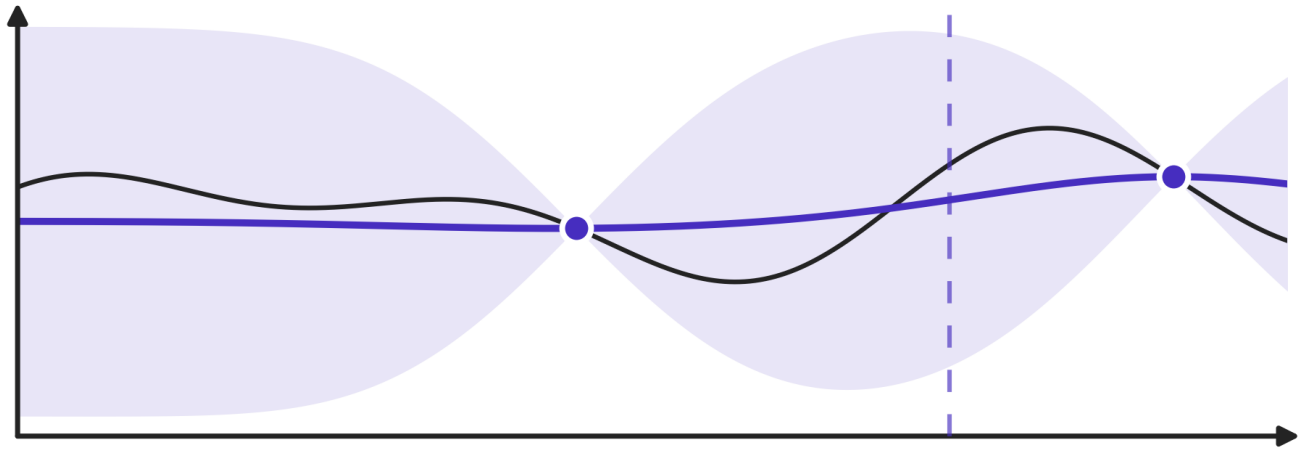
Bayesian Optimization of Attosecond Pulse Generation with Intense Lasers

In numerical simulations



Dominika Maslarova

Nuclear Physics under the Low-Energy, High-Intensity Frontier
ECT*, Villazzano • Aug 10 - 14, 2026

Plasma Theory Group

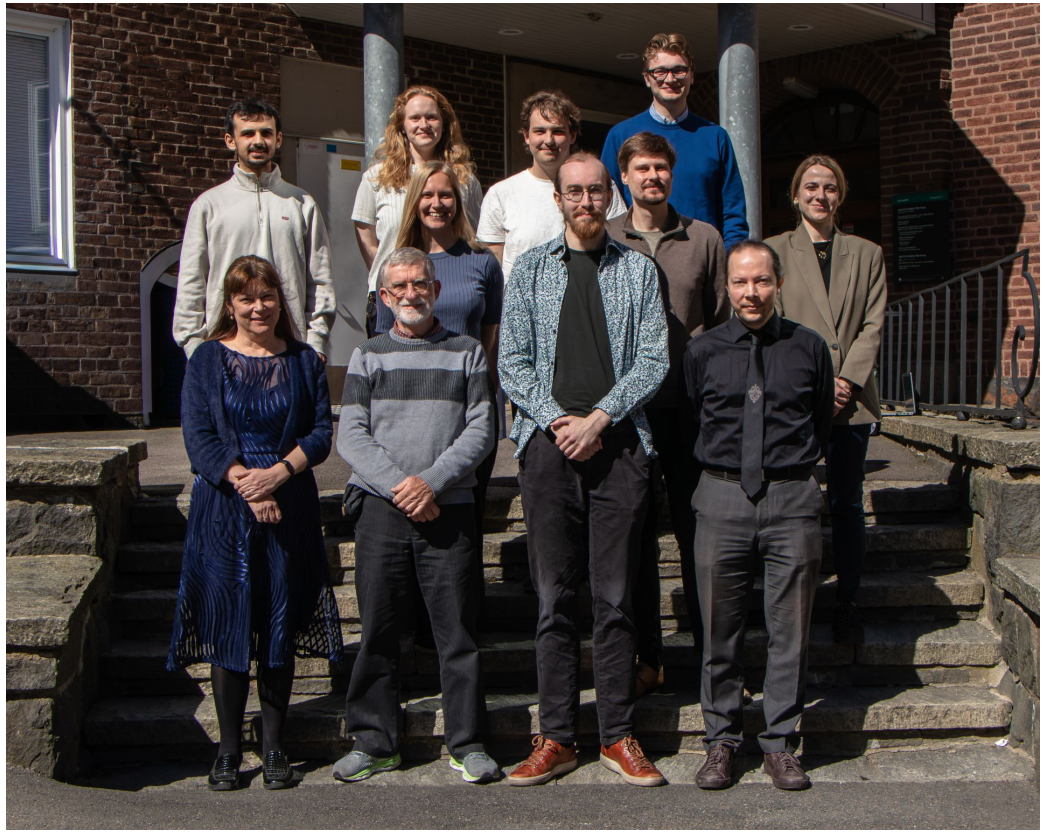
CHALMERS

Department of Physics and Astronomy
Division of Subatomic, High Energy and Plasma Physics

- 1 professor
  Tünde Fülöp
- 1 associate professor
- 4 postdocs
- 3 PhD students

Topics

- Fusion plasma
- Space plasma
- Laser plasma





What is Bayesian optimization?

PRIOR ASSUMPTION ABOUT DATA (FUNCTION)

What is Bayesian optimization?

PRIOR ASSUMPTION ABOUT DATA (FUNCTION)

OBSERVATIONS



What is Bayesian optimization?

PRIOR ASSUMPTION ABOUT DATA (FUNCTION)

OBSERVATIONS



POSTERIOR MODEL OF THE FUNCTION

What is Bayesian optimization?

PRIOR ASSUMPTION ABOUT DATA (FUNCTION)

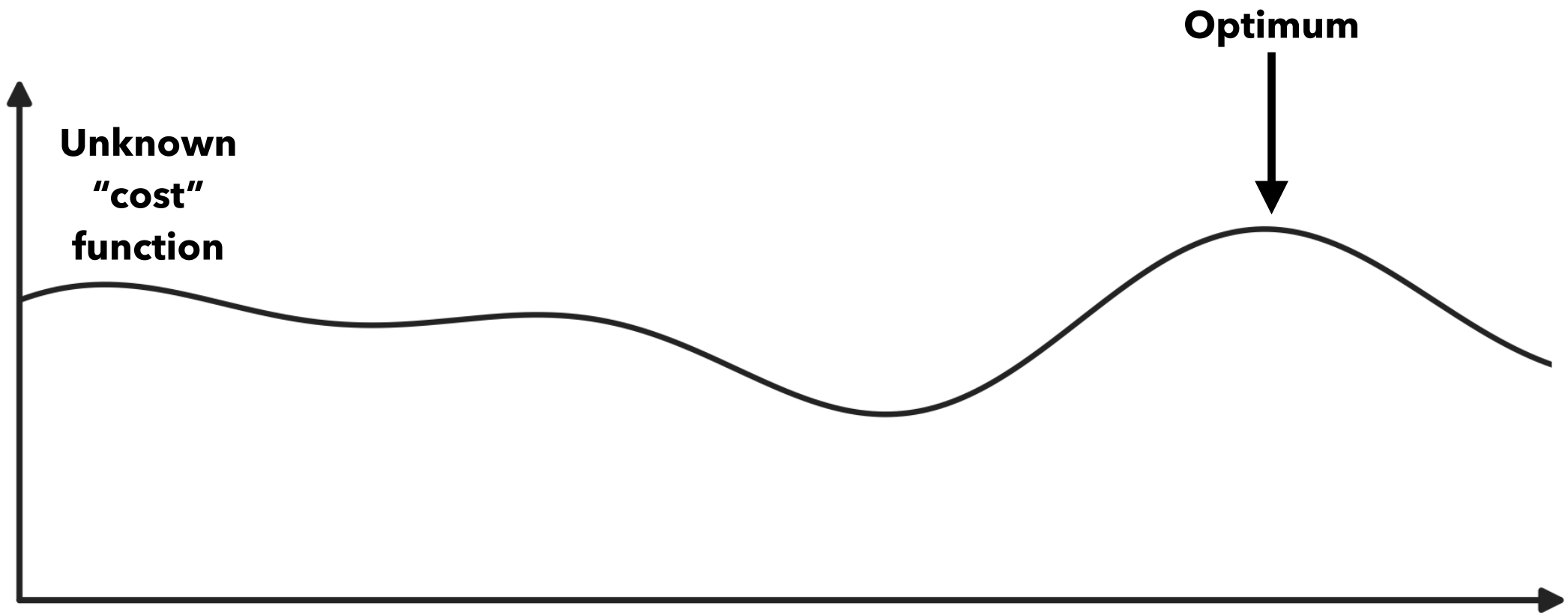
OBSERVATIONS



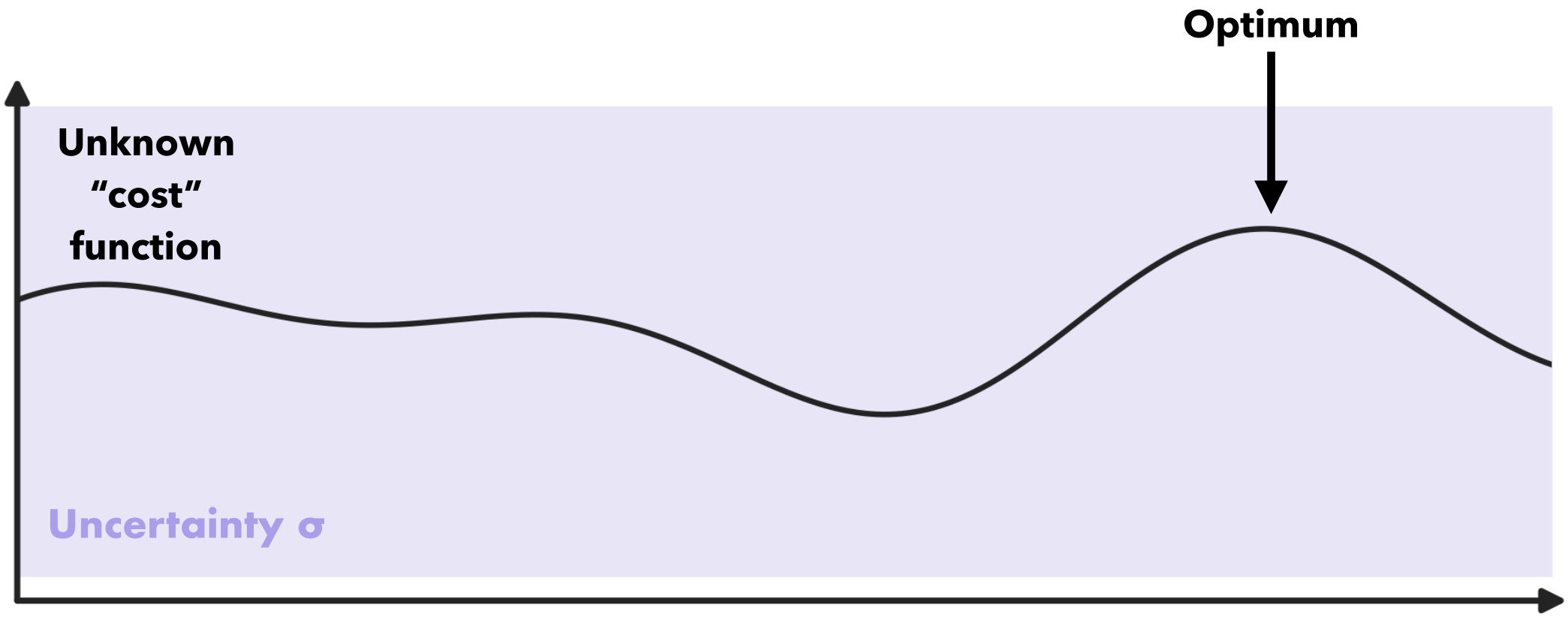
POSTERIOR MODEL OF THE FUNCTION

We look for optimum of this function

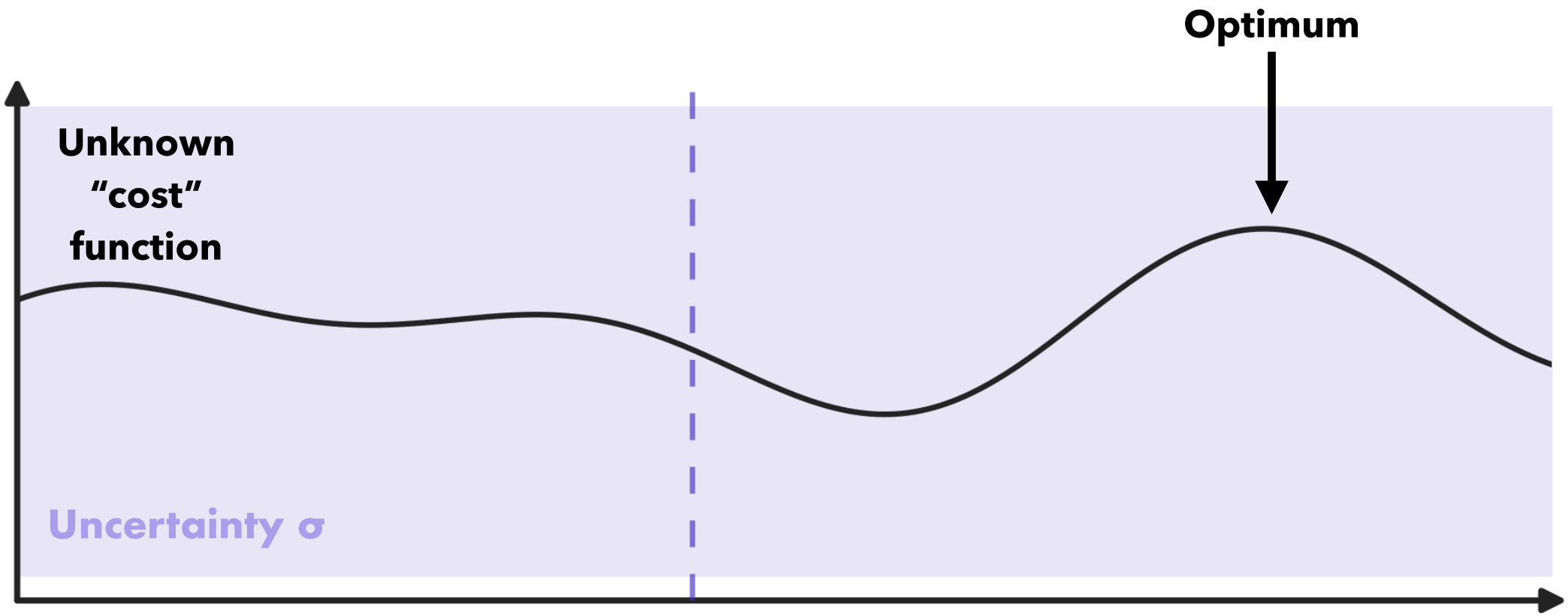
What is Bayesian optimization?



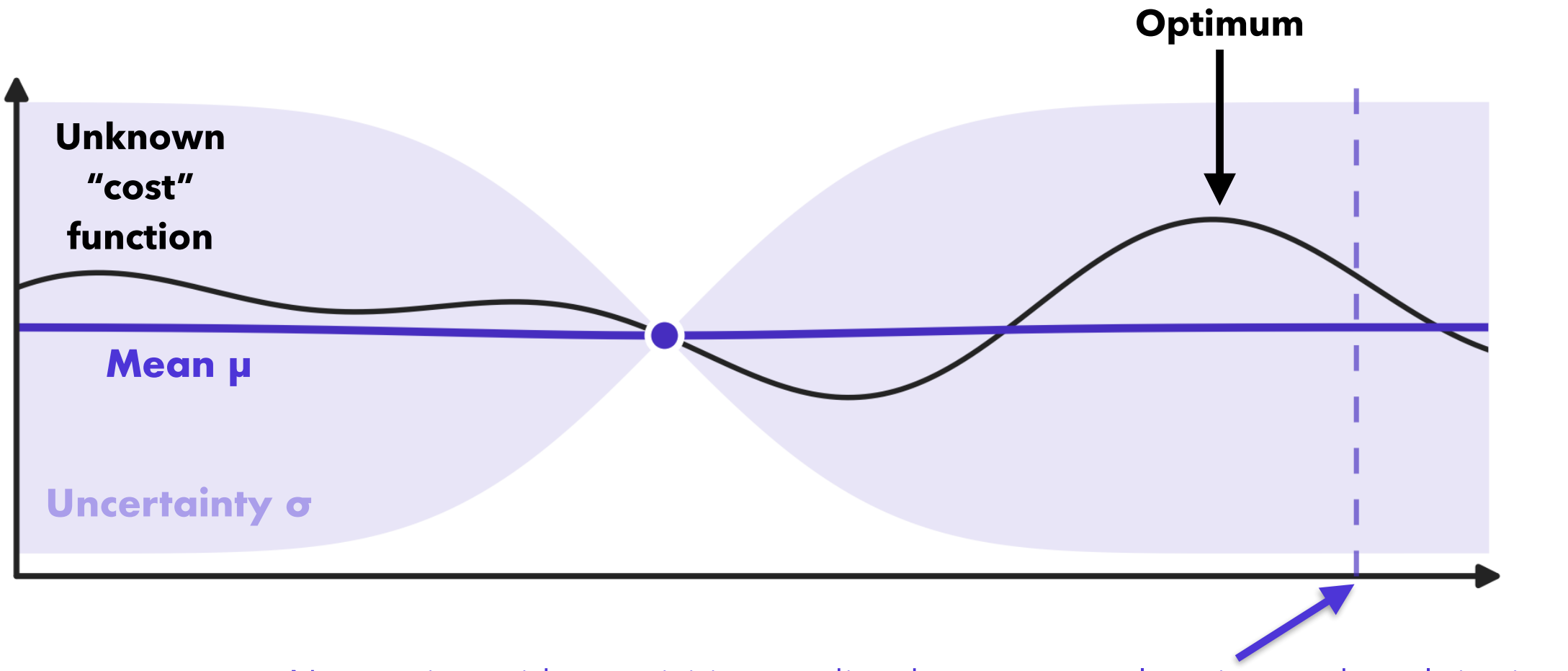
What is Bayesian optimization?



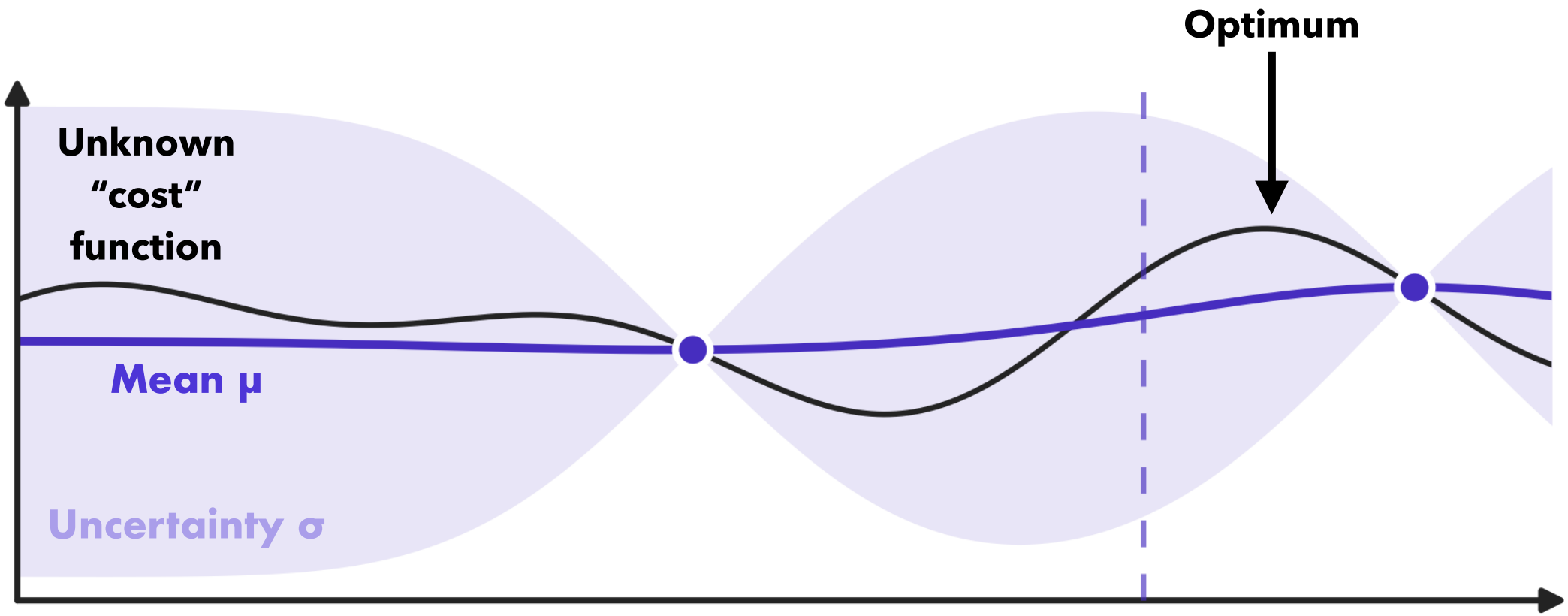
What is Bayesian optimization?



