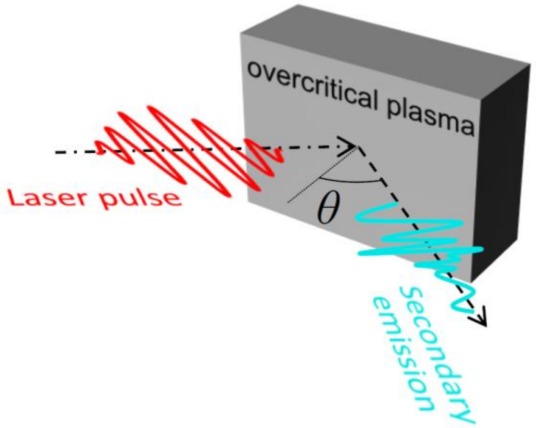
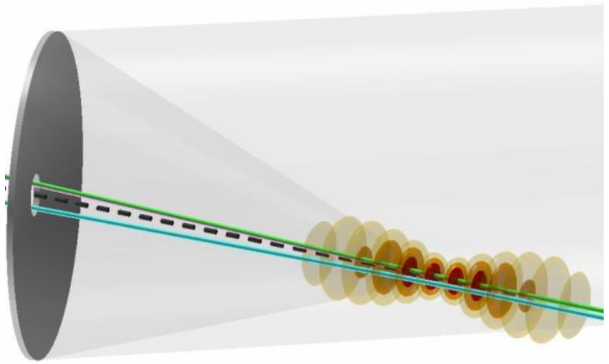
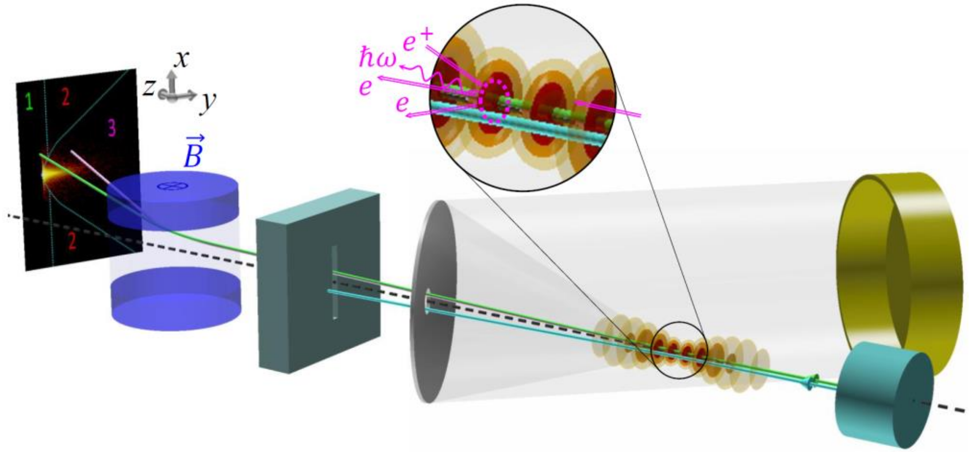
LOBACHEVSKY UNIVERSITY

UNIVERSITY OF GOTHENBURG

Institute of Applied Physics
Russian Academy of Sciences







Towards simulation-governed ML-based analysis of laser-plasma interactions

Yury Rodimkov, Valentin Volokitin, Elena Panova,
Iosif Meyerov *Lobachevsky University (Russia)*

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

Evgeny Efimenko
IAP RAS (Russia)

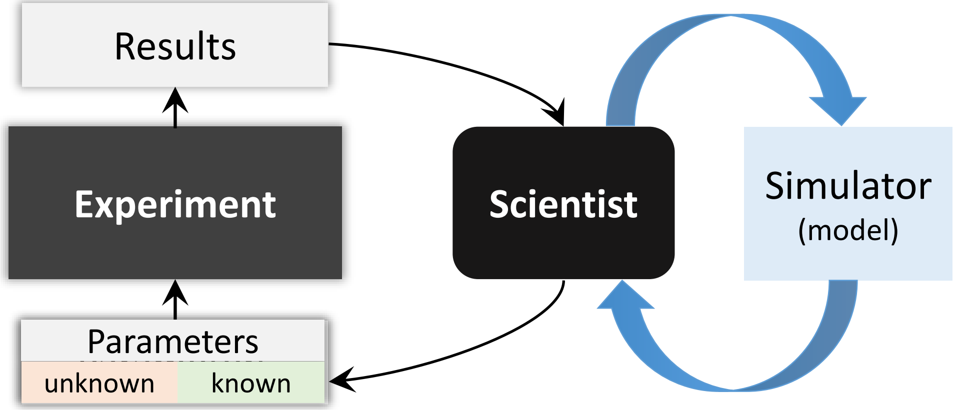
Alexey Polovinkin
Adv Stat & Machine Learning, LTD, Intel (USA)

László Veisz
Umeå University (Sweden)

2022-01-26

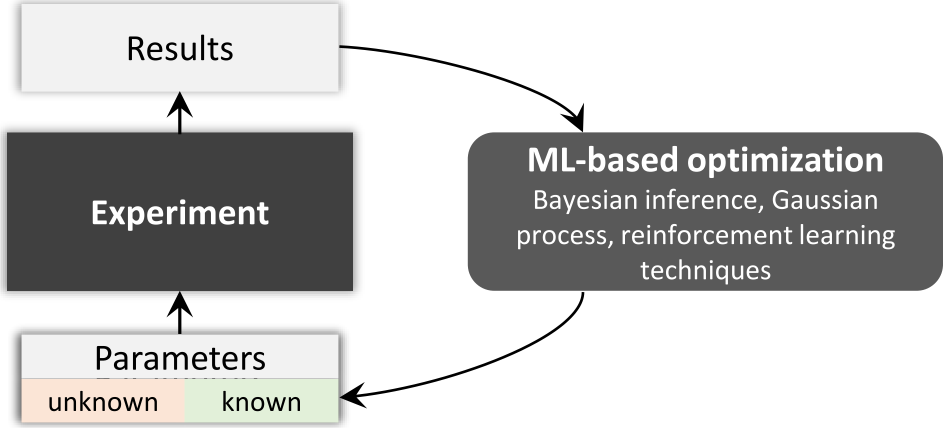
What if experiment and theory do not agree?

