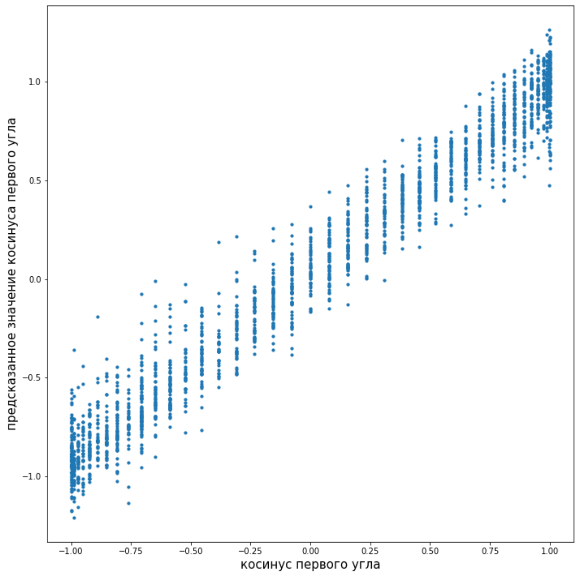
Более детальное рассмотрение указывает на появление необратимости при некоторых значениях параметров:



$\cos\phi_1 = -1$ $\quad$ $\cos\phi_1 = 1$



Разбиение пространства параметров на две части (по знаку угла наклона) устраняет проблему:



Summary, intentions and open questions

Problem 1 (laser-particle collider):

- Tolerance of FCNN and CNN to noise suggests that the noise can be used to enhance generalizability.
- What would mean the application of non-linear PCA for this problem?

Problem 2 (laser-matter converter):

- Use of PCA, pruning and data pre-processing can increase accuracy, but not dramatically.
- We intend to identify whether the FCNN pre-trained by analytical model can achieve better performance.
- We intend to use transfer learning to enforce generalizability and to mitigate transition to real experimental data.
- Are there any XAI concepts relevant to combating the difficulties with stochasticity.

Problem 3 (tight focusing):

- One can subdivide parameter space for better learning and use ML to identify the specialization region.
  Can one automate this process?
- One can use $(\psi_{tilt} \cos\phi, \psi_{tilt} \sin\phi)$ instead of (ψ_{tilt}, ϕ) or corresponding loss function. Can one automate the search of such replacements?
- We intend to identify features by visualizing the direction of maximized change of an output parameter (via backward gradient propagation). Is it sensible? Are there better ways?
- Can one apply non-linear PCA to the accuracy levels in multidimensional space of parameters in order to explain/visualize, where the problem is ill-posed?