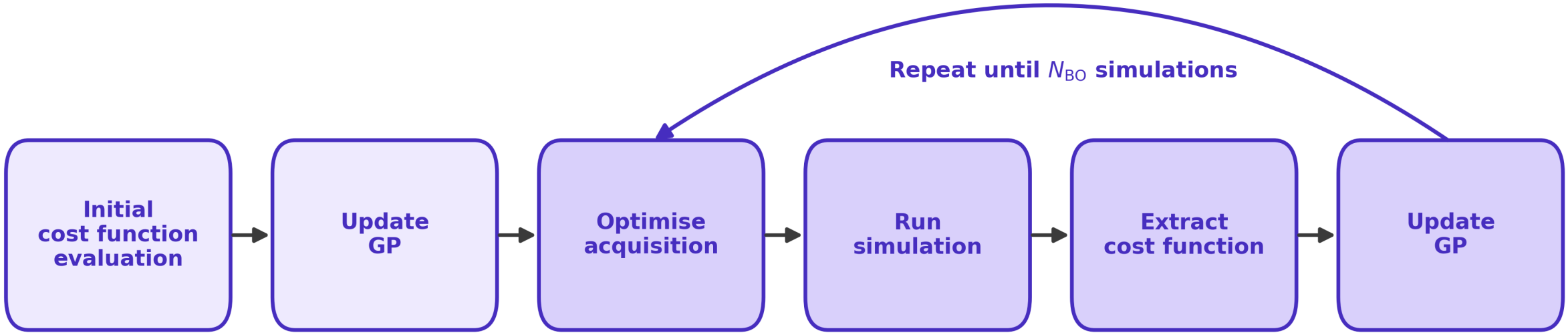
What is Bayesian optimization?



What is Bayesian optimization?



Bayesian optimization with simulations works in a loop...



I will talk about Bayesian optimization in attosecond science with lasers

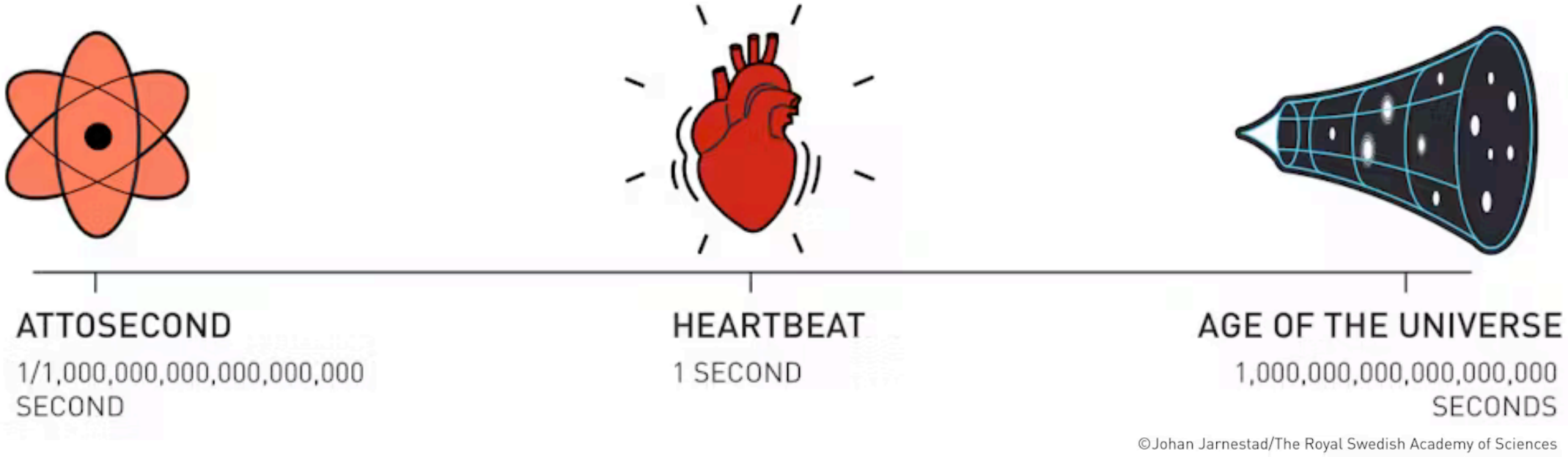
I. Introduction to attosecond pulses

II. Attosecond pulses from laser wakefield acceleration:
Batch Bayesian optimization

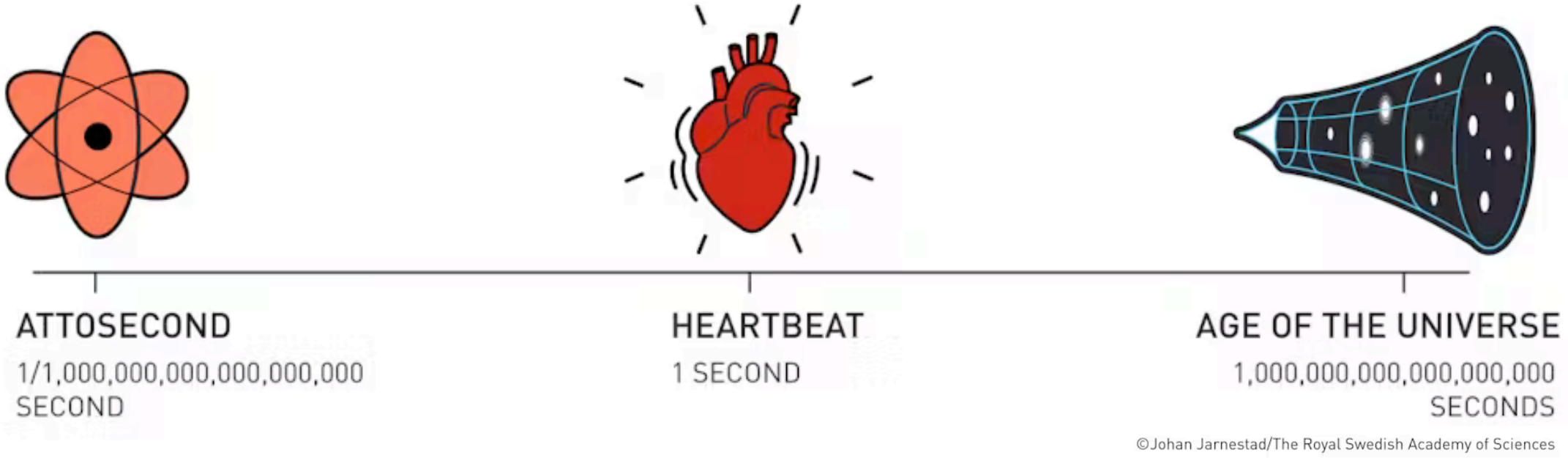
III. Attosecond pulses from high harmonic generation in gases:
Physics-informed models

INTRODUCTION TO ATTOSECOND PULSES

What is attosecond?



What is attosecond?

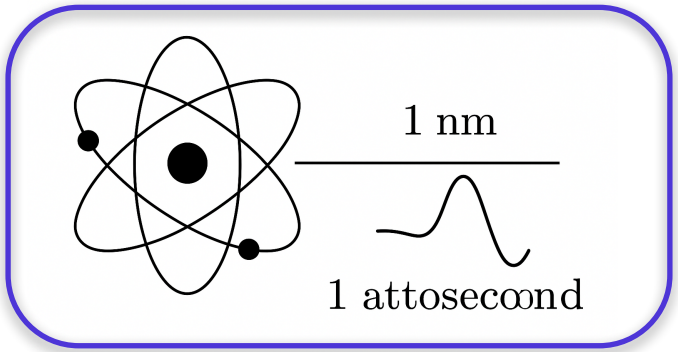


SHORTEST PULSE EVER CREATED AND MEASURED: 43 as

T. Gaumnitz et al. *Opt. Express* 25, 27506-27518 (2017)

Why attoseconds?

- Probing matter on scales of electron motion in atoms
- Timing and controlling ionization dynamics



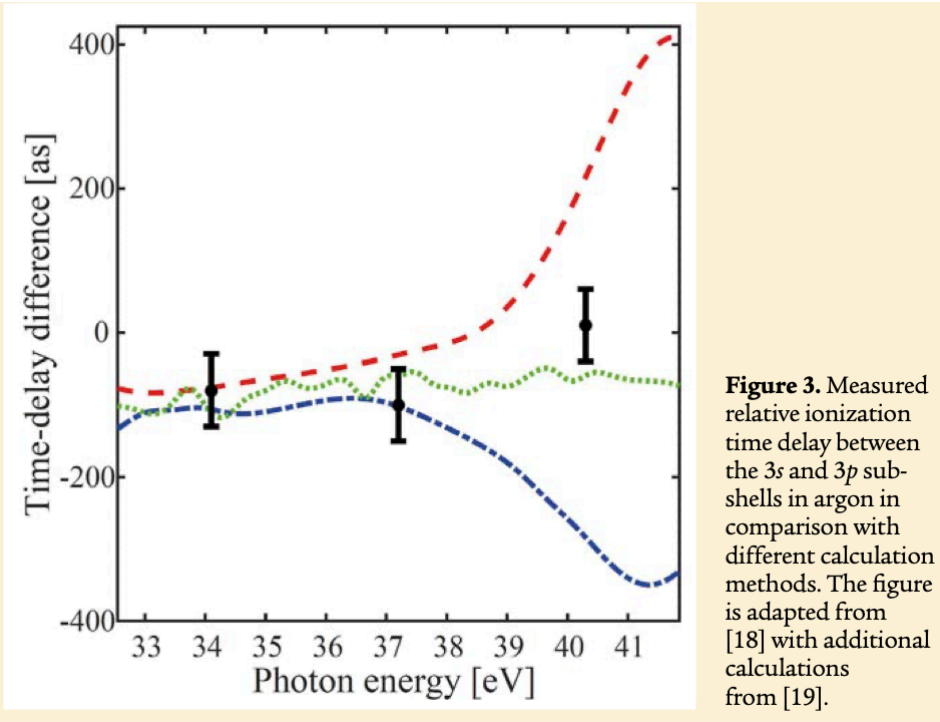
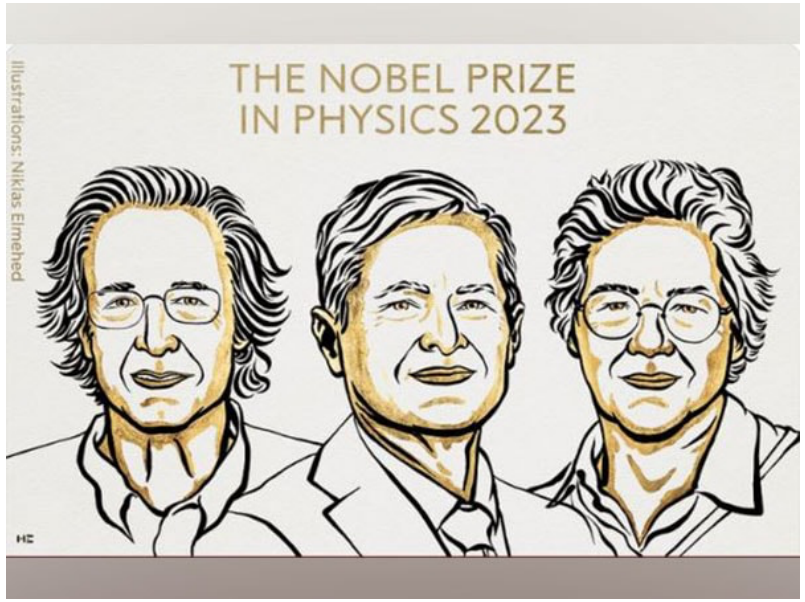


Figure 3. Measured relative ionization time delay between the 3*s* and 3*p* sub-shells in argon in comparison with different calculation methods. The figure is adapted from [18] with additional calculations from [19].

ARNOLD, C. L., et al. *Photoniques*, 2018, 28-35.



Niklas Elmehed © Nobel Prize Outreach

LASER WAKEFIELD ACCELERATION

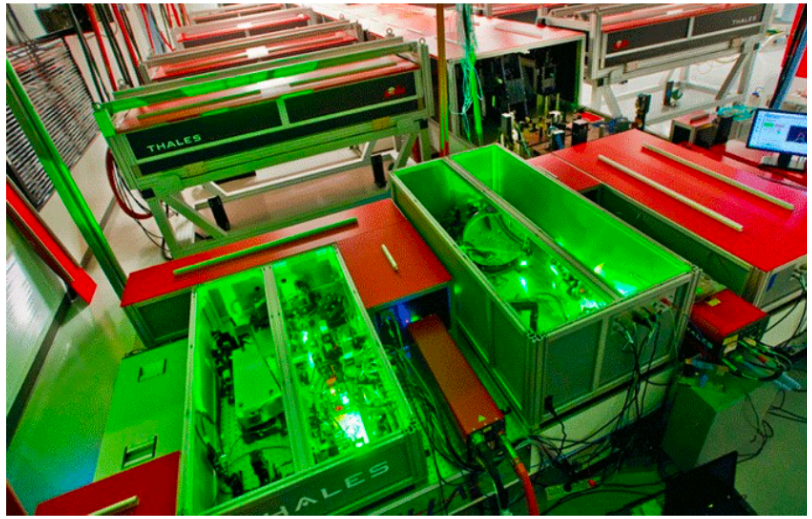
Laser wakefield acceleration accelerates electrons on very short scales

- Laser wakefield accelerator is a very **compact particle** (electron) **accelerator**
- Acceleration happens in **PLASMA** = ionized state of matter containing charged particles



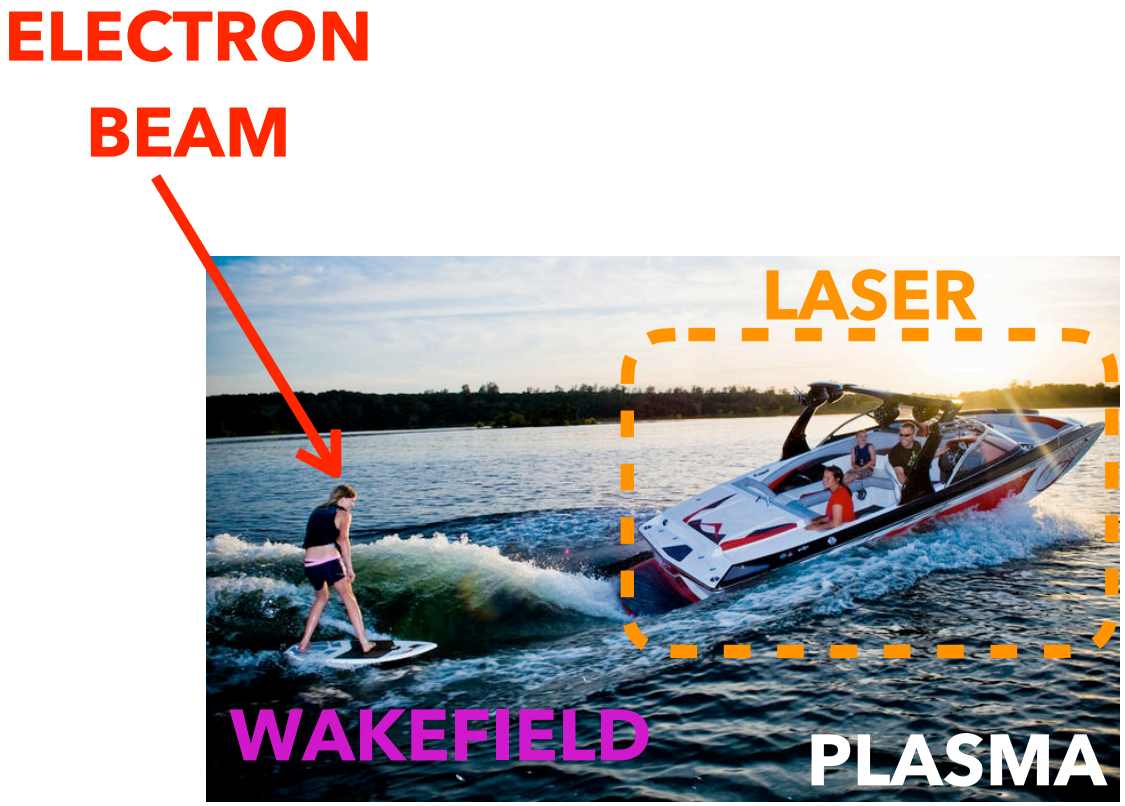
CONVENTIONAL ACCELERATOR

~1000x SMALLER!