Simulation-based strategy



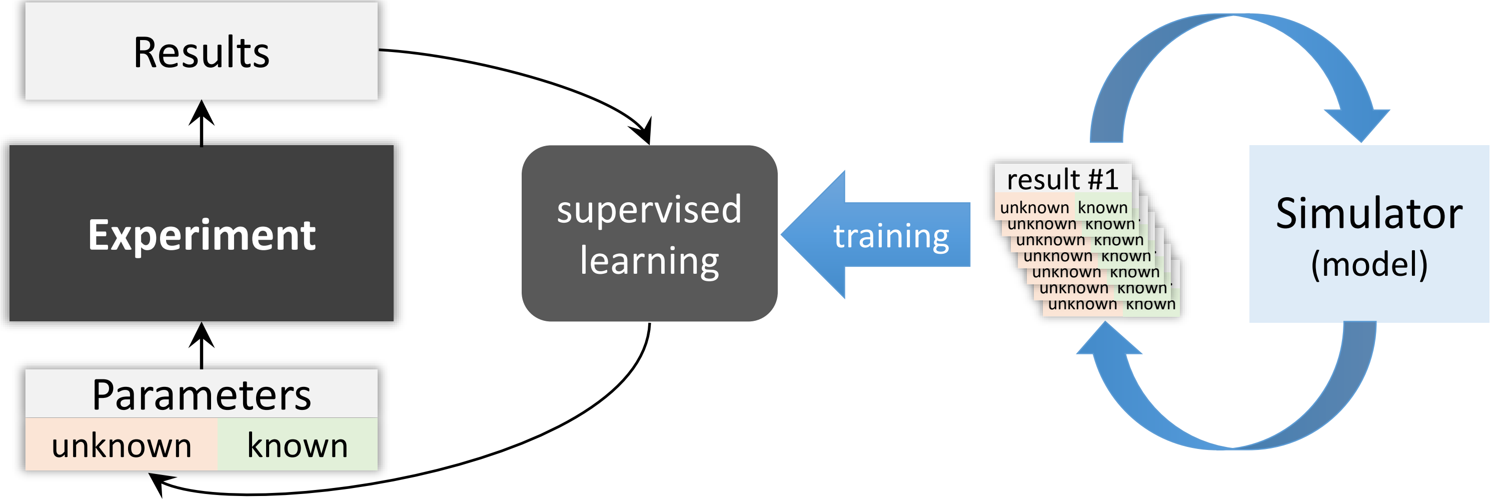
… extend model and/or diagnostics

ML-based strategy



… do optimization for tunable parameters

ML-based diagnostics:



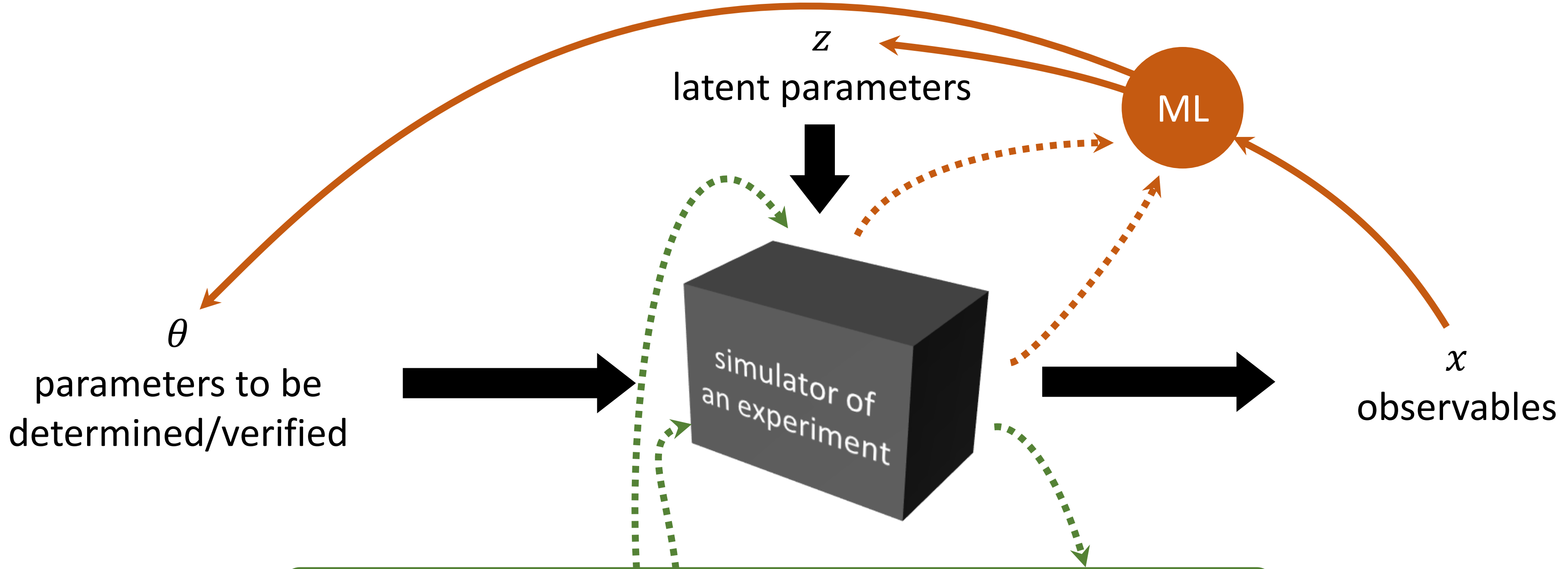
Problems:

- simulated results can differ
- and in an unknown way!