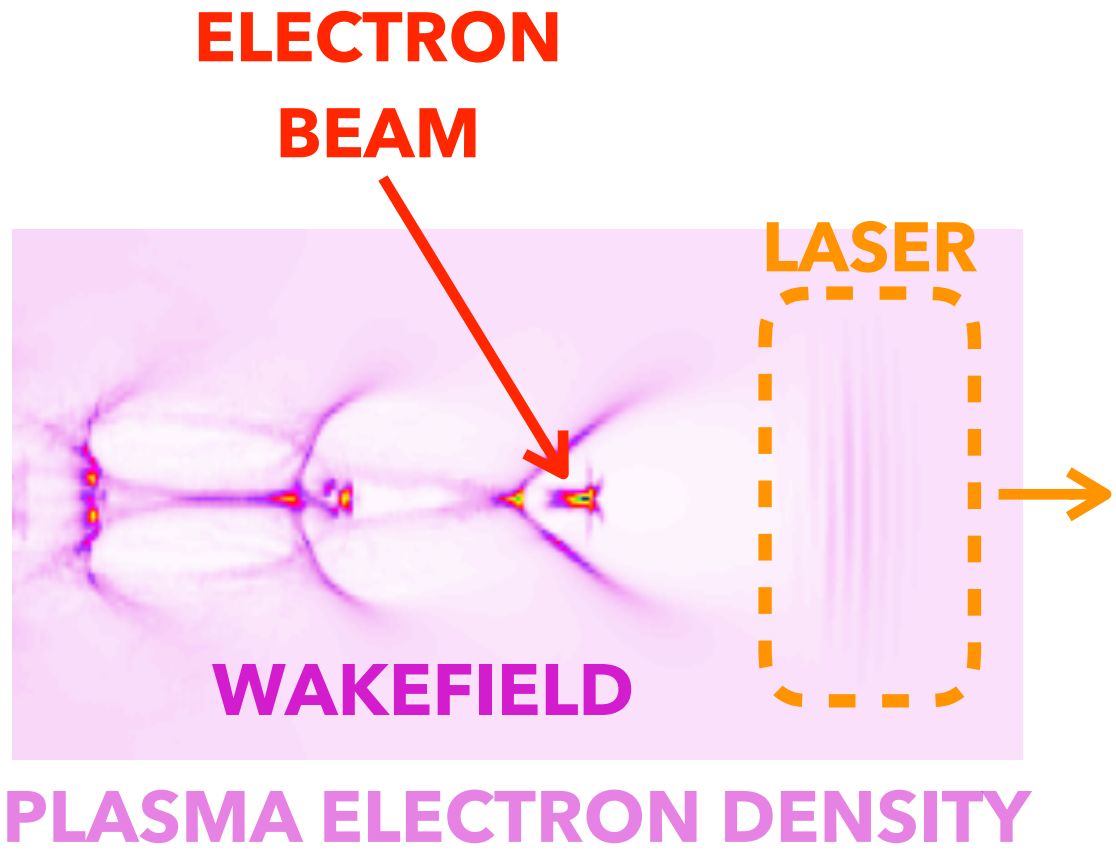


LASER ACCELERATOR

In laser wakefield accelerator, electrons surf on a plasma wave

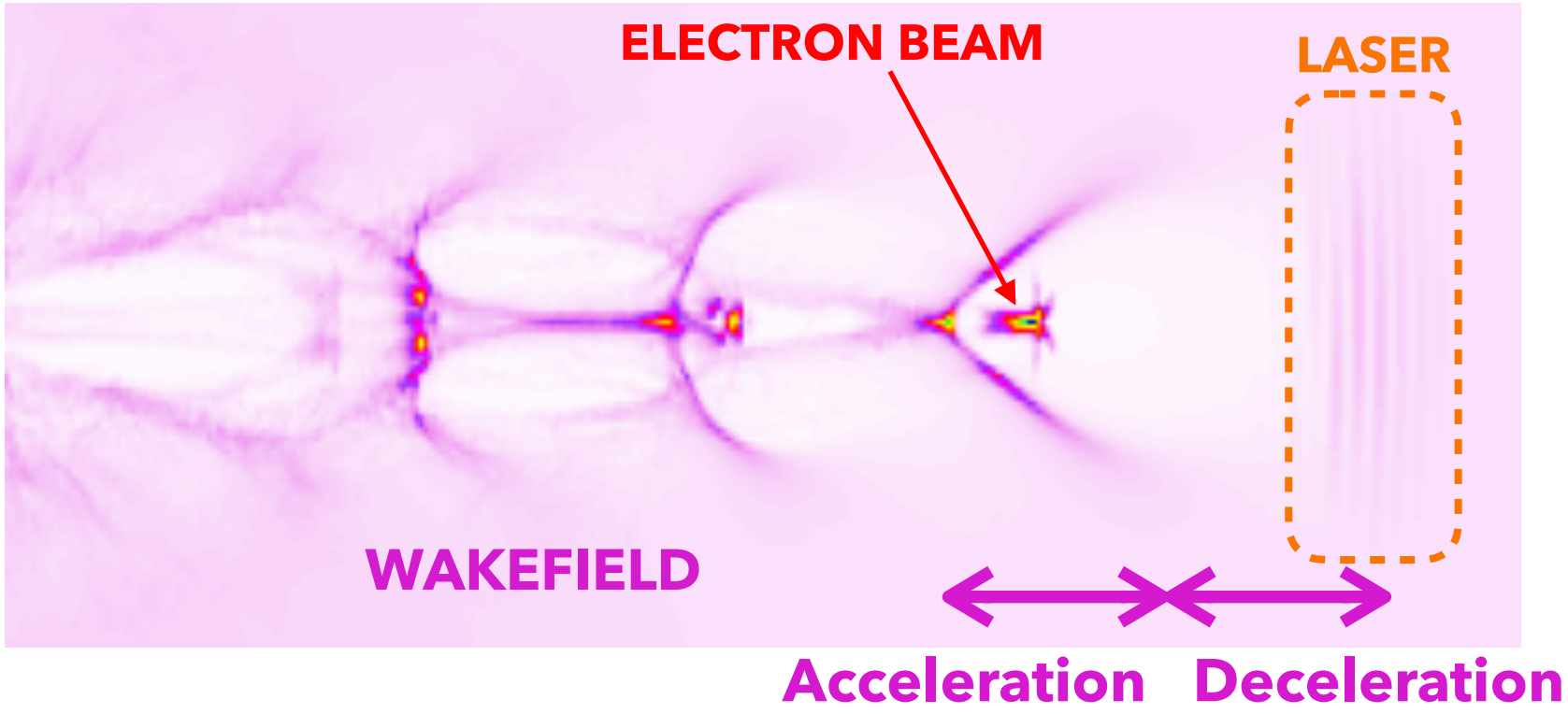


Picture credit: https://www.precision-performance.com/blog/wakesurfing--how-to-get-started-9429



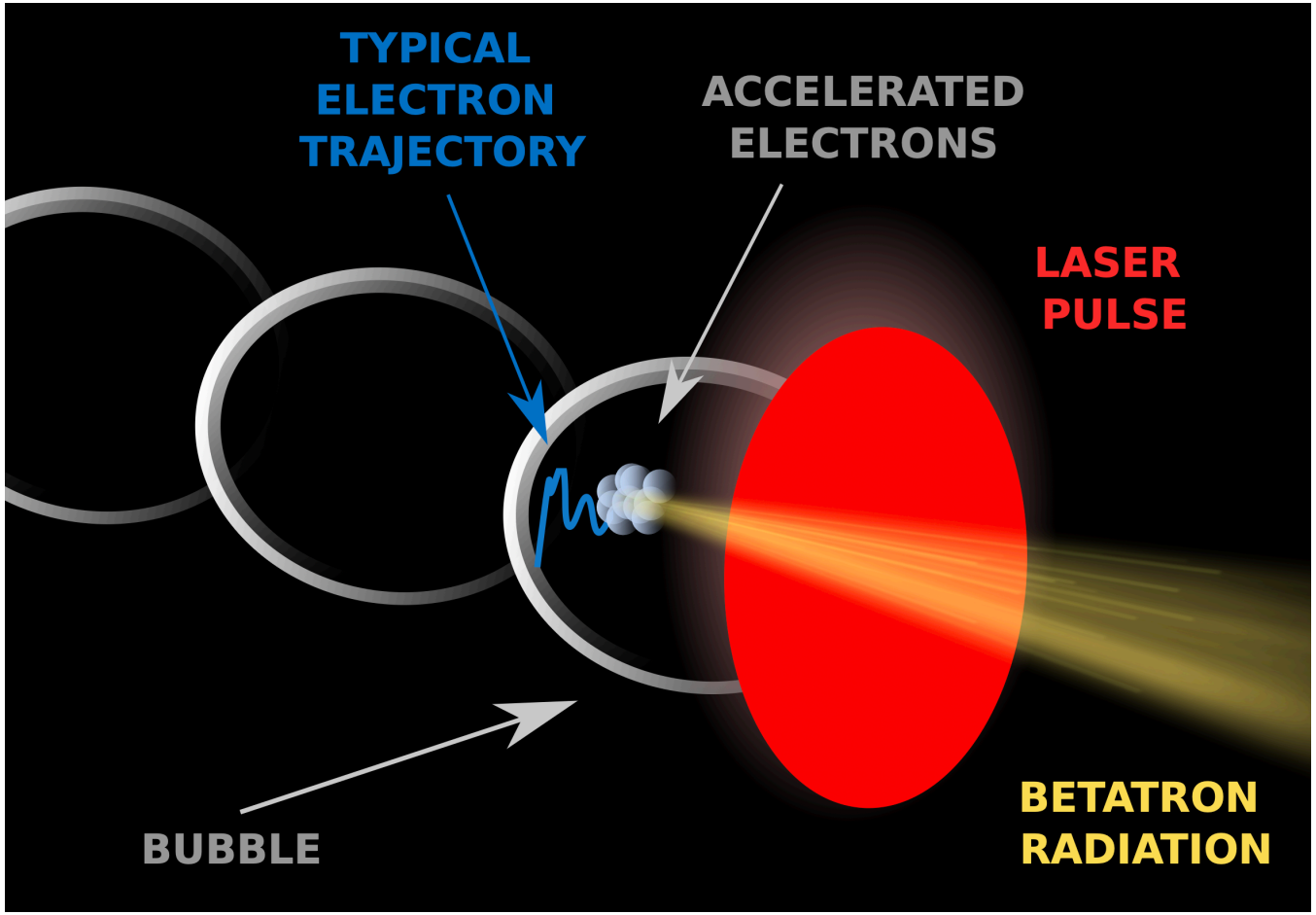
Laser wakefield acceleration accelerates electrons on very short scales

- Laser wakefield acceleration uses an **intense laser pulse** to **drive a plasma wave (wakefield)**
- Strongest accelerating gradients are in the nonlinear regime (**bubble regime**)
  $a_0 \approx 0.85(\text{intensity}\,[10^{18}\text{Wcm}^{-2}]\cdot\text{wavelength}[\mu\text{m}])^{1/2} \gg 1$
- An ion cavity (bubble) formation



Electron oscillations in wakefield generate betatron radiation

- Electrons undergo **transverse oscillations**
- They produce **femtosecond (10^{-15} s) X-ray betatron radiation bursts**
- Simulations have shown that the durations can be decreased down to attoseconds (10^{-18} s)
- Applications would benefit from the **increase of the radiation gain**

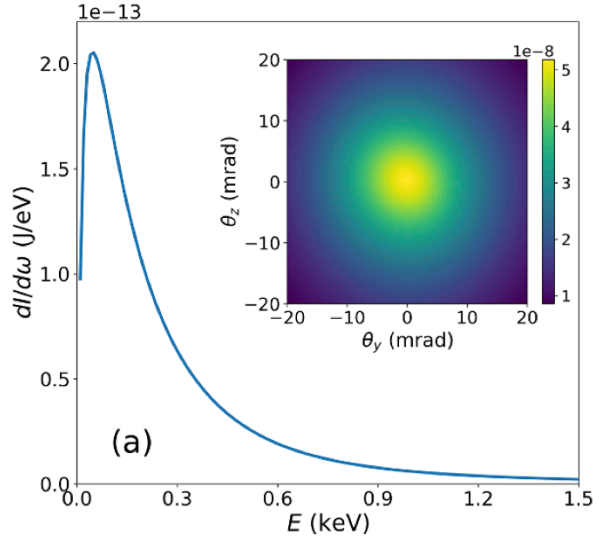


Betatron pulses could go toward the attosecond regime

- LWFA electron beams **typically have durations on femtosecond timescales** corresponding to the typical dimensions of the wakefield structure

- It is also possible to **push the electron beam durations to attoseconds:**
  - Luttikhof, M., Khachatryan, A., Van Goor, F. & Boller, K.-J. *Phys. Rev. Lett.* 105, 124801 (2010),
  - Tomassini, P. et al. In *Laser Acceleration of Electrons, Protons, and Ions IV*, vol. 10240, 81–98, (SPIE, 2017),
  - Tooley, M. et al. *Phys. Rev. Lett.* 119, 044801 (2017),
  - Zhao, Q. et al. *Plasma Phys. Control. Fusion* 61, 085015 (2019),
  - Kim, J., Wang, T., Khudik, V. & Shvets, G. *Phys. Rev. Lett.* 127, 164801 (2021),
  - Deng, A., Li, X., Luo, Z., Li, Y. & Zeng, J. *Opt. Express* 31, 19958–19967 (2023),
  - Tomassini, P., Horny, V. & Doria, D. *Instruments* 7, 34 (2023),
  - Sun, T., Zhao, Q., Wan, F., Salamin, Y. I. & Li, J.-X. *Phys. Rev. Lett.* 132, 045001 (2024),
  - Horný, V., Krus, M., Yan, W. & Fülöp, T. *Sci. Reports* 10, 15074 (2020),
  - Ferri, J., Horný, V. & Fülöp, T. *Plasma Phys. Control. Fusion* 63, 045019 (2021).

- **Betatron radiation can inherit the attosecond scale:**
  - Horný, V., Krus, M., Yan, W. & Fülöp, T. *Sci. Reports* 10, 15074 (2020),
  - Ferri, J., Horný, V. & Fülöp, T. *Plasma Phys. Control. Fusion* 63, 045019 (2021).

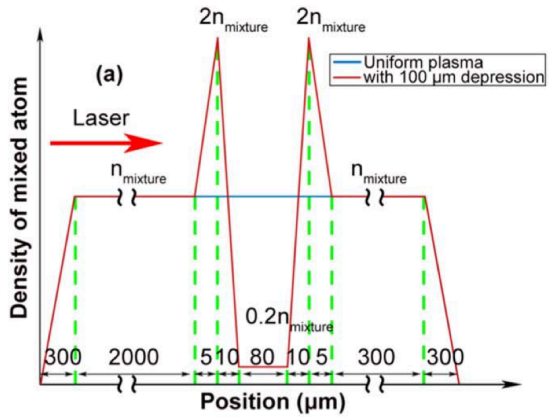


Ferri, Julien, Vojtech Horný, and Tünde Fülöp. *Plasma Physics and Controlled Fusion* 63.4 (2021): 045019.

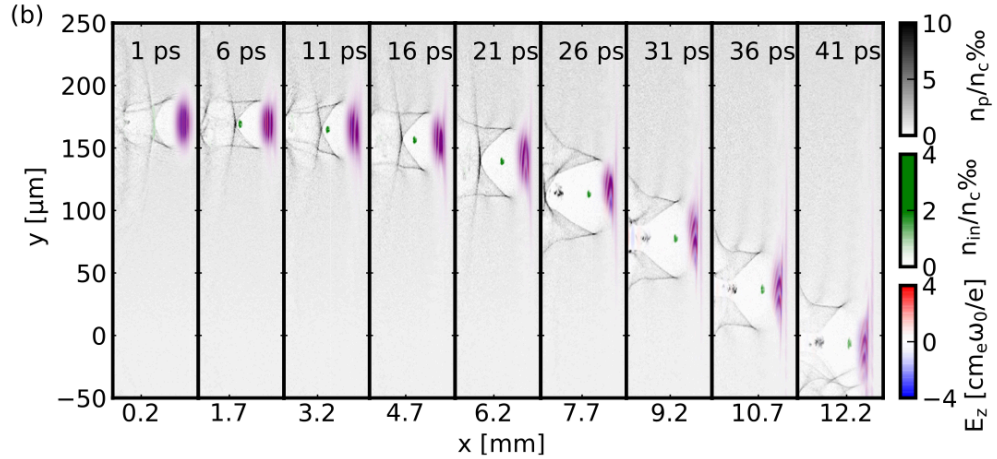
Betatron pulses could go toward the attosecond regime

- Practical applications could benefit from increasing the betatron gain
- Betatron gain can be boosted by **increasing the oscillation amplitude, energy or charge of the electron beam**
- Several longitudinal and transverse density modulations have been proposed to boost these properties in fs betatrons

- Guo, B. et al. Plasma Phys. Control. Fusion 61, 035003, (2019)



- Ma, Y. et al. Phys. Plasmas 25, 113105, (2018)



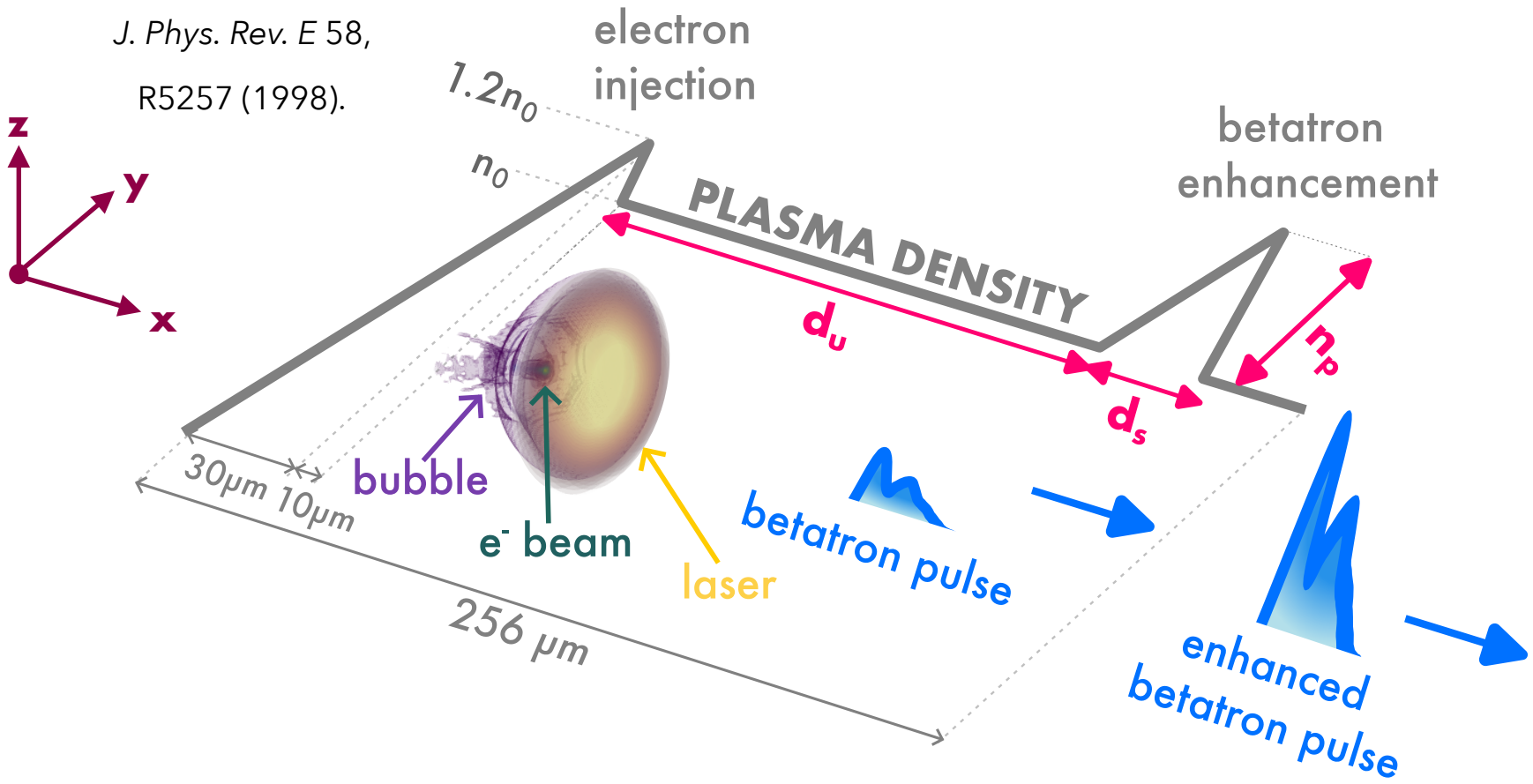
- Ferri, J. & Davoine, X. *Phys. Rev. Accel. Beams* 21, 091302,
- Ta Phuoc, K. et al. *Phys. Plasmas* 15, 063102, (2008),
- Ferri, J. et al. *Phys. Rev. Lett.* 120, 254802 (2018),
- Gautier, J. et al. arXiv:2506.22021 (2025), …

We tailored the properties of a density spike to enhance radiation

Bulanov, S. et al.,

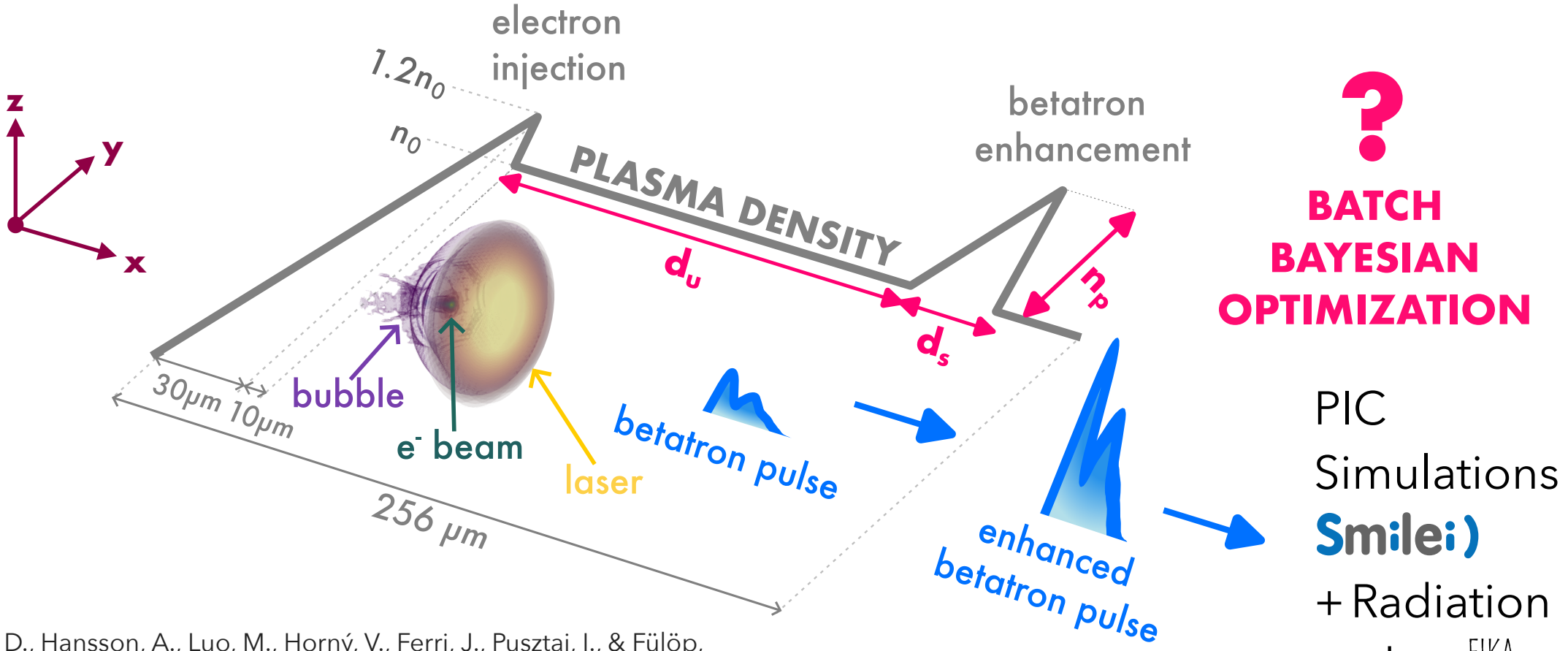
J. Phys. Rev. E 58,

R5257 (1998).