Methods:

- develop tolerance to differences
- extend simulator
- transfer learning

The layout of an inverse problem with latent parameters



ABC: for each x_e: $(z_i, \theta_i\ ) \sim \pi(z, \theta)$, simulate $x_i(z_i, \theta_i\ )$, summary statistics $s(x)$, accept if $K\big(s(x_i), s(x_e)\big) < h$.

Problem: Infer θ from experimental data

Approaches and difficulties:

1. Approximate Bayesian Computation (ABC): large dimensionality of x and z make the likelihood function intractable (requires integration over all possible outcomes)
2. ML: explanation and reliability; irreversibility due to probabilistic or/and stochastic nature of the process; difference between experiment and simulation

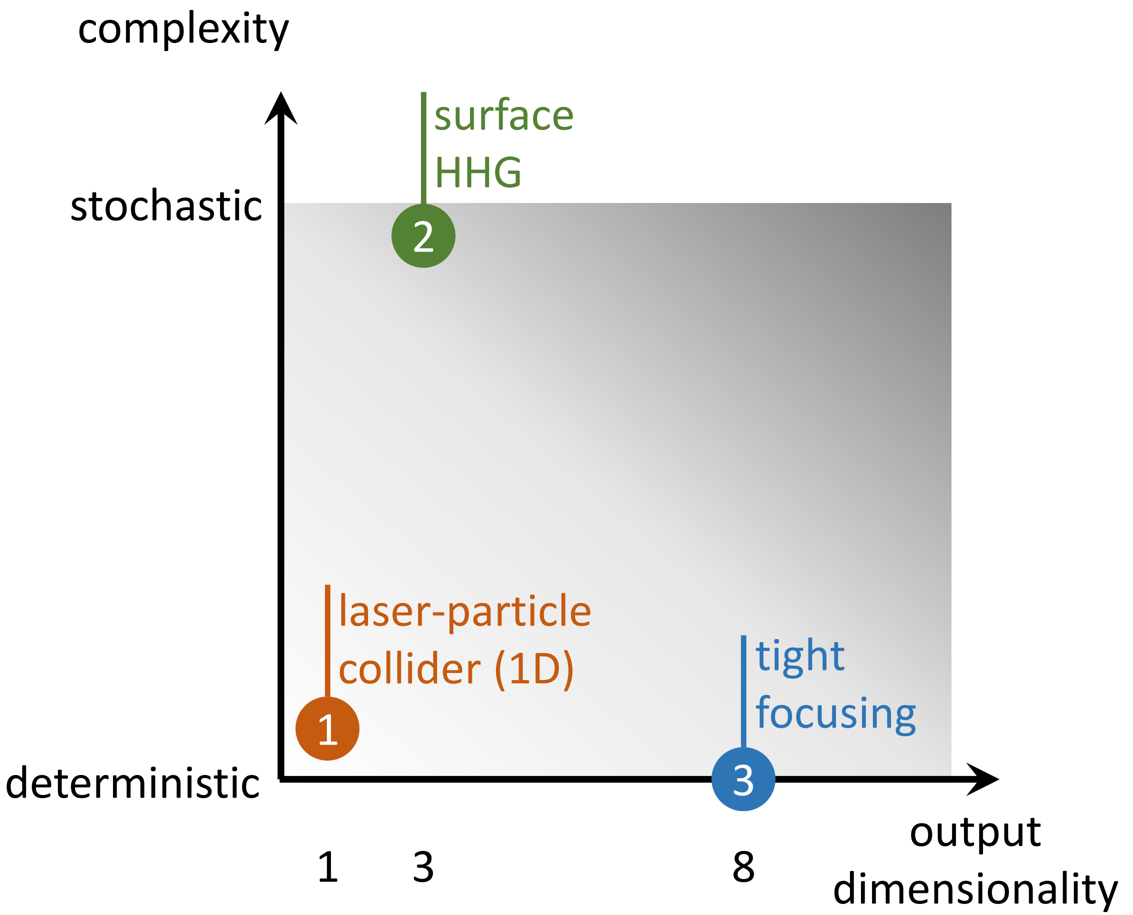
The overview of the activity in terms of ML

Incremental improvements:

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

Game-changing improvements:

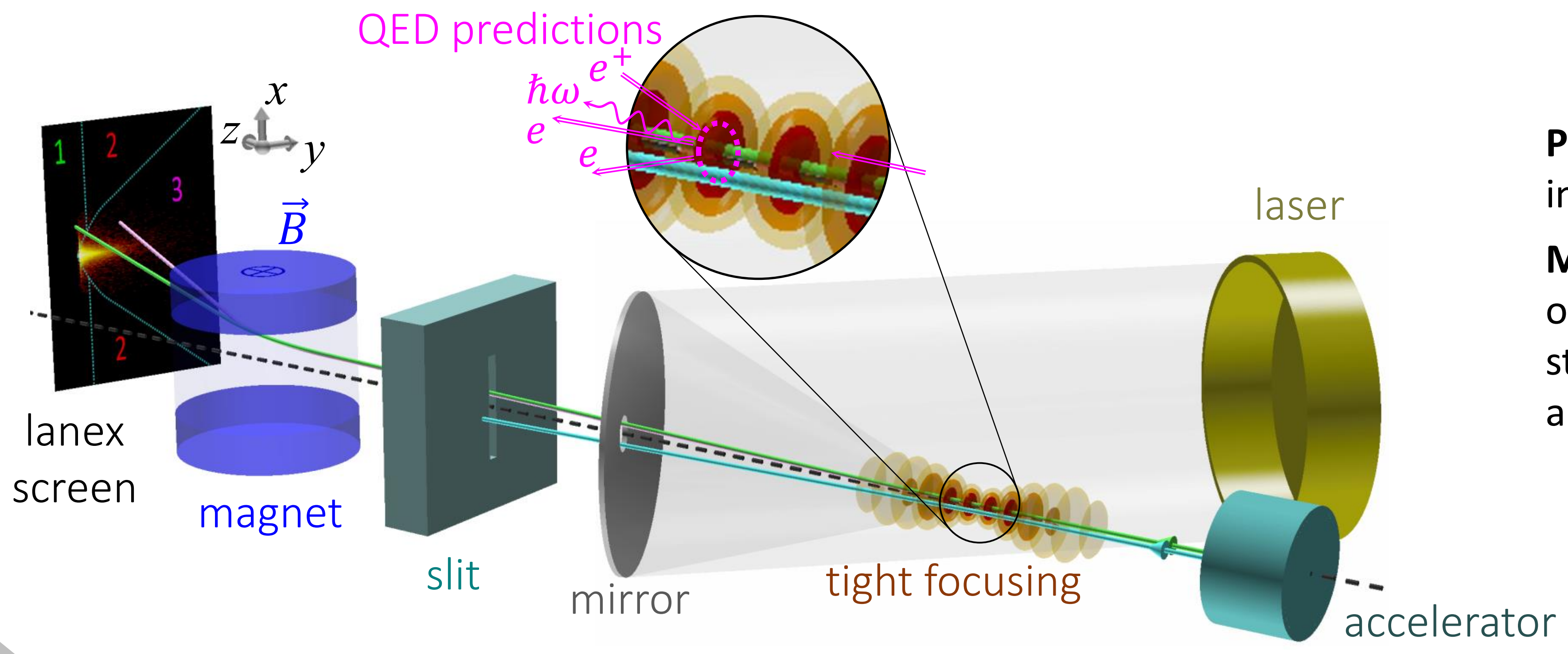
- Generalizability (simulations → experiment)
- Overcoming irreversibility (detect, explain)
- Reliability (retrieve sufficient summary statistics; identify indicative features)



Methodology:

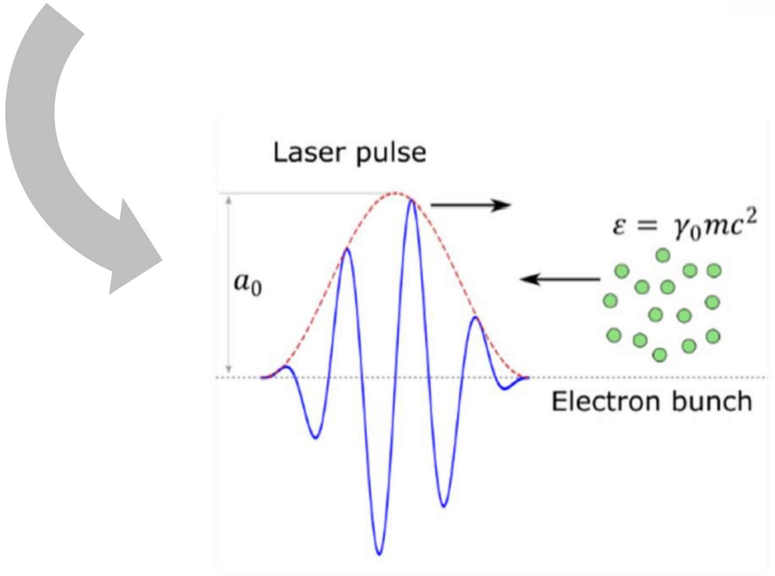
1. Identify ML models tolerant to noise (varied by binning strategy): use noise to enhance generalizability.
2. Transfer learning: (1) pre-train using simplified analytical models (uncostly data) to accentuate indicative features; (2) generalize using ab-initio simulations (cheap data); (3) fine-tune using actual experiment (expensive data).
3. Improve ML model invariance by using simulation-based generative (composabale) model for training.

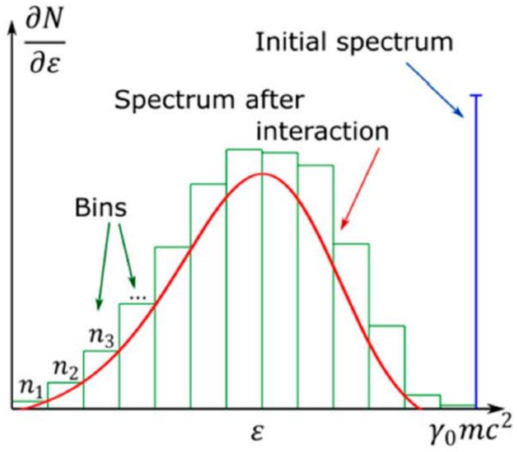
Problem #1: peak field determination in experiments on SFQED

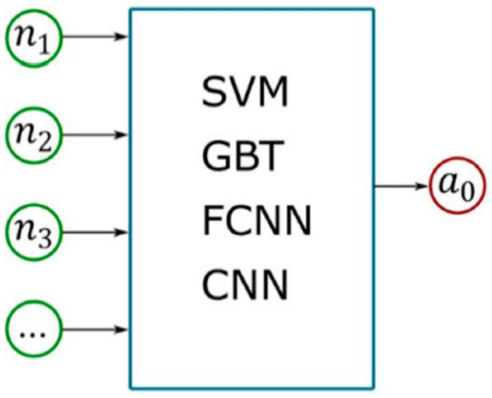


Problem: determine peak laser intensity achieved.