Ferri, J., Horný, V. & Fülöp, T. *Plasma Phys. Control. Fusion* 63, 045019 (2021).

We tailored the properties of a density spike to enhance radiation



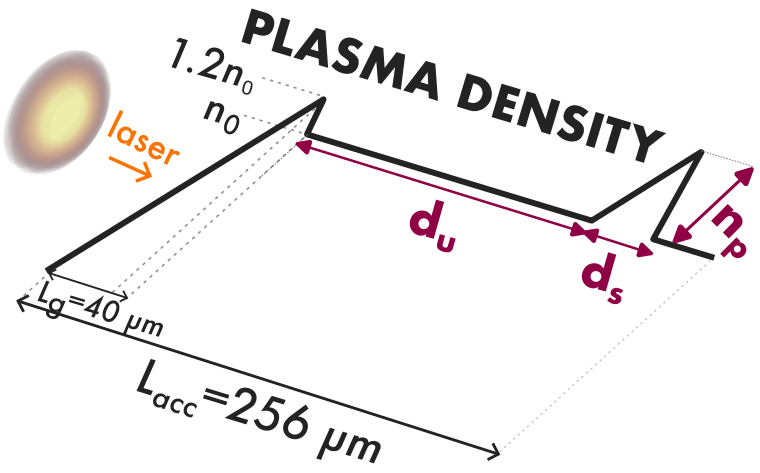
Maslarova, D., Hansson, A., Luo, M., Horný, V., Ferri, J., Pusztai, I., & Fülöp, T. (2026). Batch Bayesian optimization of attosecond betatron pulses from laser wakefield acceleration. *Communications Physics*, *9*(1), 92.

We tailored the properties of a density spike to enhance radiation

PICK INPUT POINT $\{d_u, d_s, n_p\}$
WHAT IS THE OPTIMAL ONE?

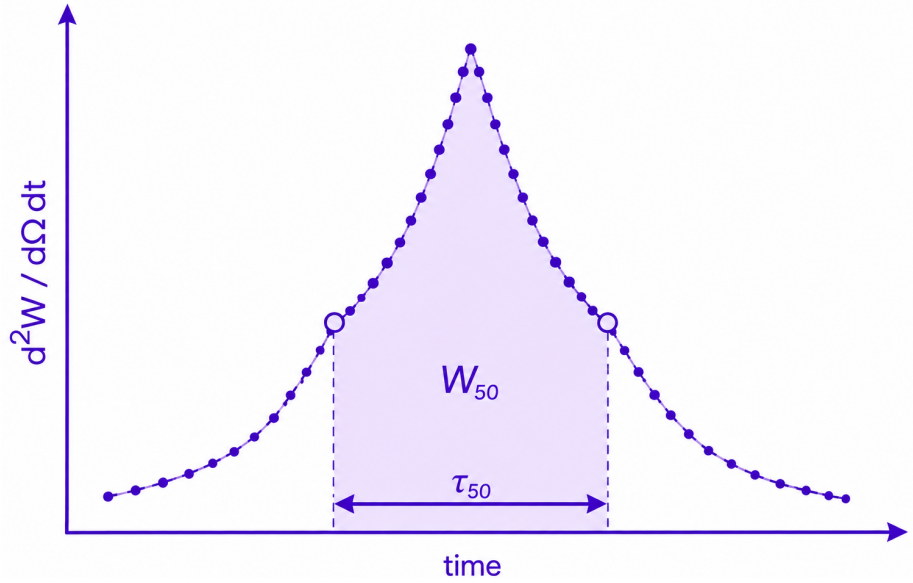
- **Spike position $d_u \in [0, L_{acc} - L_g - 5\ \mu m]$**
- **Spike length $d_s \in [5\ \mu m, L_{acc} - L_g]$**
- **Density peak $n_p \in (1, 4n_0]$**

$$d_u + d_s + L_g \leq L_{acc}$$



OPTIMIZE THE COST FUNCTION

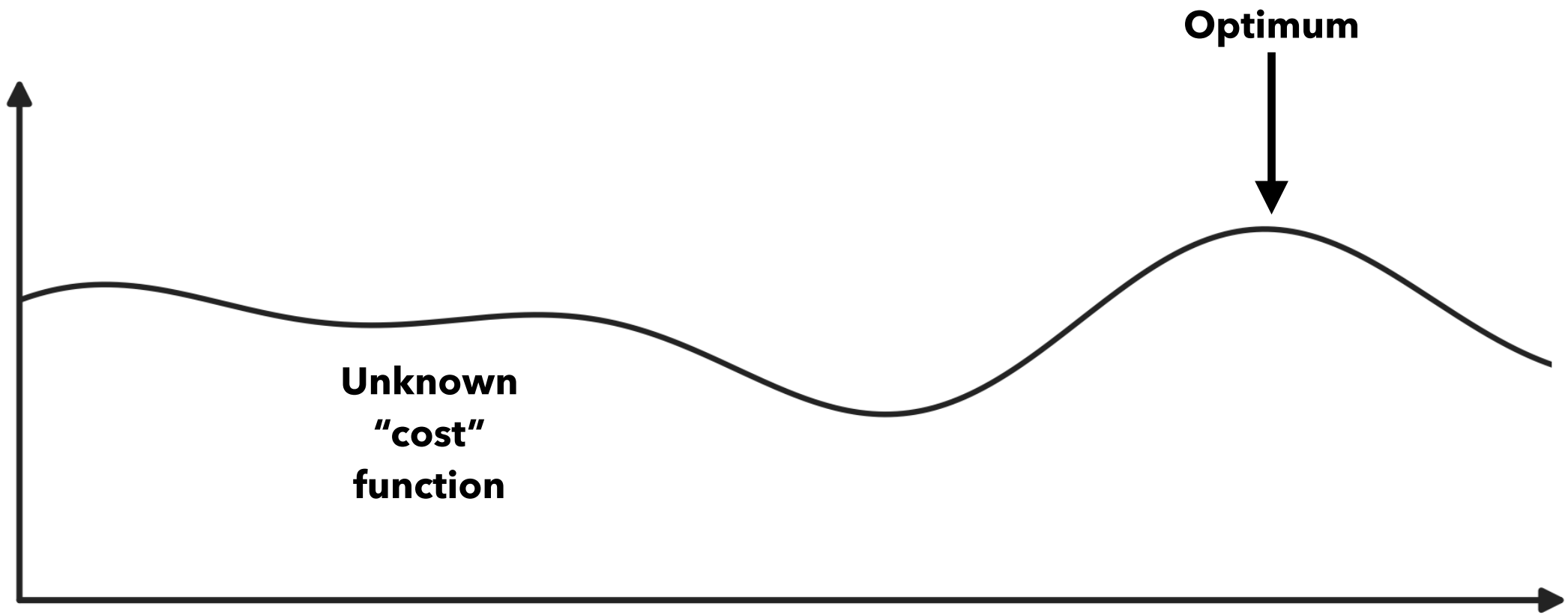
$C(W_{50}, \tau_{50}) = W_{50} / \tau_{50}$



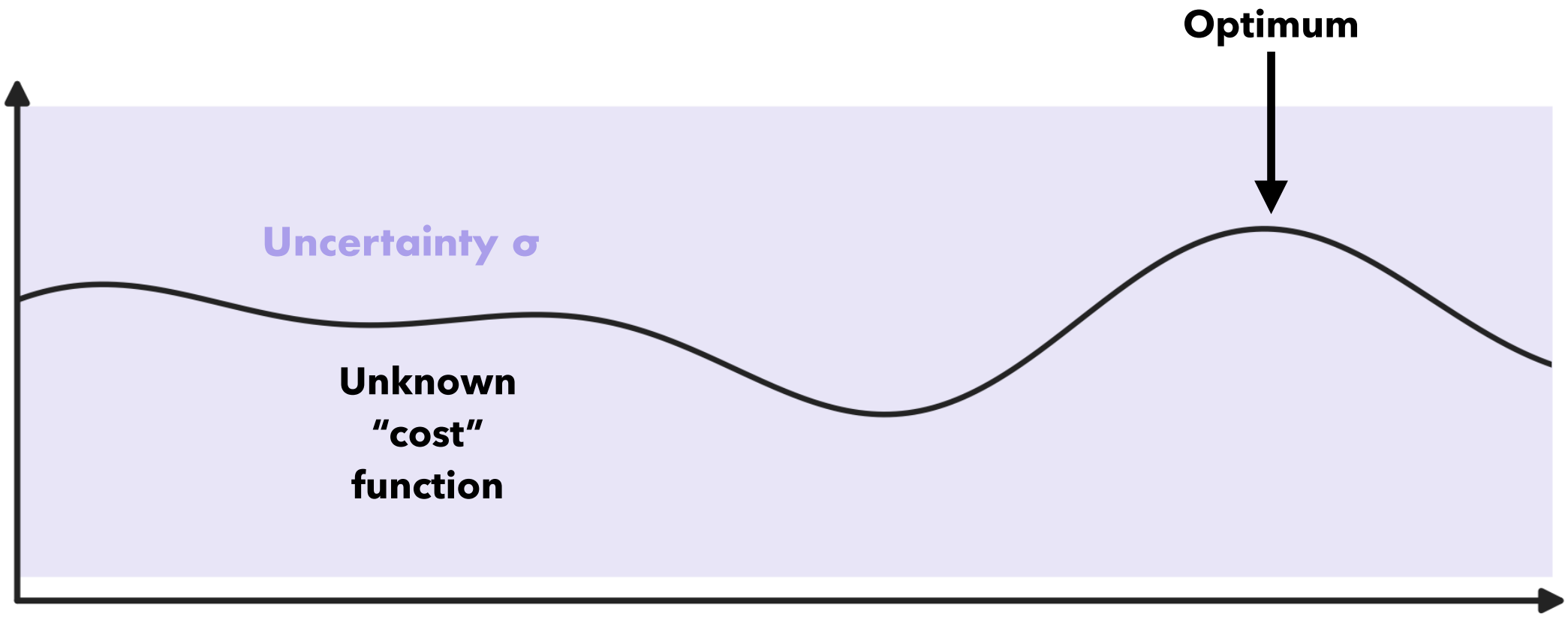
W_{50} - energy per solid angle contained within the central 50% of $d^2W / d\Omega\, dt$ on axis (between 25% and 75% of the cumulative curve)

τ_{50} - the time interval over which W_{50} is emitted.

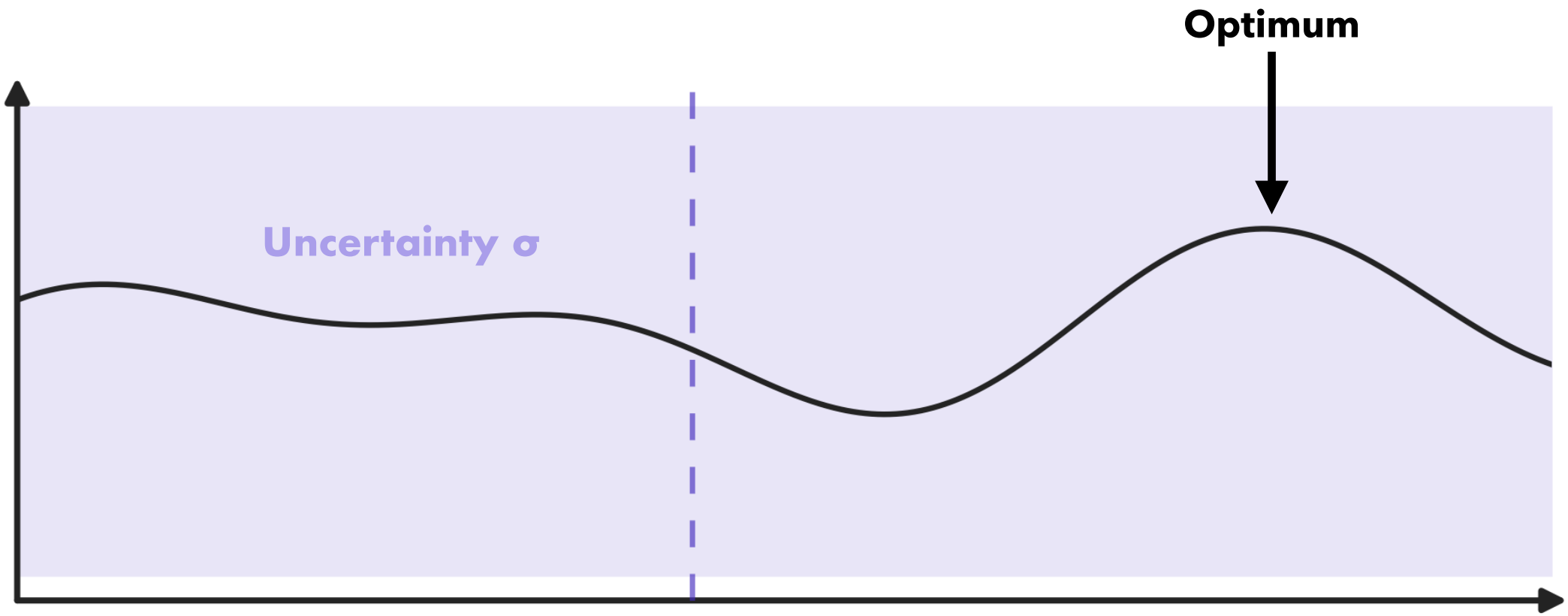
What is Bayesian optimization?



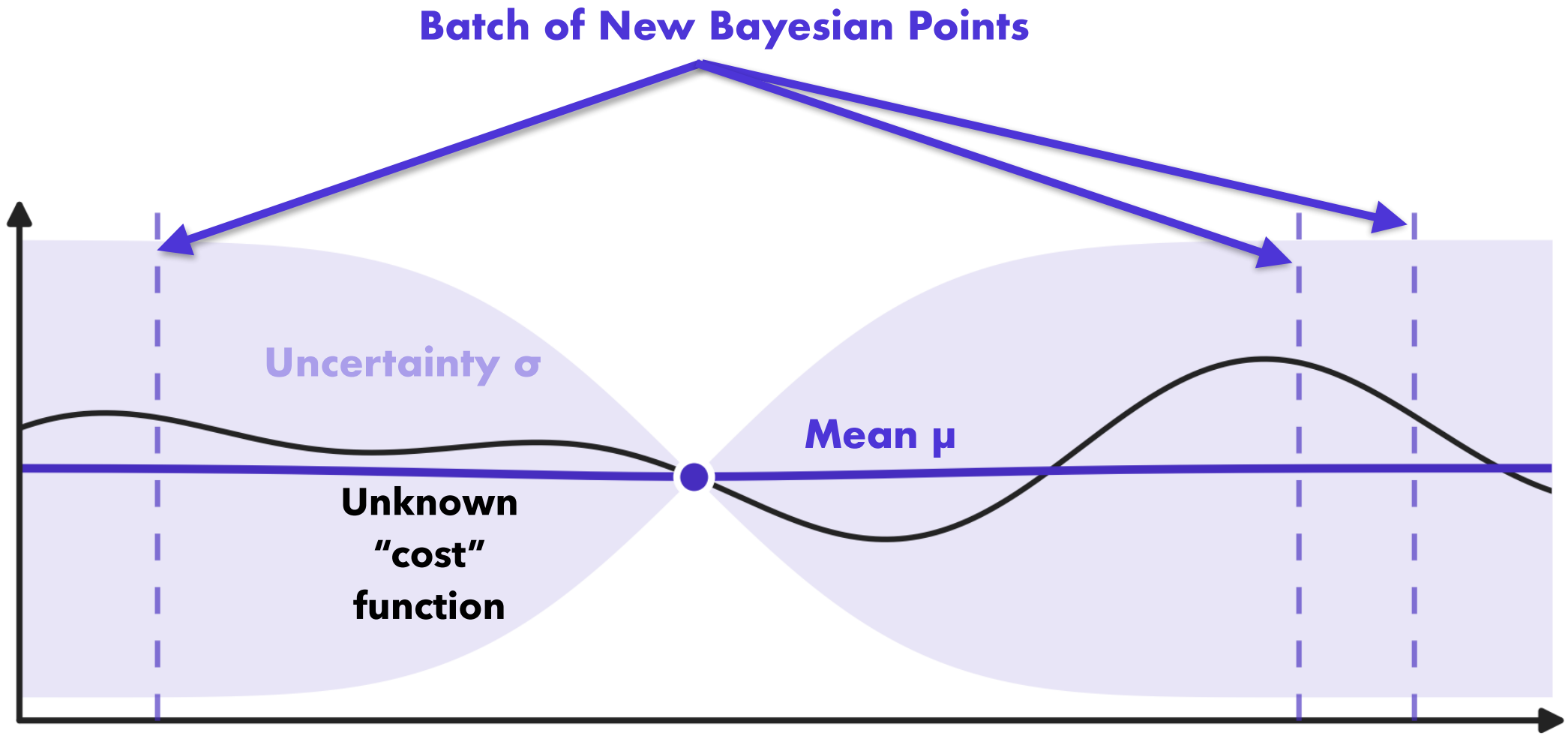
What is Bayesian optimization?



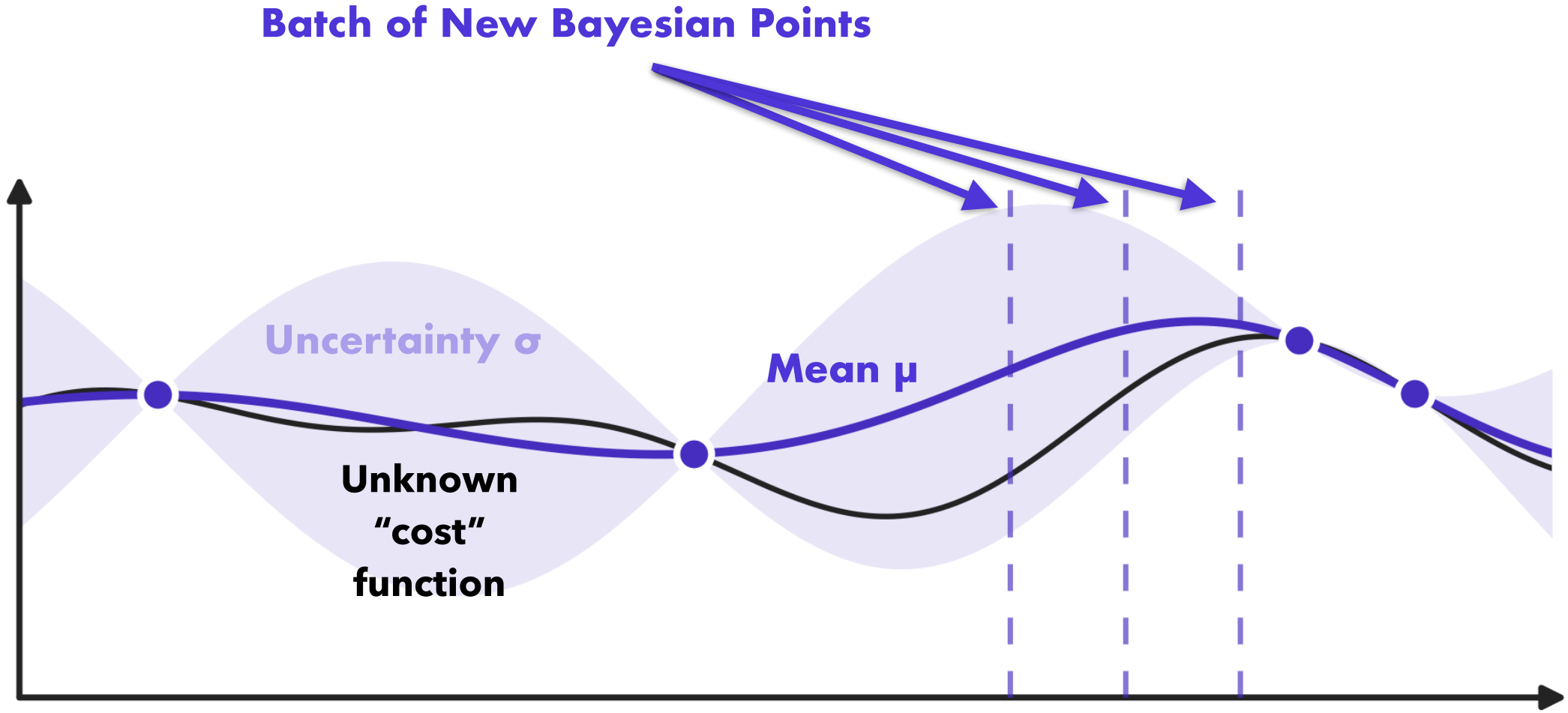
What is Bayesian optimization?



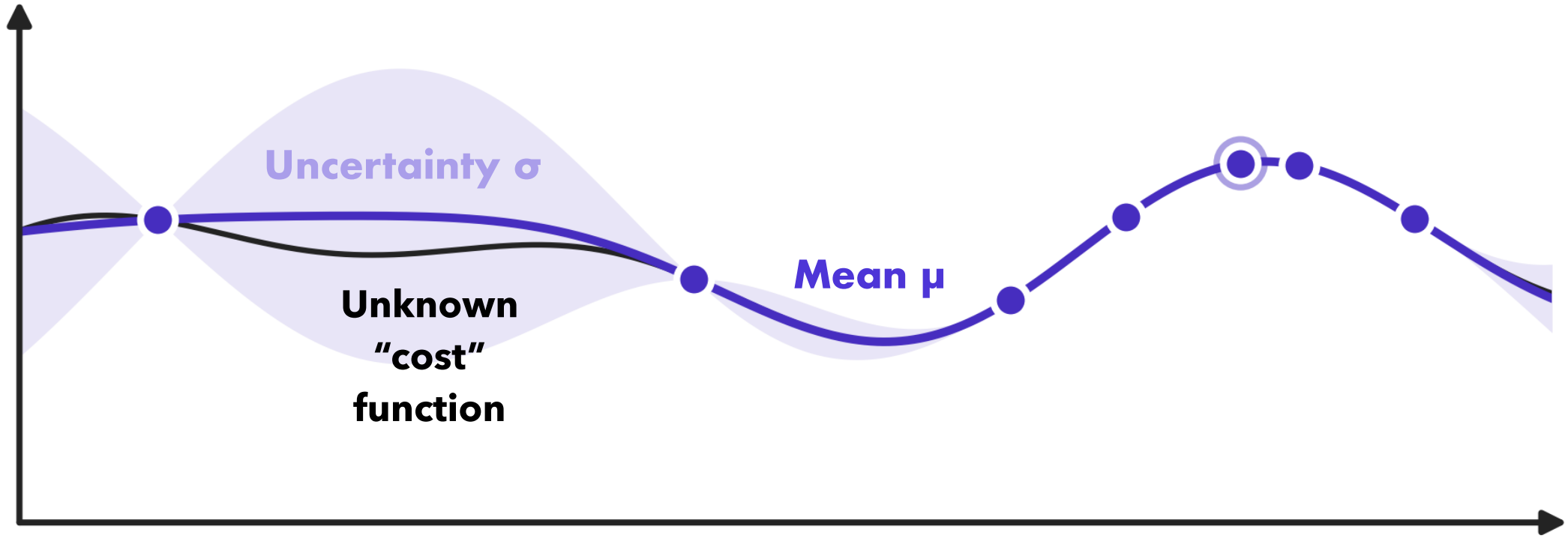
We used Batch Bayesian Optimization



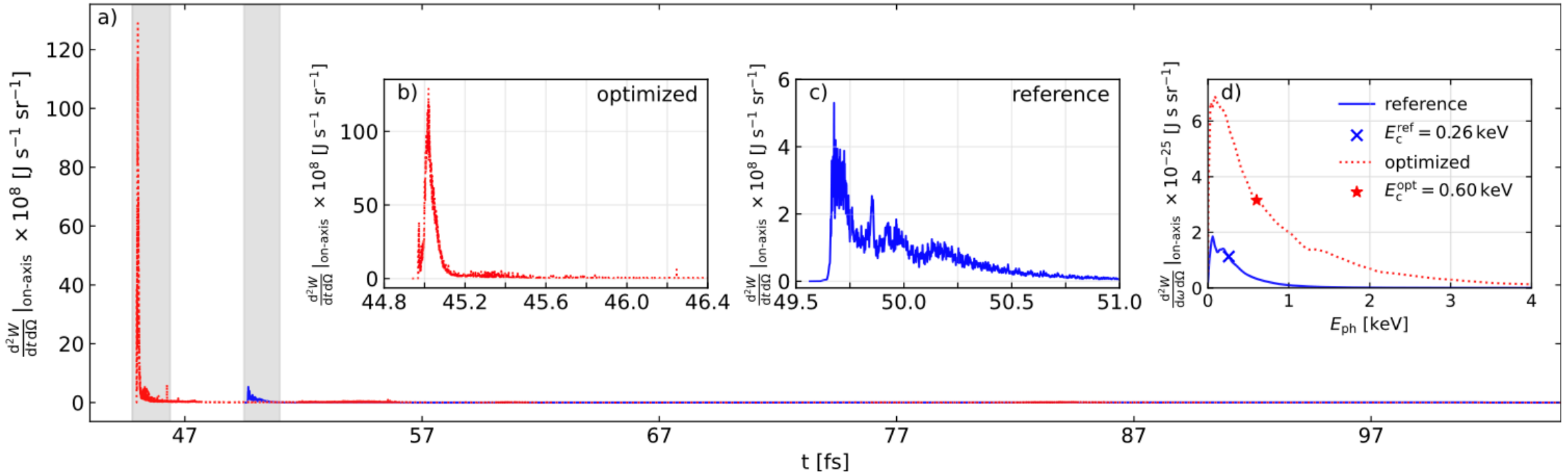
We used Batch Bayesian Optimization



We used Batch Bayesian Optimization

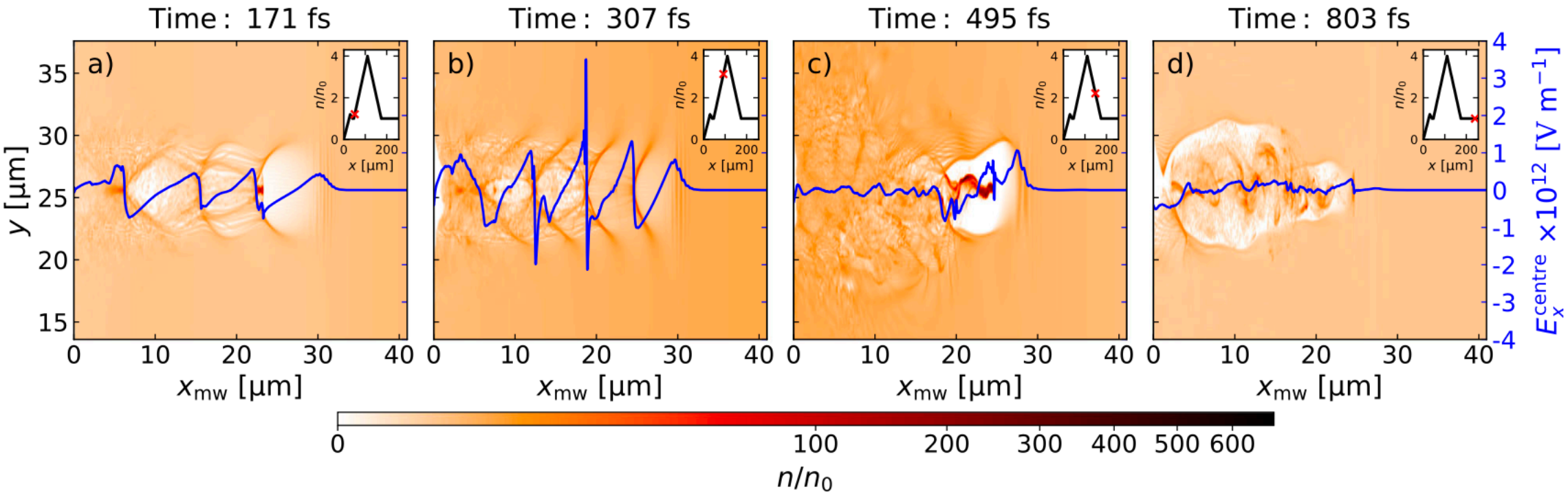


The best case improved the radiation significantly



- Batch Bayesian optimization improved the cost function more than 60x
- The energy in the central part increased by 6x
- The peak value of the radiation increased by 25x

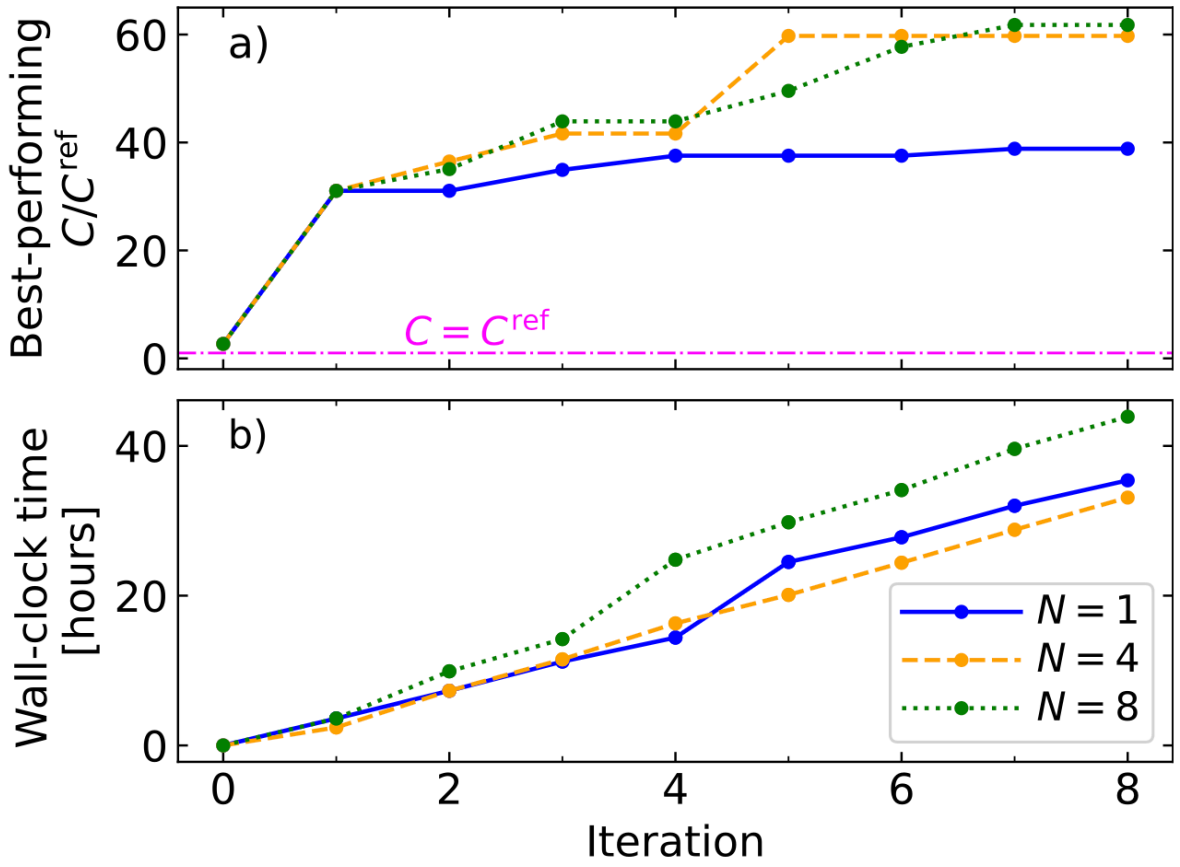
We tailored the properties of a density spike to enhance radiation



- Spike located few microns after the injection, approximately 120 micron-long showed the best performance
- Enhancement is due to a rich second electron injection

Different size of batches performs differently

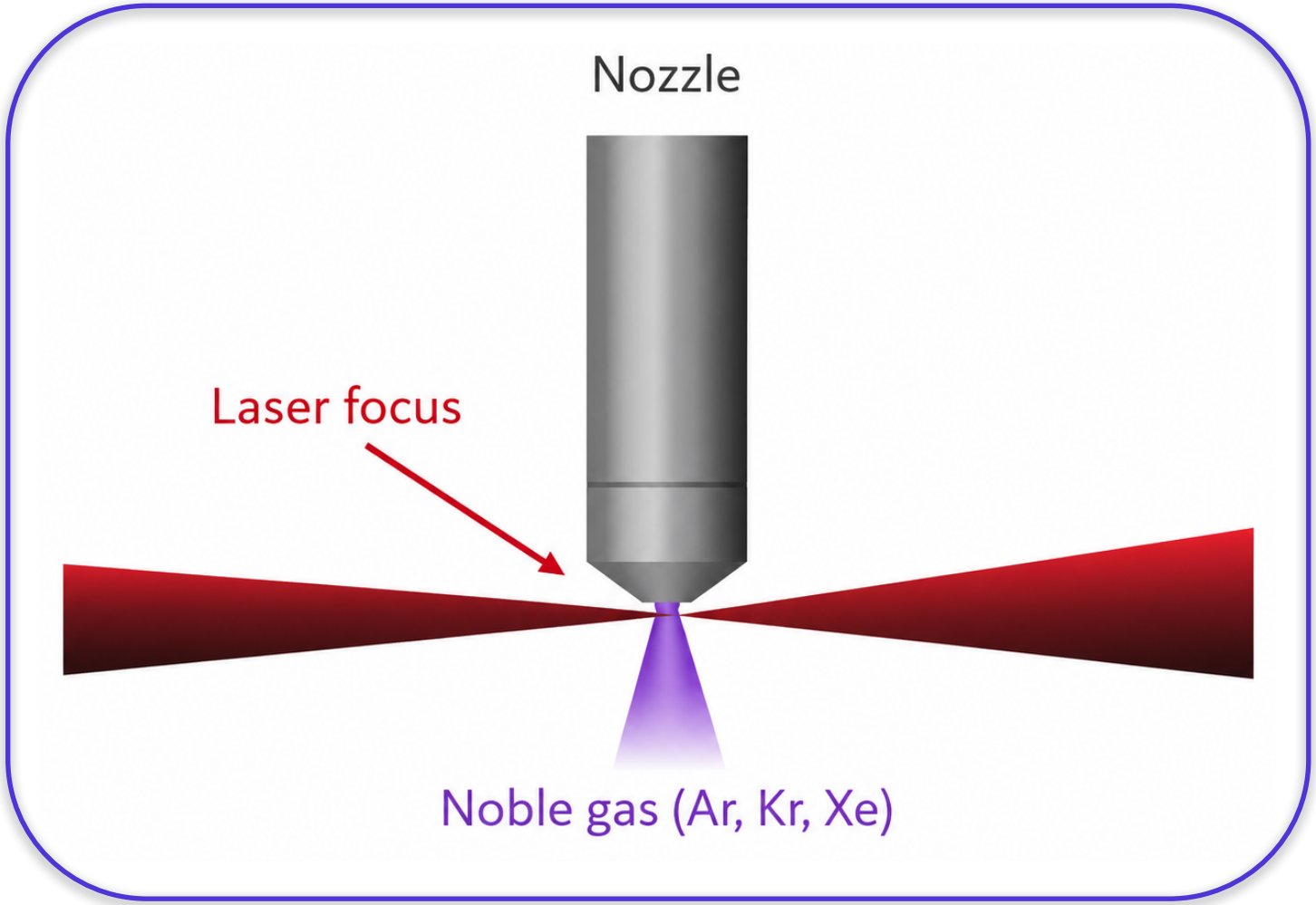
Having similar simulations times, optimization in batches performs faster