Motivation: experimental tests of strong-field QED require strong fields of known amplitude.



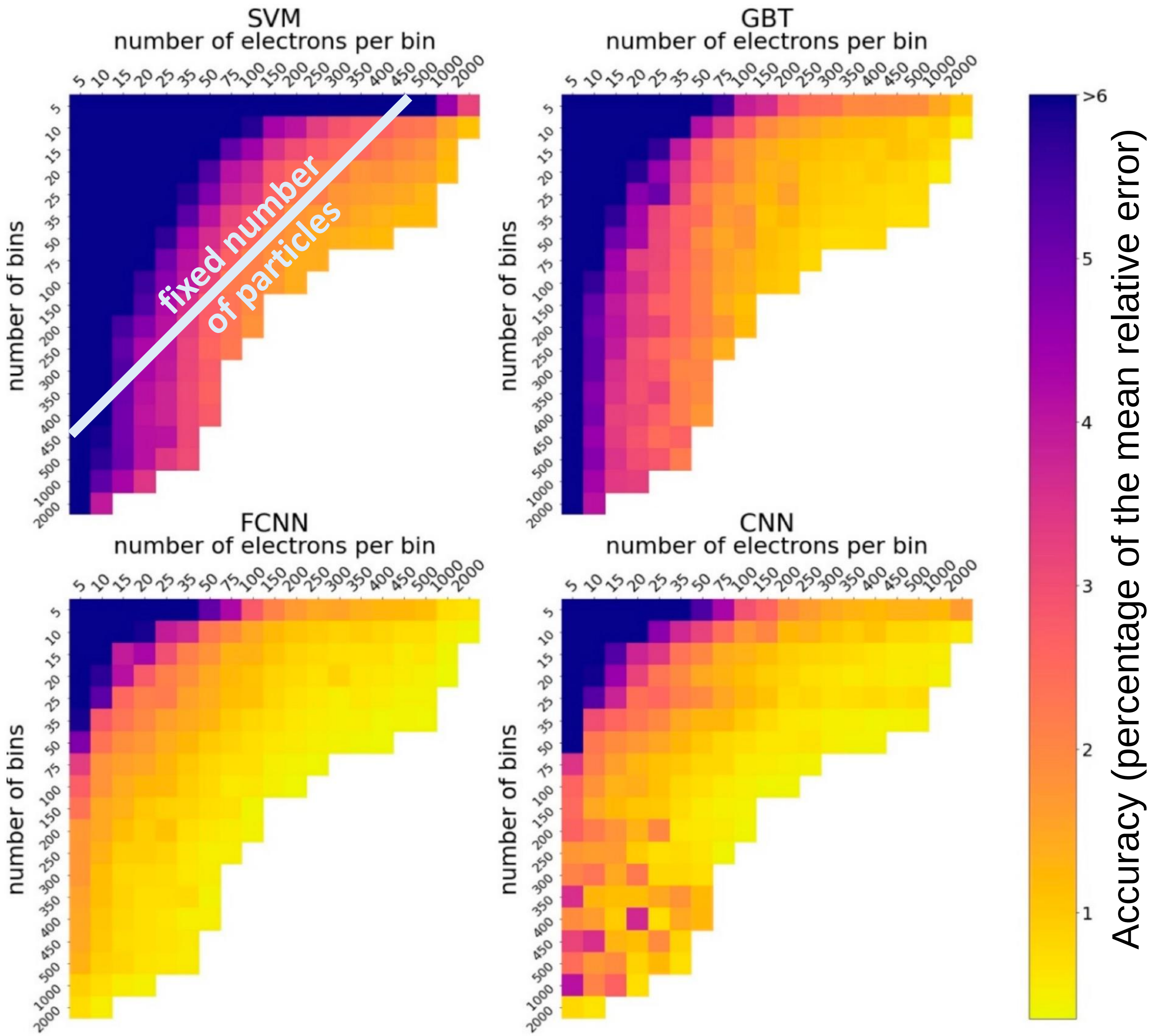




Problem 1: results

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.

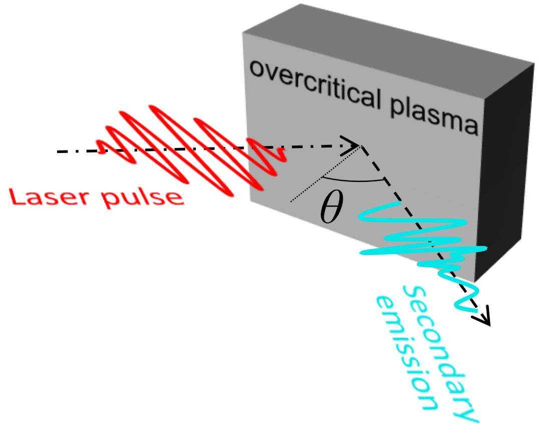


Y. Rodimkov et al. ML-Based Analysis of Particle Distributions in High-Intensity Laser Experiments: Role of Binning Strategy, Entropy, 23 (1), 21 (2020)

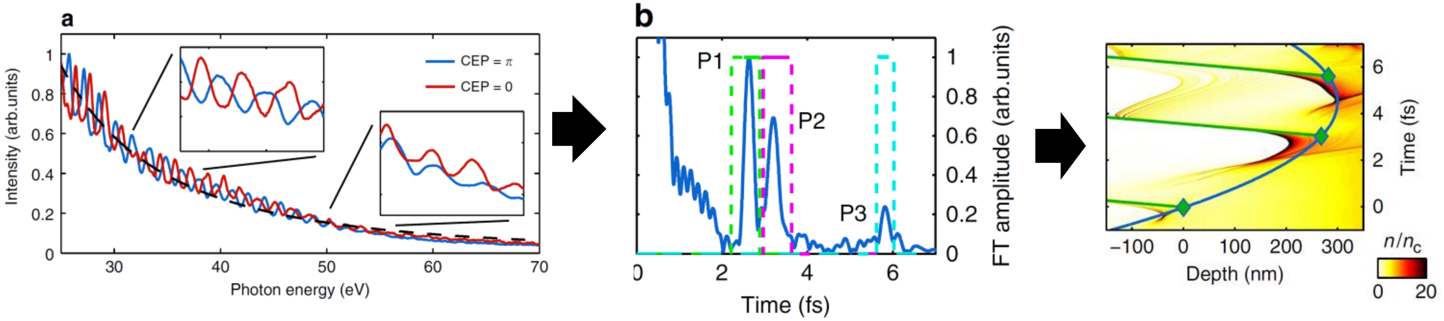
Problem #2: problem statement

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).



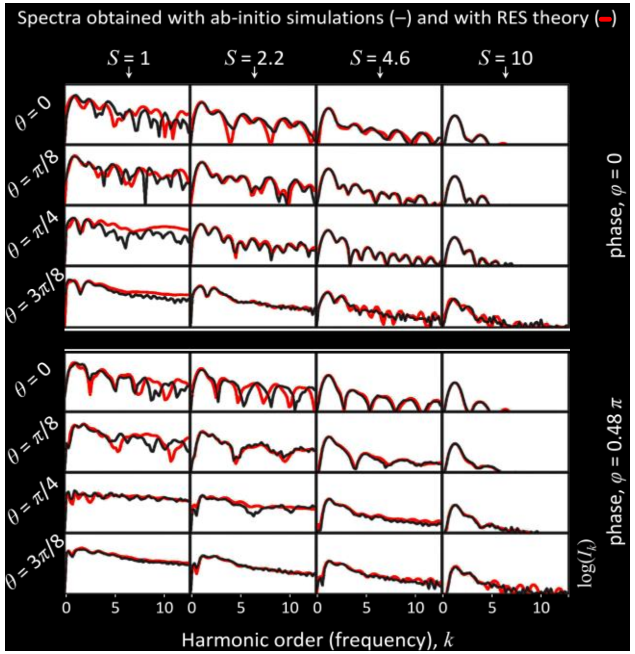
Spectral interferometry (designed for 2-3 cycle pulses, known parameters):



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 (~ 1 min, ~ 10^5 cases), experiment (~ 10^3 cases)
- determine (highlight?) indicative features