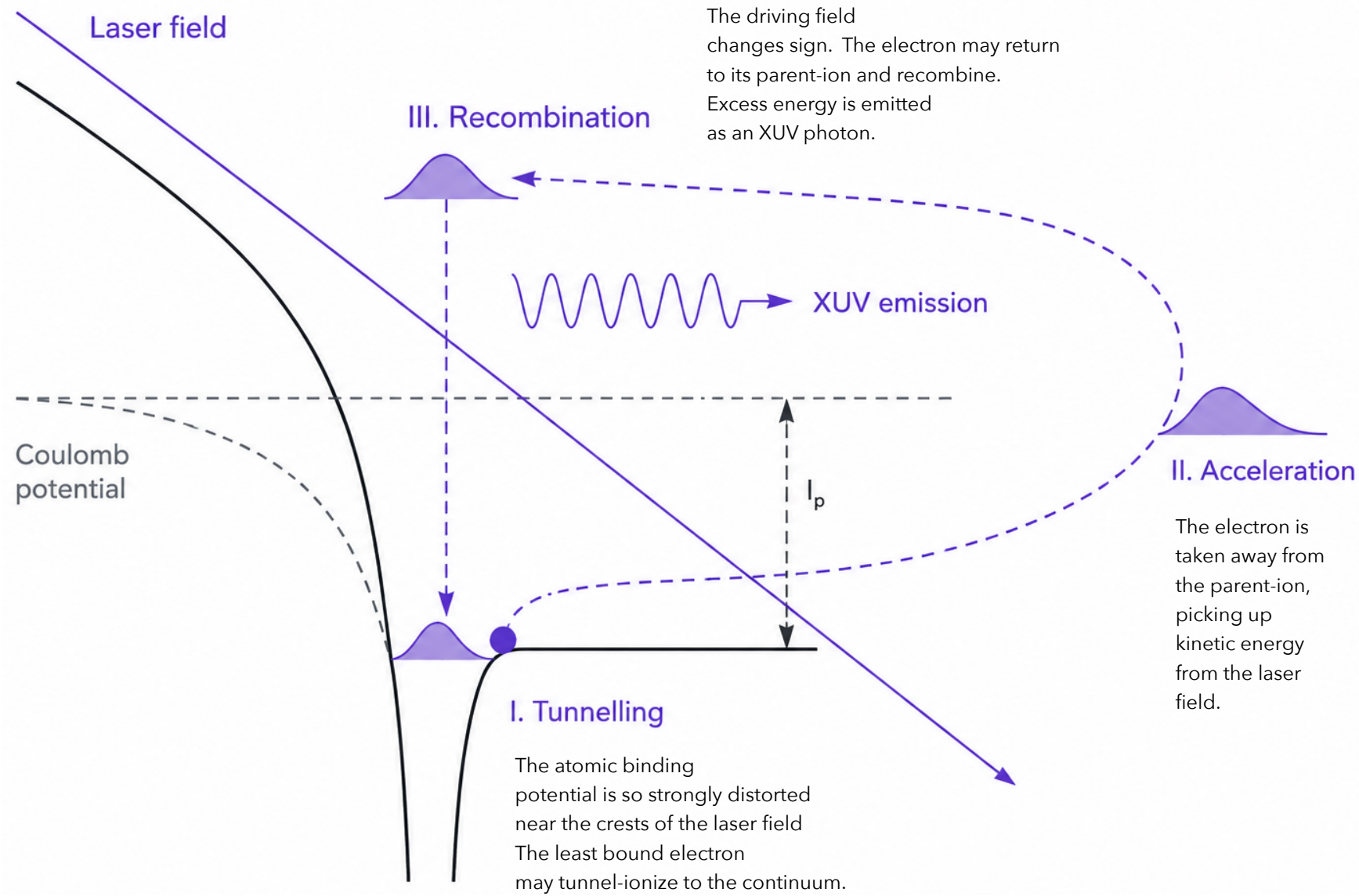
HIGH HARMONIC GENERATION IN GASES

High Harmonic Generation can be generated in gases with IR lasers

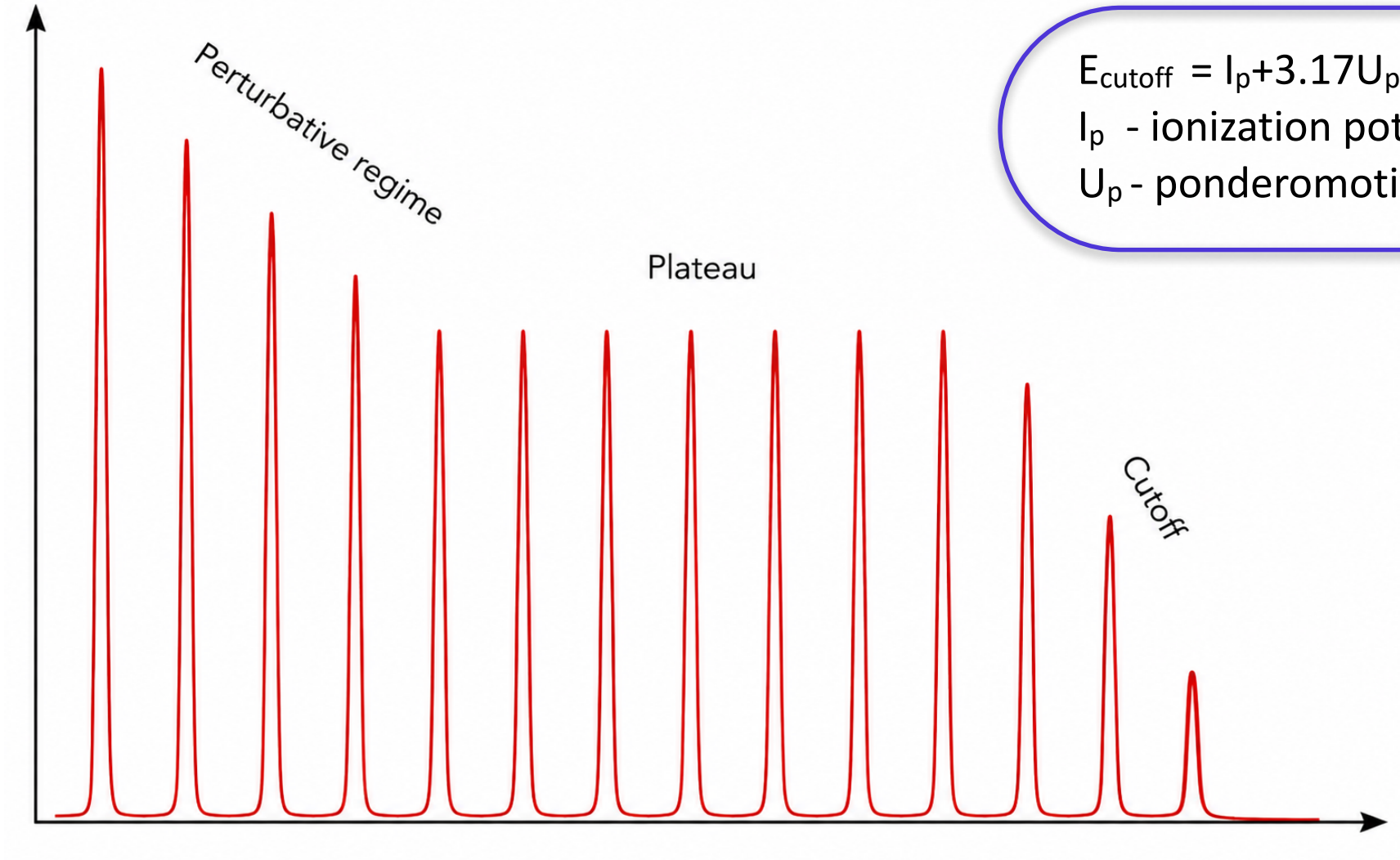
- Intense IR laser pulses interact with a **noble gas** to generate XUV/soft X-ray radiation
- Typical laser intensity: ~10^{14}–10^{15} W/cm^2
- **Key parameters:** laser intensity, wavelength and pulse duration, gas species, pressure and interaction length
- Challenge: increase the radiation yield while preserving attosecond duration



High Harmonic Generation can be described in 3 steps



Harmonics generate a plateau profile



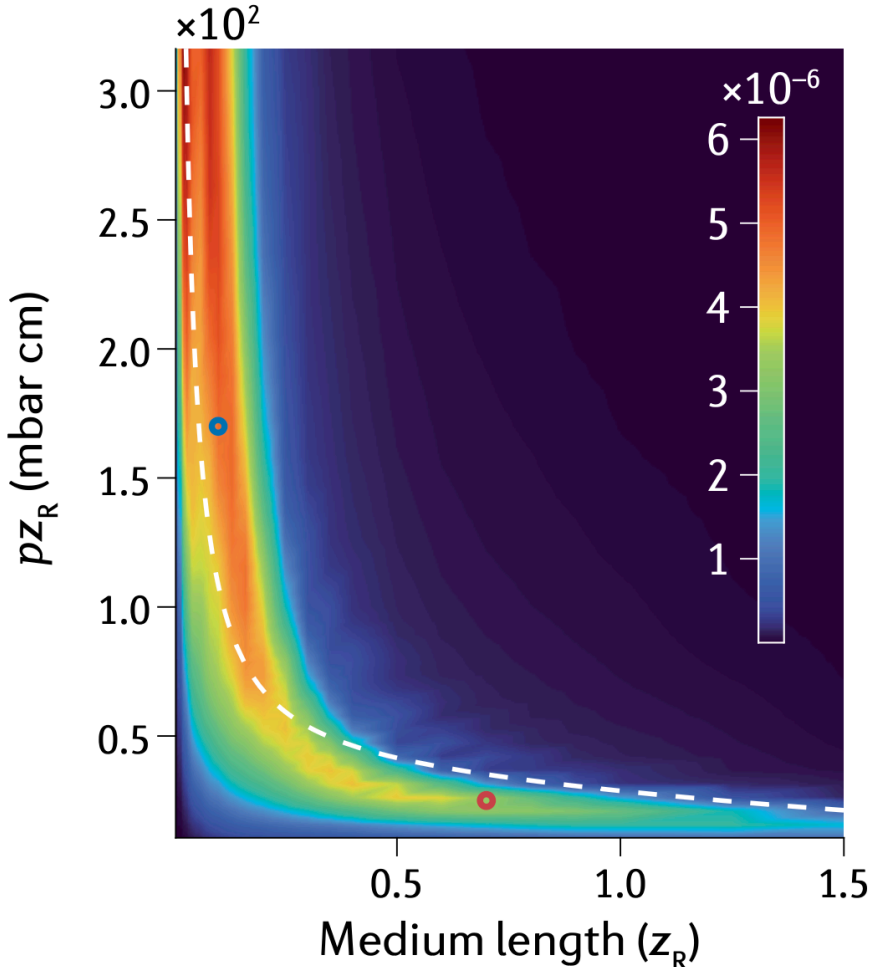
$E_{cutoff} = I_p + 3.17U_p$

I_p - ionization potential

U_p - ponderomotive potential

Best conversion efficiency lies on a hyperbola

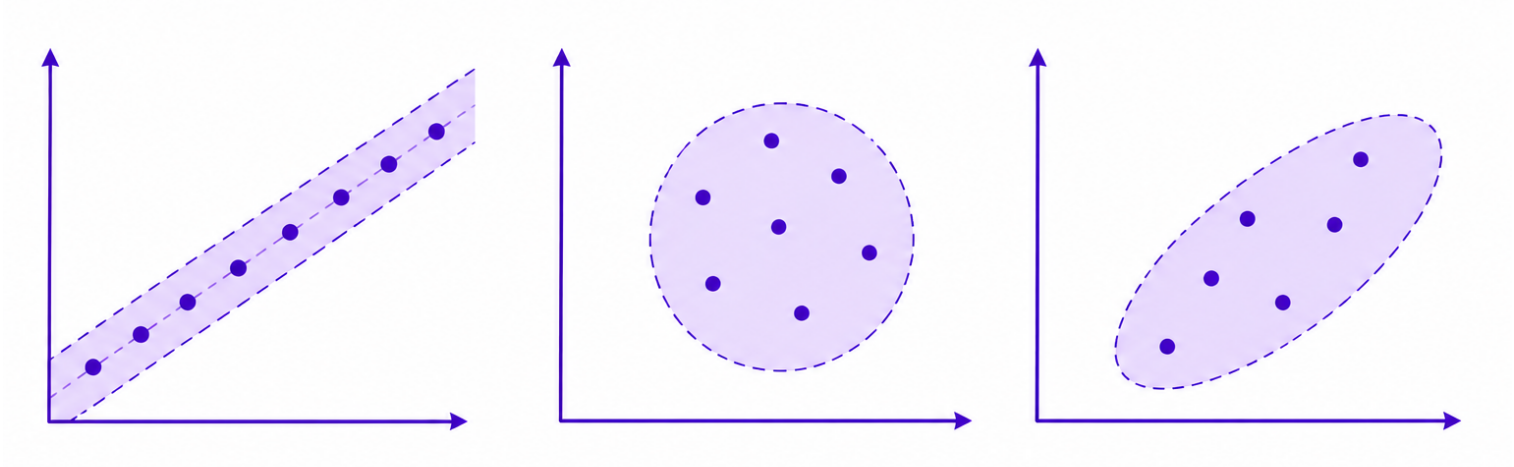
- HHG conversion efficiency depends strongly on the **gas pressure p** and **medium length L**
- There is no single optimal pressure or interaction length: many combinations can produce efficient HHG
- The optimal combinations approximately follow a **hyperbolic relation**, i.e. **pL ≈ const.**



Weissenbilder, R., Carlström, S., Rego, L. *et al.* How to optimize high-order harmonic generation in gases. *Nat Rev Phys* **4**, 713–722 (2022).

Bayesian optimization assumes that Euclidean-close points are similar…

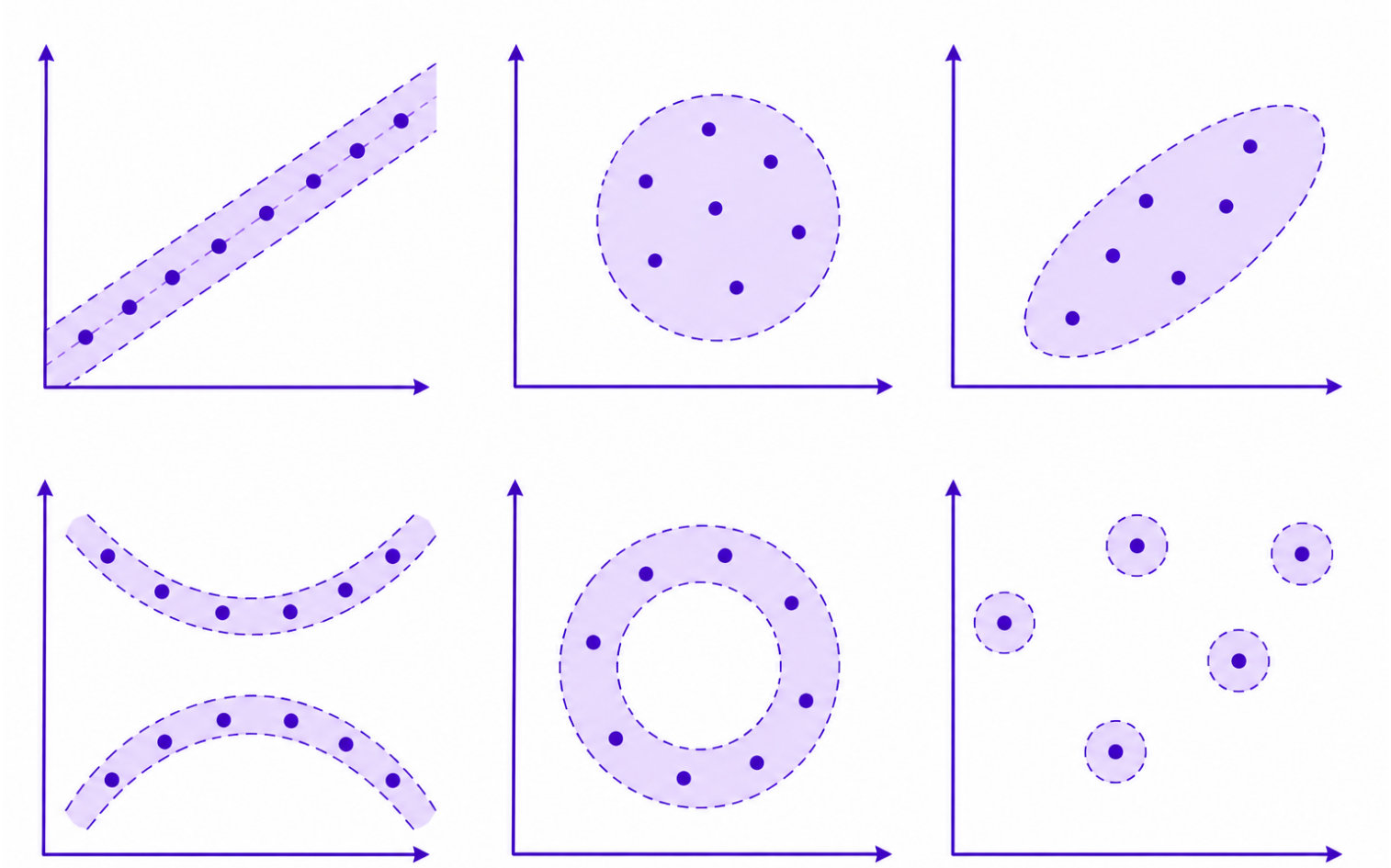
… problems for structures like hyperbola





Bayesian optimization assumes that Euclidean-close points are similar…

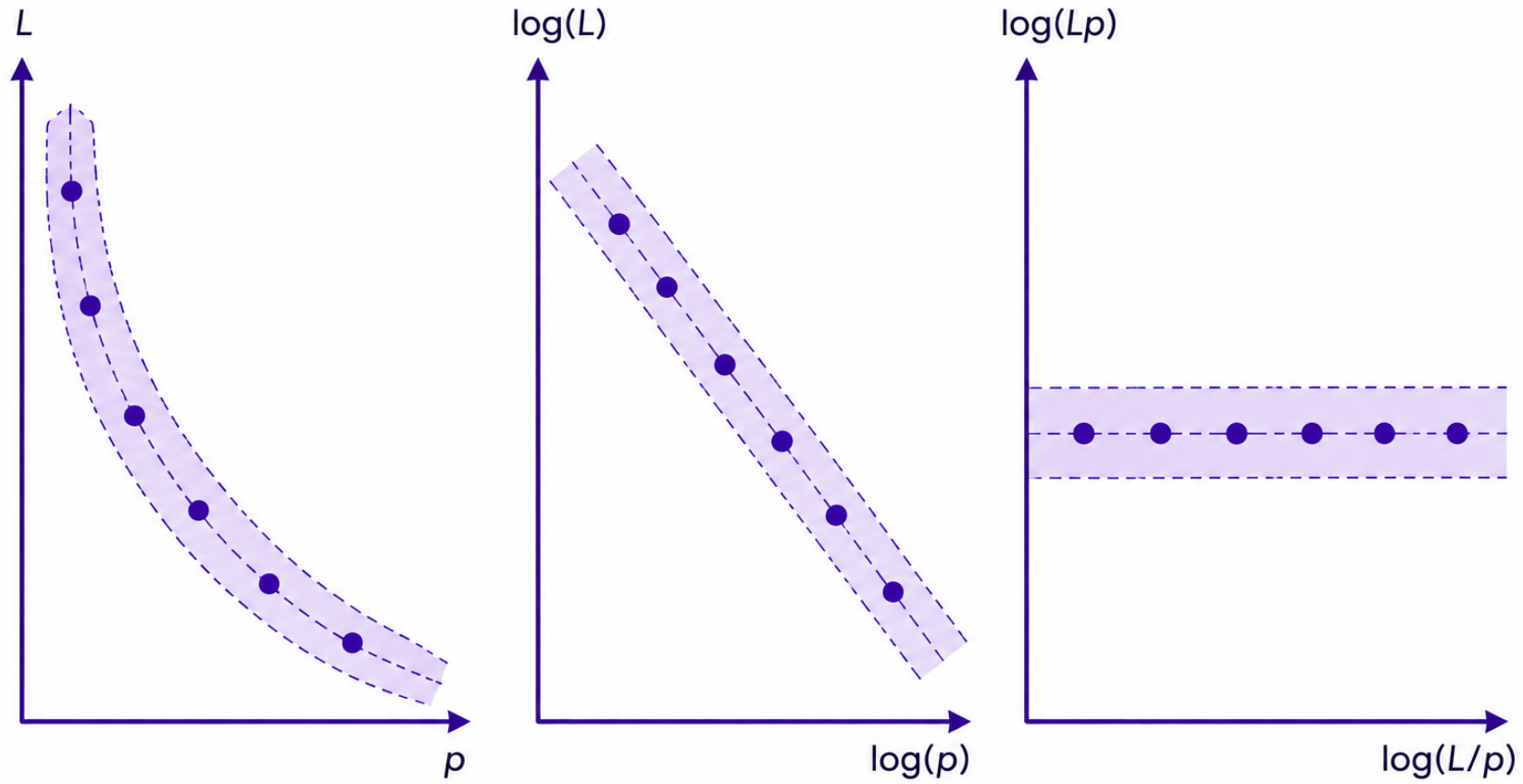
… problems for structures like hyperbola





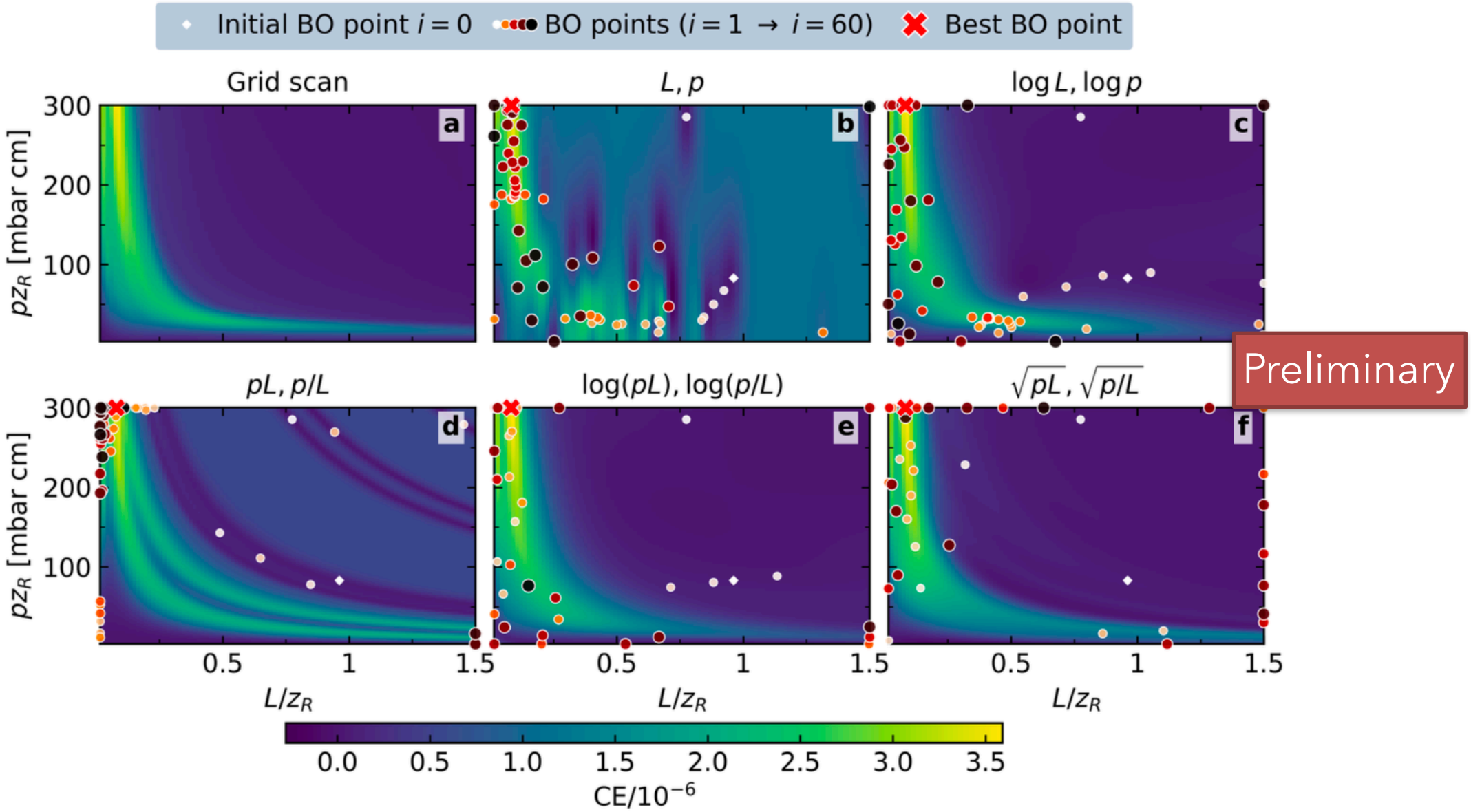


Transformation of the coordinates easily change the Euclidean distance...



=> even small knowledge about the physical system can improve Bayesian optimization

We observe a great match with 2D grid scan after transformation

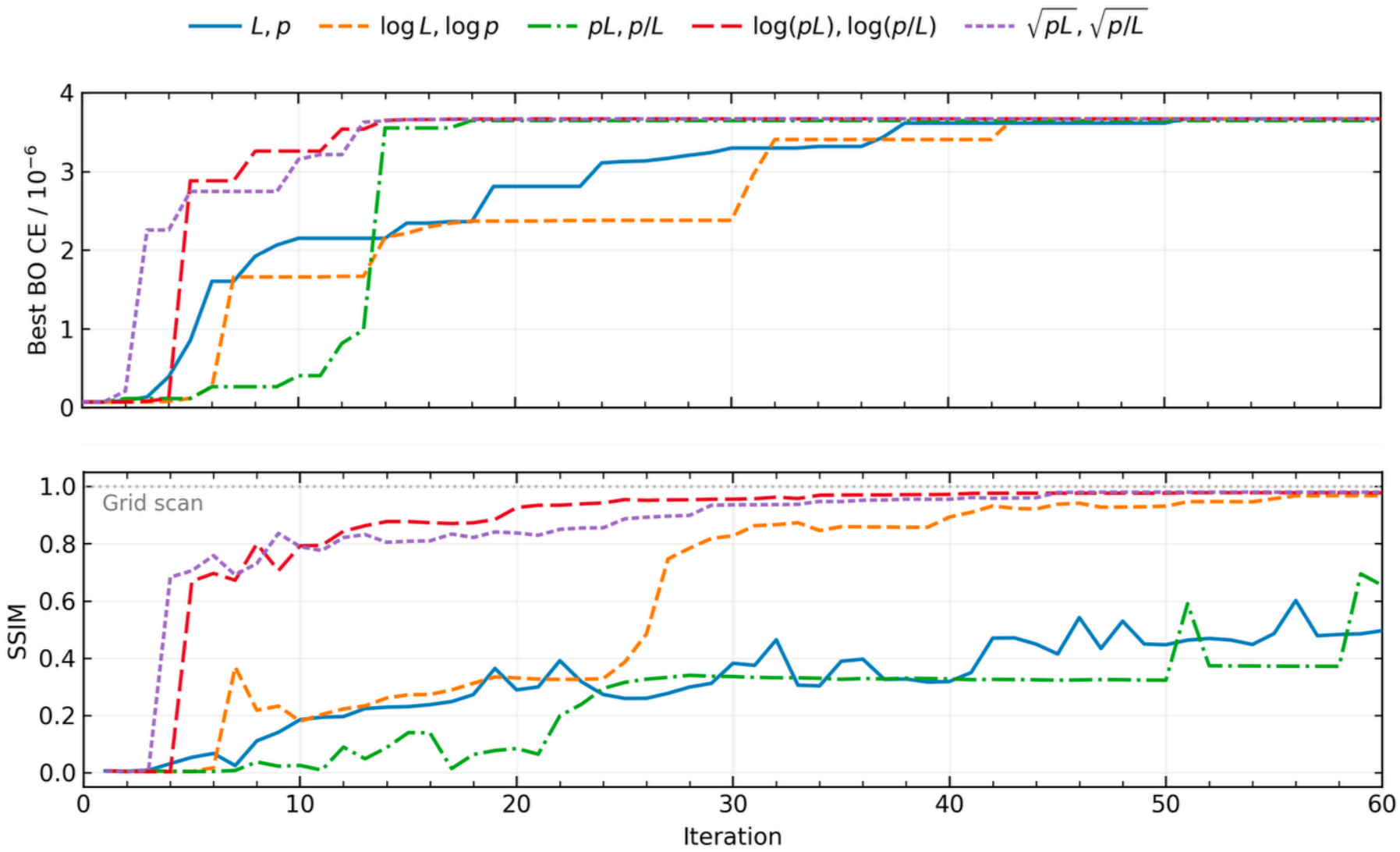


Preliminary

We observe a great match with 2D grid scan after transformation

Convergence of optimization with different parameter spaces

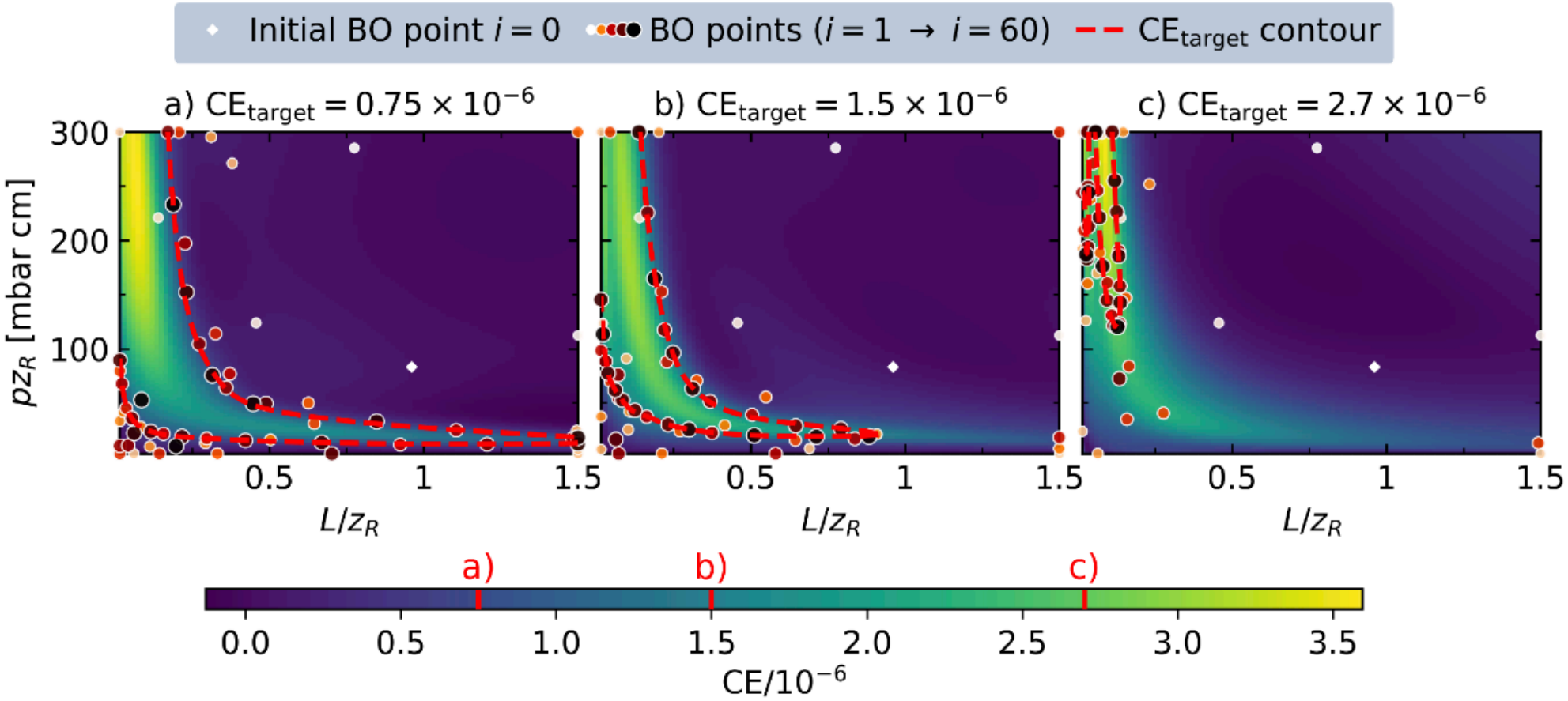
SSIM compared to a grid scan (10 000 simulations)



SSIM compares two images based on their brightness, contrast, and structural similarity,

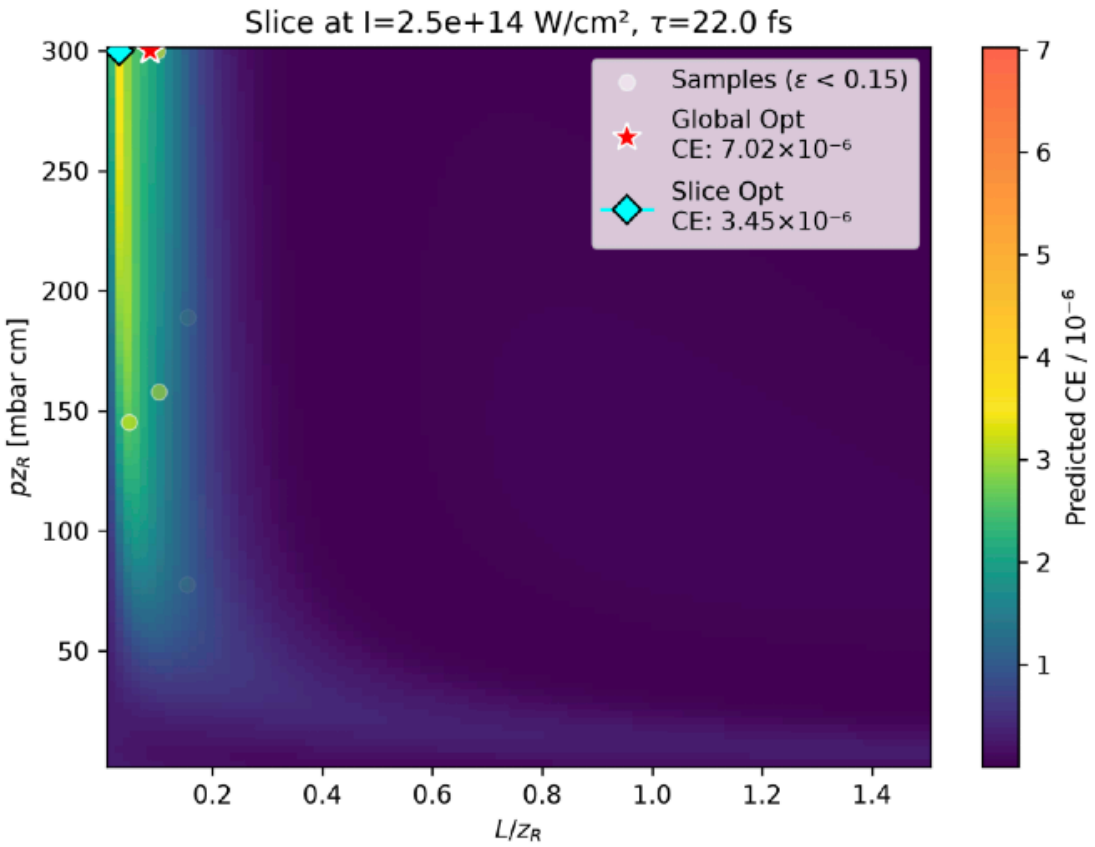
We can also look for areas of good performance (not only optimum)

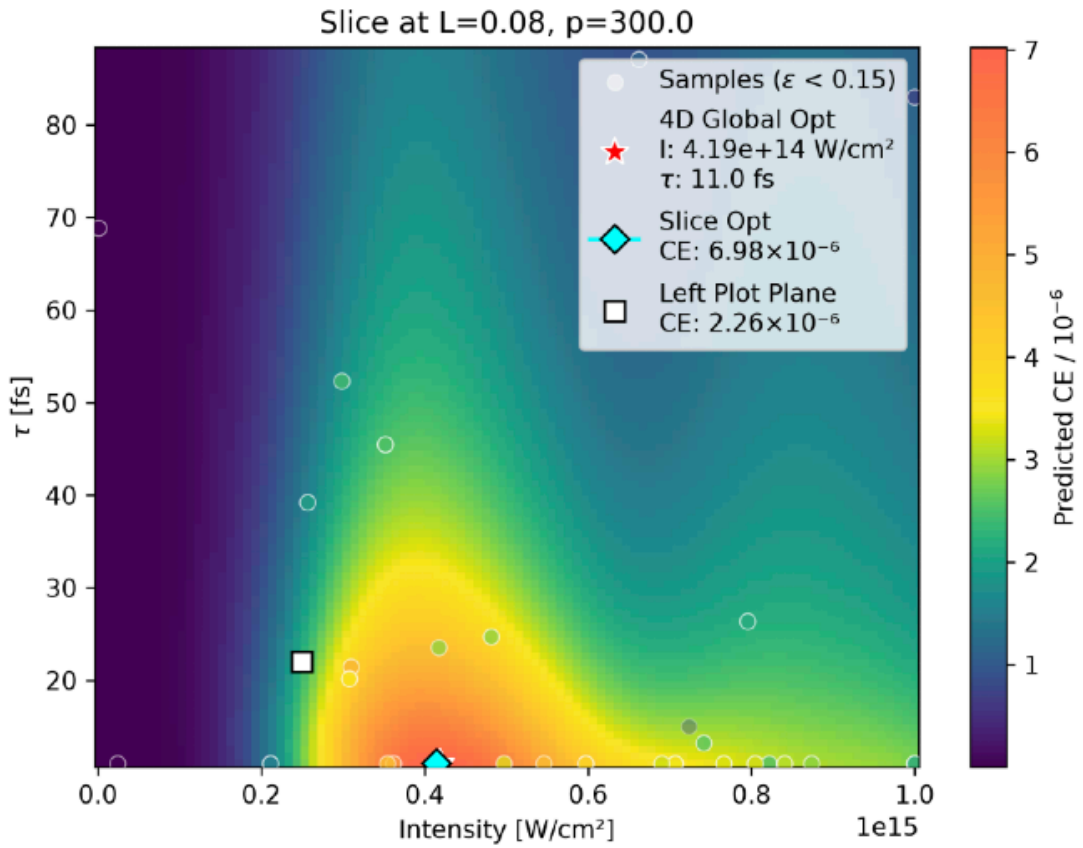
- We force the conversion efficiency to be close to the prescribed threshold
- As a result, the algorithm preferentially samples points near the estimated boundary of the high-efficiency region, allowing the region satisfying CE > CE_{target}



The optimization can be extended to 4D optimization

- There is also a prediction that high conversion efficiency for laser intensity I scales with duration τ as 1 / sqrt (τ)





Discussion

- Bayesian optimization (BO) can be a very strong tool in proposing new concepts
- It can be efficiently used for simulations, experiments…
- Physical knowledge can improve the performance
- In general, BO is essentially a statistical tool and the approach varies from problem to problem
- We used it to improve knowledge and performance of the attosecond light pulse generation with high-intensity lasers

Back up slides

In Bayesian optimization, Gaussian process serves as…

- **Probabilistic regression model**
- Learns a "cost" function from observed data
- For every input, predicts a **distribution of functions f(x) ~ GP(μ,k)**
- **Mean, μ(x)** → expected mean value of the function at point x
- **Kernel, k(x,x')** → Measure the similarity between input points x and x' and help Gaussian processes learn patterns from data
- **Uncertainty, σ(x)** → confidence in the prediction
- Uncertainty is typically **low near observed points** and **higher in unexplored regions**

Gaussian process

$$m_{post}(x*) = m(x*) + K(x*, X)[K(X, X) + \sigma_n^2 I]^{-1} \times (y - m(X))$$

$$\sigma_{post}^2(x*) = K(x*, x*) - K(x*, X)[K(X, X) + \sigma_n^2 I]^{-1} \times K(X, x*)$$

- $X=[x_1, x_2, \ldots, x_n]$ = points where simulations are run.
- $y=[y_1, y_2, \ldots, y_n]^T$ = simulation results.
- $x*$ = new point where you want to predict.
- $m(x*)$ = prior mean at the new point.
- $m(X)$ = prior means at observed points.
- $K(X,X)$ = covariance between all observed points, calculated using the kernel.
- $K(x*, X)$ = covariance between the new point and observed points.
- σ_n^2 = observation noise (often 0)

Kernels

Measure the similarity between input points and help Gaussian Processes learn patterns from data.