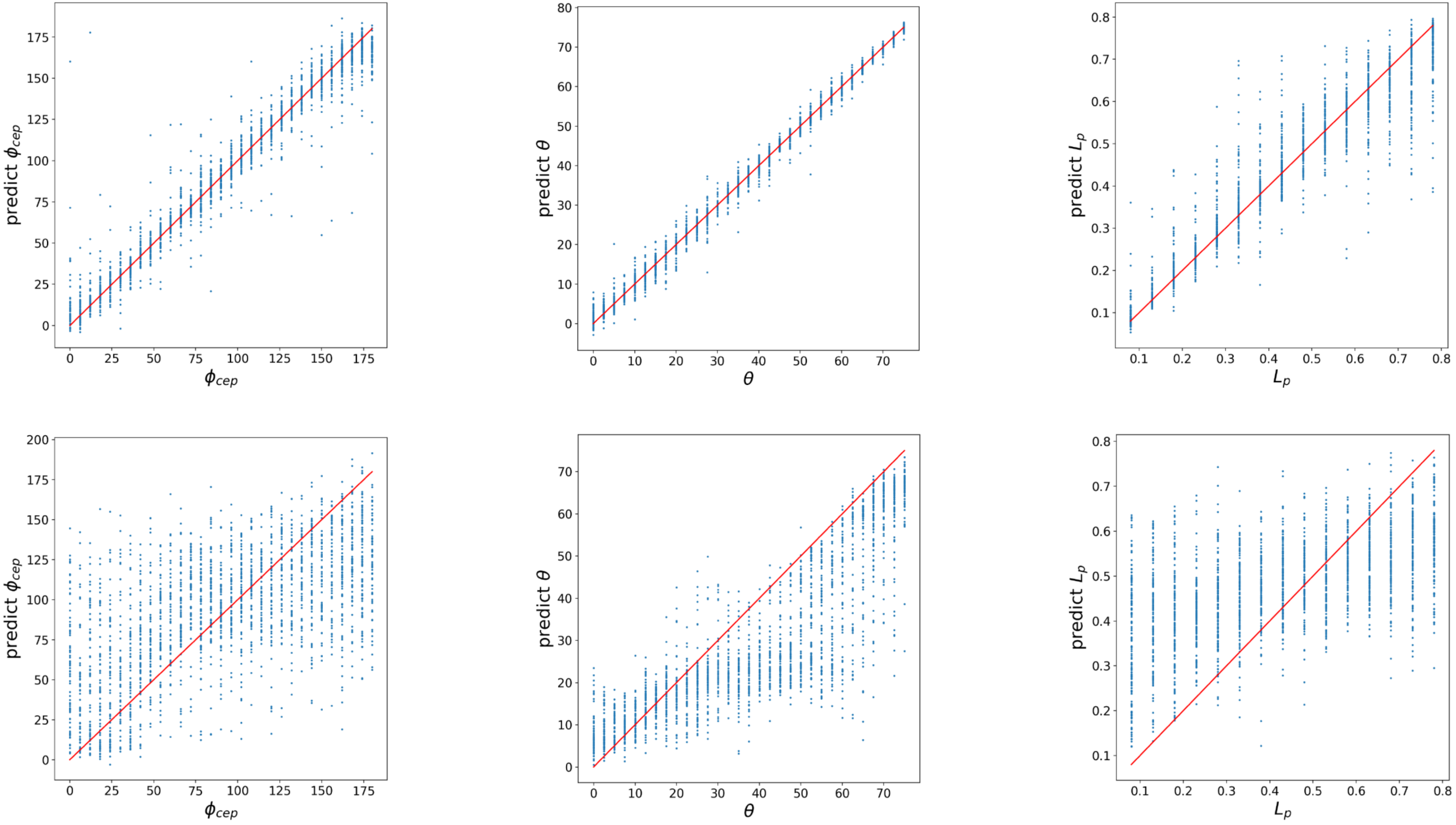


Gonoskov et al., Sci. Rep. 9 (1), 1-15 (2019)
RES model: Gonoskov, Phys. Plasmas (2018)

Problem #2: results, FCNN trained with PIC simulations

clean data

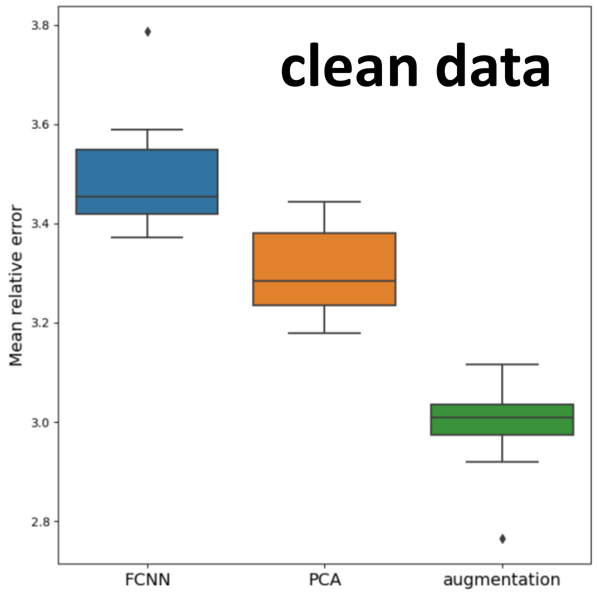
“noisy” data

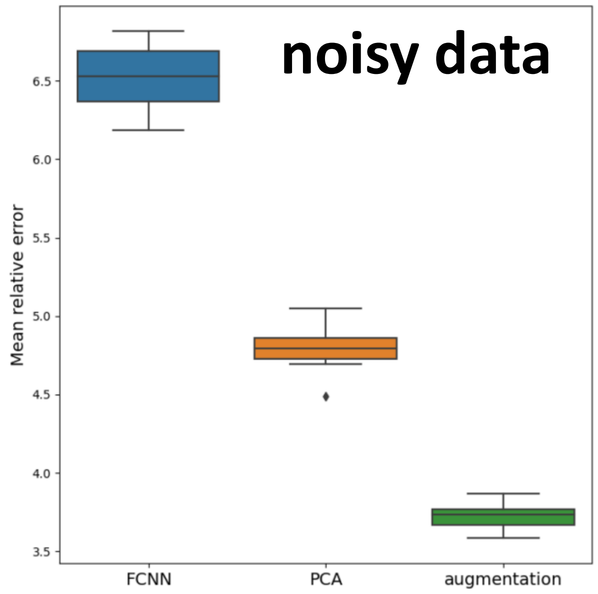


Problem #2: tolerance to noise

Accuracy measures: $MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|\widehat{y_i} - y_i|}{\max(\widehat{y})}$, $R^2 = 1 - \sum_{i=1}^{n}\frac{(\widehat{y_i} - y_i)^2}{(y_i - \overline{y})^2}$

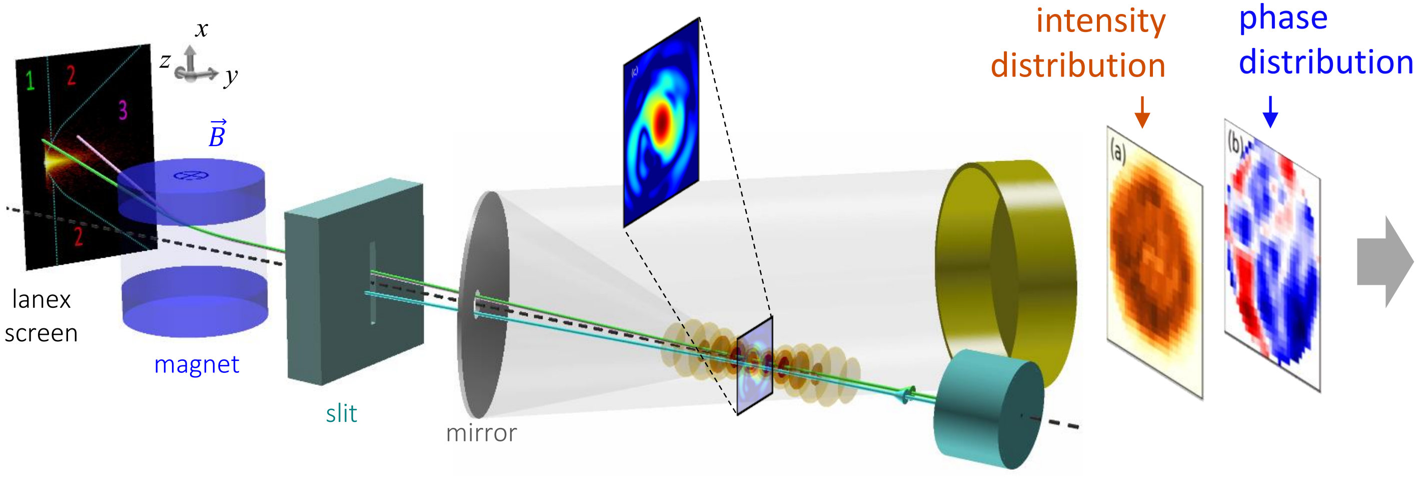
Model	Metrics	Φ_{CEP}	θ	L_p
Baseline model trained on clean data, tested on clean data	MAPE	3.823	1.781	4.890
	R^2	0.952	0.991	0.904
Baseline model trained on clean data, tested on noisy data	MAPE	18.675	13.297	18.601
	R^2	0.327	0.638	0.304
Baseline model trained on noisy data, tested on noisy data	MAPE	7.638	3.179	8.743
	R^2	0.841	0.968	0.756
PCA preprocessing	MAPE	5.385	2.434	6.537
	R^2	0.944	0.988	0.893
Adding noise to data gradually	MAPE	4.170	1.911	5.095
	R^2	0.946	0.988	0.897