Gaussian process (GP) is a probabilistic model that approximates the cost function using a mean function to predict the expected value at any point and a covariance function (kernel) to describe correlations between different points in the input space. The kernel determines the covariance matrix between all evaluated points, which is essential for making predictions in GP regression. For the batch Bayesian optimization implemented in our work, we used the Matérn kernel[1], defined as

$$K(X,X') = \sigma_K^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\, r(X,X')\right)^{\nu} K_\nu \left(\sqrt{2\nu}\, r(X,X')\right), \tag{S1}$$

where K_ν is the modified Bessel function of the second kind, Γ denotes the gamma function, and $X = (d_\mathrm{u}, d_\mathrm{s}, n_\mathrm{p})$ and $X' = (d'_\mathrm{u}, d'_\mathrm{s}, n'_\mathrm{p})$ are points in the three-dimensional parameter space. The scaled distance $r(X,X')$ is given by

$$r(X,X') = \sqrt{\frac{(d_\mathrm{u} - d'_\mathrm{u})^2}{\ell_{d_\mathrm{u}}^2} + \frac{(d_\mathrm{s} - d'_\mathrm{s})^2}{\ell_{d_\mathrm{s}}^2} + \frac{(n_\mathrm{p} - n'_\mathrm{p})^2}{\ell_{n_\mathrm{p}}^2}}, \tag{S2}$$

with σ_K^2 the variance and ℓ_{d_u}, ℓ_{d_s}, ℓ_{n_p} the characteristic length scales associated with the three parameters. The smoothness of the kernel is controlled by ν. In our work, we set $\nu = 5/2$, which corresponds to a kernel whose sample functions are twice differentiable. The kernel hyperparameters σ_K^2, ℓ_{d_u}, ℓ_{d_s} and ℓ_{n_p} were chosen by maximizing the Gaussian-process log marginal likelihood, which is a standard way to select parameters so that the surrogate model captures the data without overfitting.

Betatrons: Different size of batches performs differently

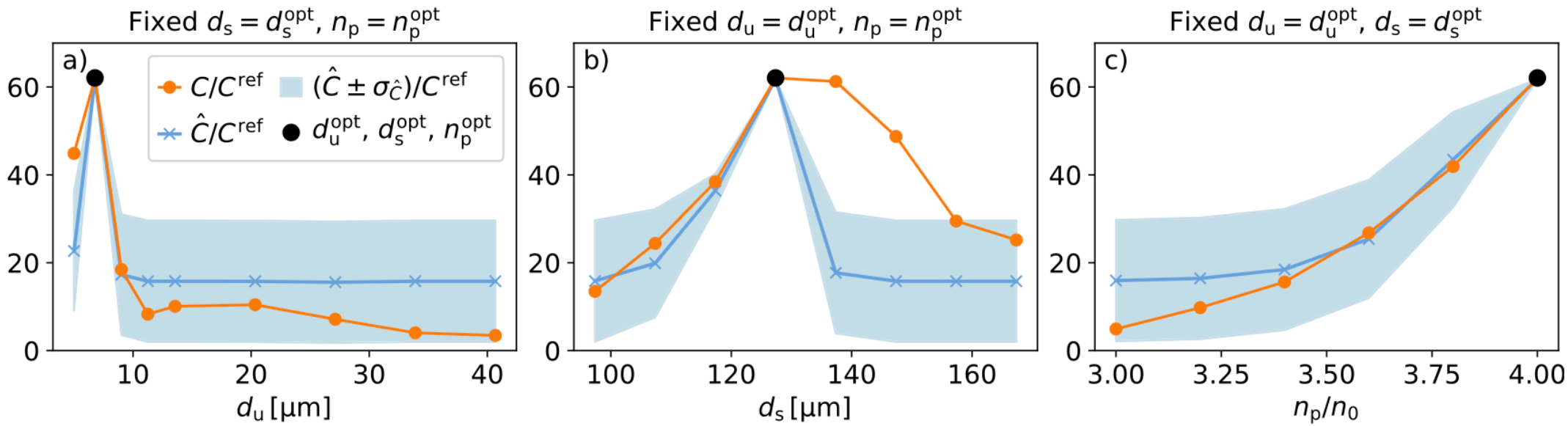


Fig. 3 | Values of the normalized cost function around the best-performing point found by BBO. The values are shown for **a)** varying d_u, fixed $d_s = d_s^{opt}$, $n_p = n_p^{opt}$, **b)** varying d_s, fixed $d_u = d_u^{opt}$, $n_p = n_p^{opt}$, and **c)** varying n_p/n_0, fixed $d_u = d_u^{opt}$, $d_s = d_s^{opt}$. The normalized cost function C/C^{ref} is shown with orange lines and circular markers at the evaluated points. The normalized model prediction $\widehat{C}/C^{ref}$ for $N = 8$ from the final (8th) iteration, is shown with blue lines with crosses at evaluated points. Colored shaded regions indicate the area within predicted standard deviation $\sigma_{\widehat{C}}$. The large black circle marks the best-performing case.

Betatrons: Spatial profiles

Kernel definition

Gaussian process (GP) is a probabilistic model that approximates the cost function using a mean function to predict the expected value at any point and a covariance function (kernel) to describe correlations between different points in the input space. The kernel determines the covariance matrix between all evaluated points, which is essential for making predictions in GP regression. For the batch Bayesian optimization implemented in our work, we used the Matérn kernel[1], defined as

$$K(X,X') = \sigma_K^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\, r(X,X')\right)^{\nu} K_\nu \left(\sqrt{2\nu}\, r(X,X')\right), \tag{S1}$$

where K_ν is the modified Bessel function of the second kind, Γ denotes the gamma function, and $X = (d_\mathrm{u}, d_\mathrm{s}, n_\mathrm{p})$ and $X' = (d'_\mathrm{u}, d'_\mathrm{s}, n'_\mathrm{p})$ are points in the three-dimensional parameter space. The scaled distance $r(X,X')$ is given by

$$r(X,X') = \sqrt{\frac{(d_\mathrm{u} - d'_\mathrm{u})^2}{\ell_{d_\mathrm{u}}^2} + \frac{(d_\mathrm{s} - d'_\mathrm{s})^2}{\ell_{d_\mathrm{s}}^2} + \frac{(n_\mathrm{p} - n'_\mathrm{p})^2}{\ell_{n_\mathrm{p}}^2}}, \tag{S2}$$

with σ_K^2 the variance and ℓ_{d_u}, ℓ_{d_s}, ℓ_{n_p} the characteristic length scales associated with the three parameters. The smoothness of the kernel is controlled by ν. In our work, we set $\nu = 5/2$, which corresponds to a kernel whose sample functions are twice differentiable. The kernel hyperparameters σ_K^2, ℓ_{d_u}, ℓ_{d_s} and ℓ_{n_p} were chosen by maximizing the Gaussian-process log marginal likelihood, which is a standard way to select parameters so that the surrogate model captures the data without overfitting.

Betatrons: Electron spectra and charge

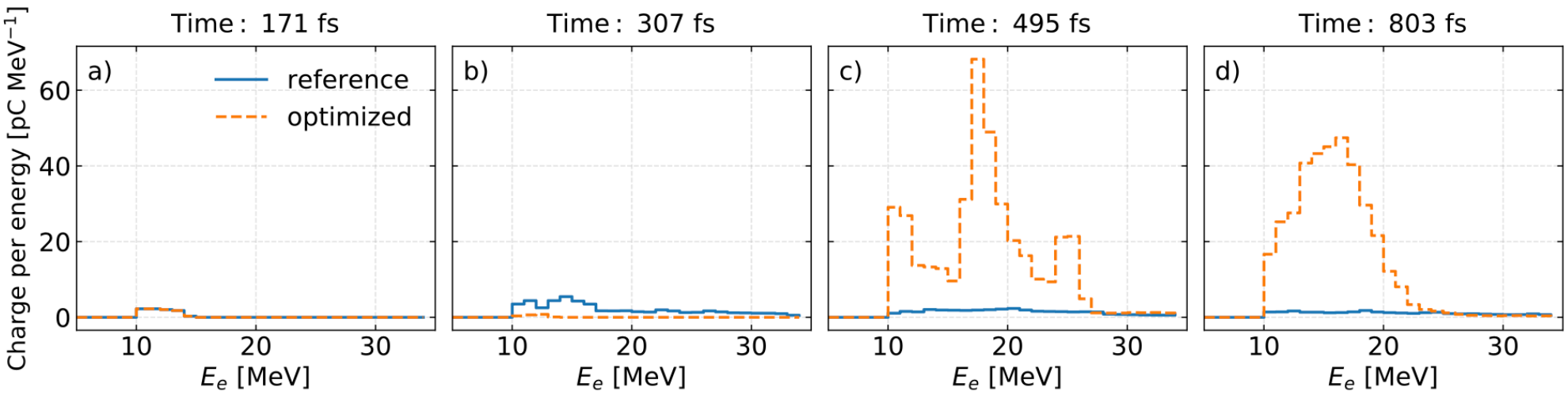
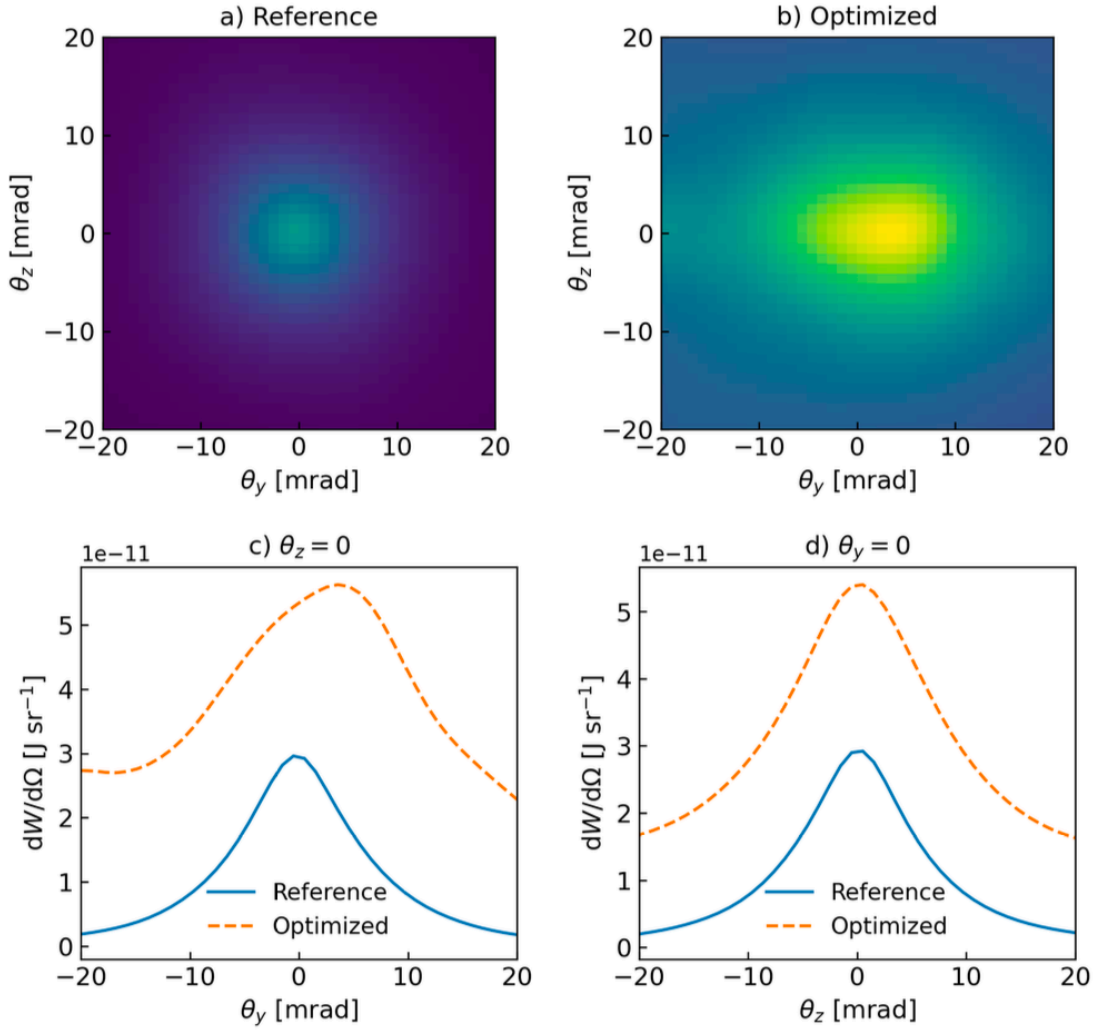


Figure S2. Energy spectra depicted at different simulation times a) 171 fs, b) 307 fs, c) 495 fs, d) 803 fs for the reference (blue solid) and the best-performing optimized (orange dashed) cases. The times shown above the panels are measured from when the pulse centre reaches the simulation box entrance.

	171 fs	307 fs	495 fs	803 fs
reference	9 pC	52 pC	36 pC	28 pC
optimized	9 pC	2 pC	397 pC	370 pC

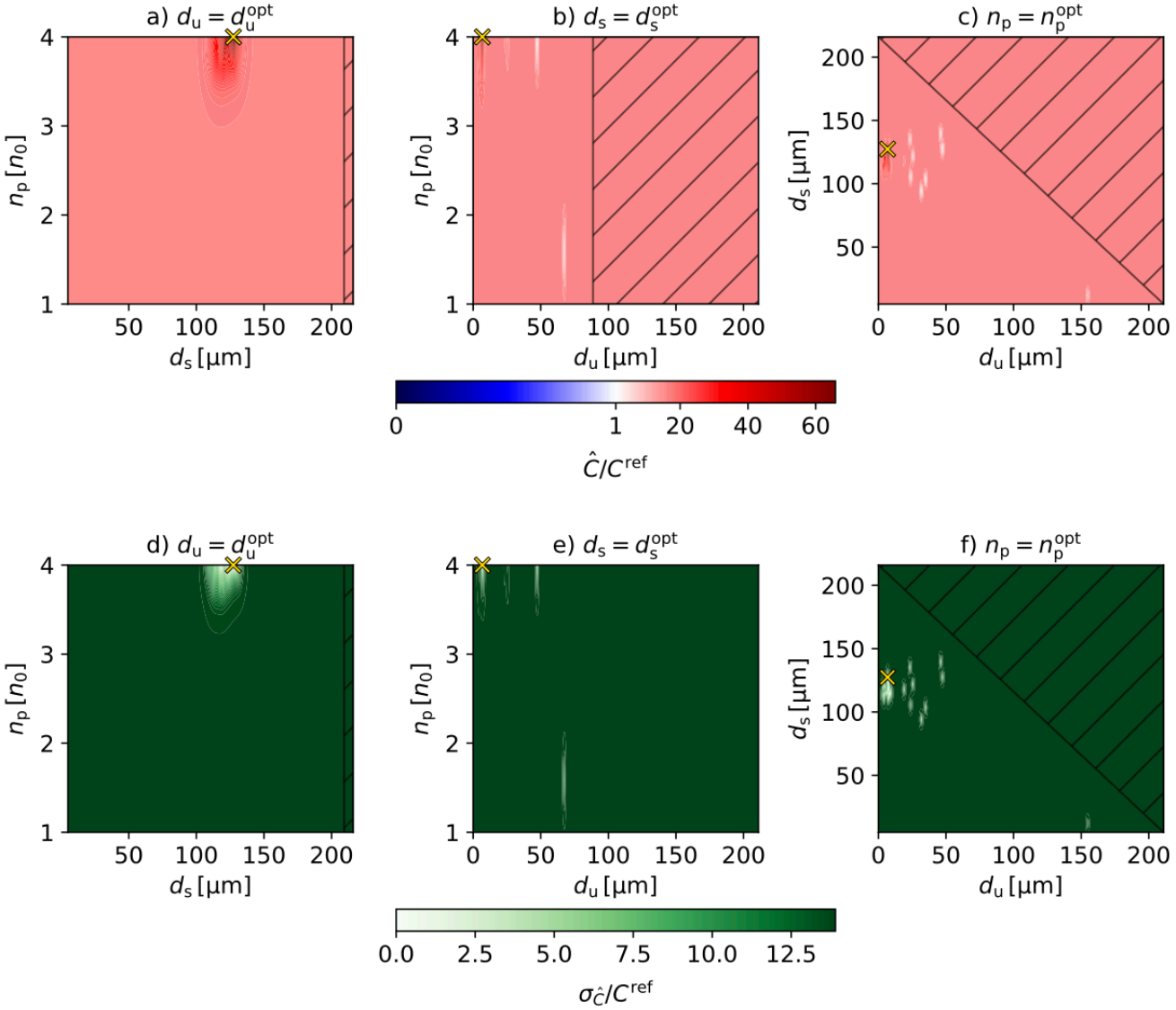
Table S1. Total charge of electrons with energies $E_e \geq 10$ MeV at different simulation times for the reference and optimized cases, corresponding to the spectra in Fig. S2.

Betatrons: Radiation spectra



Maslarova, D., Hansson, A., Luo, M., Horný, V., Ferri, J., Pusztai, I., & Fülöp, T. (2026). Batch Bayesian optimization of attosecond betatron pulses from laser wakefield acceleration. *Communications Physics*, *9*(1), 92.

Betatrons: Model at the best performing point



HHG: Acquisition function Expected improvement

$$a^i_{\mathrm{EI}}(\mathbf{x}) = (\mu^i_{\mathrm{CE}}(\mathbf{x}) - \mathrm{CE}^i_{\mathrm{best}})\Phi(\gamma^i) + \sigma^i_{\mathrm{CE}}(\mathbf{x})\phi(\gamma^i),$$

$$\gamma^i = \frac{\mu^i_{\mathrm{CE}}(\mathbf{x}) - \mathrm{CE}^i_{\mathrm{best}}}{\sigma^i_{\mathrm{CE}}(\mathbf{x})}.$$

Here, γ^i measures how far the predicted mean lies above or below the best conversion efficiency observed so far, $\mathrm{CE}^i_{\mathrm{best}}$, scaled by the GP posterior standard deviation at the ith iteration. The functions Φ and ϕ represent the cumulative distribution and probability density functions of the standard normal distribution, respectively. To prevent numerical underflow in regions of high certainty or poor performance, the numerically stable logarithmic variant, LogEI ($a^i_{\mathrm{LogEI}} = \log a^i_{\mathrm{EI}}$), was employed [23] via the BoTorch library.

HHG: Acquisition function Straddle

$$a^i_{\mathrm{Str}}(\mathbf{x}) = \beta\sigma^i_{\mathrm{CE}}(\mathbf{x}) - \left|\mu^i_{\mathrm{CE}}(\mathbf{x}) - \mathrm{CE}^i_{\mathrm{target}}\right|,$$

HHG: cutoffs

$10^{14} – 3\times10^{14}$ W/cm^2

Gas	Typical useful HHG region	Approx. cutoff scale
Xe	~15–50 eV	~30–60 eV
Kr	~20–70 eV	~40–80 eV
Ar	~20–100 eV	~40–100 eV
Ne	~30–150+ eV	~80–150+ eV
He	~40–200+ eV	~100–200+ eV

Betatrons: PIC simulations

- The laser has a wavelength 800 nm
- Peak intensity 2.53×10^{19} W cm^{-2}
- Normalized vector potential $a_0 = 3.44$
- Pulse energy 37 mJ
- Gaussian spatial and temporal profile
- Linear polarization along the y-axis.
- Pulse is focused at the plasma entrance to a spot size of 3.8 μm
- Duration 8.3 fs (full width at half maximum (FWHM) of the intensity
- The plasma density rises linearly from zero to $1.2n_0$, with $n_0 = 2\times10^{19}$ cm^{-3}
- Box has dimensions Lx × Ly × Lz = 40.96 μm × 51.2 μm × 51.2 μm.
- A moving-window technique was applied, with a velocity of c in the x direction.
- The simulation grid was discretized with a resolution of Δx × Δy × Δz = 17.8 nm × 133.3 nm × 133.3 nm.
- The simulation timestep was 57 as.
- One electron macroparticle per cell was used, with regular position initialization.
- The Maxwell equations were solved using the Lehe scheme in combination with a binomial filter.
- One macroparticle represents a set of electrons with a corresponding statistical weight.

High Harmonic Generation can be described in 3 steps