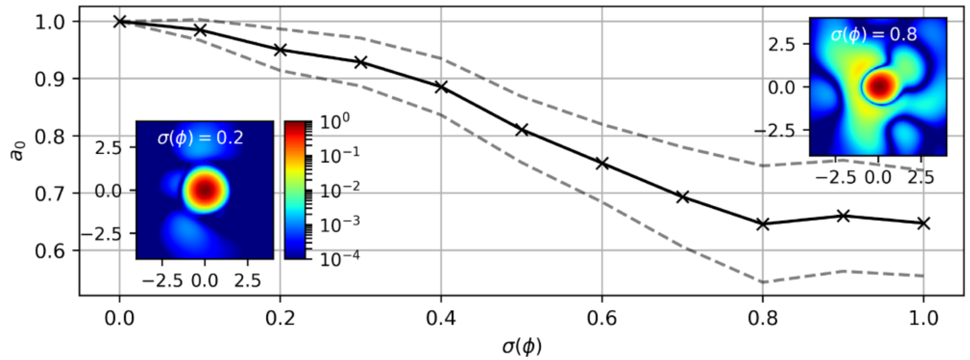


Y. Rodimkov et al. Sensors 21 (21), 6982 (2021)

Problem #3: problem statement

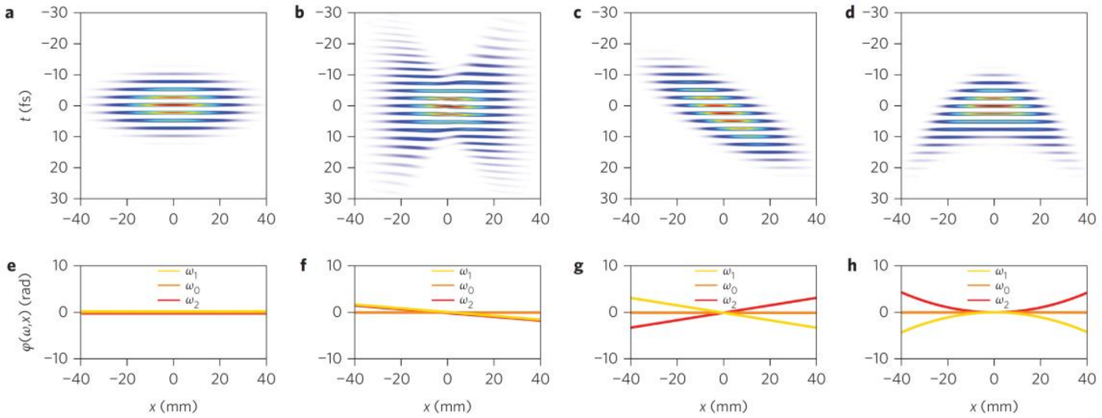


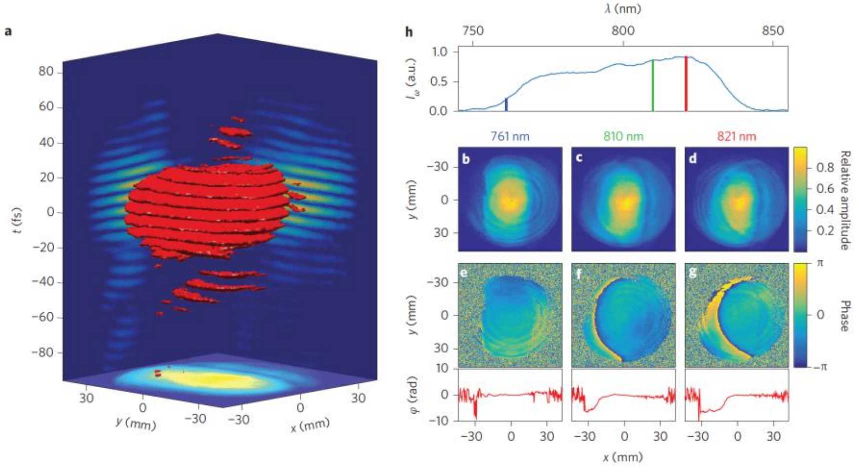
Phase deviations decrease peak field amplitude and affect its structure:



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 of short pulses) we try to use 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