

Supplementary Information for “Batch Bayesian optimization of attosecond betatron pulses from laser wakefield acceleration”

Dominika Maslarova^{1*}, Albert Hansson¹, Mufei Luo^{1,2}, Vojtěch Horný^{1,3}, Julien Ferri¹, Istvan Pusztai¹, and Tünde Fülöp¹

¹Department of Physics, Chalmers University of Technology, Göteborg, SE-41296, Sweden

²Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK

³Extreme Light Infrastructure - Nuclear Physics, Horia Hulubei National Institute for Physics and Nuclear Engineering, 30 Reactorului Street, RO-077125 Bucharest-Magurele, Romania

*E-mail: dommas@chalmers.se

Supplementary Note 1: On the optimization process

In this note, details of the batch Bayesian optimization (BBO) process are discussed. First, we show the definition of Matérn kernel used to calculate correlations between the input points. Subsequently, we show the mean values and uncertainty of the predicted model around the best-performing point found in our search.

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 kernel¹, defined as

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

where K_ν is the modified Bessel function of the second kind, Γ denotes the gamma function, and $X = (d_u, d_s, n_p)$ and $X' = (d'_u, d'_s, n'_p)$ are points in the three-dimensional parameter space. The scaled distance $r(X, X')$ is given by

$$r(X, X') = \sqrt{\frac{(d_u - d'_u)^2}{\ell_{d_u}^2} + \frac{(d_s - d'_s)^2}{\ell_{d_s}^2} + \frac{(n_p - n'_p)^2}{\ell_{n_p}^2}}, \quad (\text{S2})$$

with σ_K^2 the variance and $\ell_{d_u}, \ell_{d_s}, \ell_{n_p}$ 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 $\sigma_K^2, \ell_{d_u}, \ell_{d_s}$ 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.

Evaluation of the optimization model at the best-performing point

Here, we analyze the behaviour of the surrogate cost function \hat{C} model, which estimates the performance of the cost function C , defined in Eq. (1) of the main text. Figures S1a)-c) show 2D slices of the surrogate model, which intersect each other at the best-performing point found during the whole optimization process across all the iterations and all the batches: $d_u = d_u^{\text{opt}} = 6.8 \mu\text{m}$, $d_s = d_s^{\text{opt}} = 127.4 \mu\text{m}$ and $n_p = n_p^{\text{opt}} = 4.0 n_0$. Here, d_u is the distance between the injection of the electrons and the start of the density spike, d_s is the length of the spike, and its peak density is n_p . This model is presented for the corresponding batch size where it was found, i.e. $N = 8$, at the last (8th) iteration performed. The model shows strong improvement in predictions around the best-performing point. As depicted in Fig. 3 of the main manuscript, the estimates around this point predict the evaluation of C very well within uncertainty. Other areas show almost uniform improvement, with \hat{C}/C^{ref} roughly below 20. Some small areas around certain points show performance similar to the reference case, $\hat{C}/C^{\text{ref}} = 1$. These points correspond to the values evaluated during the BBO exploration.

Figs. S1d)-f) show a corresponding uncertainty (standard deviation) of the model at the specific points. It can be seen that the model is very confident in the vicinity of the best-performing point and in narrow regions around previously evaluated

points. However, the uncertainty remains large across most of the domain, indicating insufficient exploration. This behaviour was already apparent during the optimization: the search repeatedly sampled similar regions, and the cost values began to converge for all batch sizes after several iterations (see Fig. 2 of the main text). The exploration could possibly be boosted by using different acquisition and kernel functions, or possibly tuning parameters of these functions. A systematic investigation of such strategies lies beyond the scope of the present work.

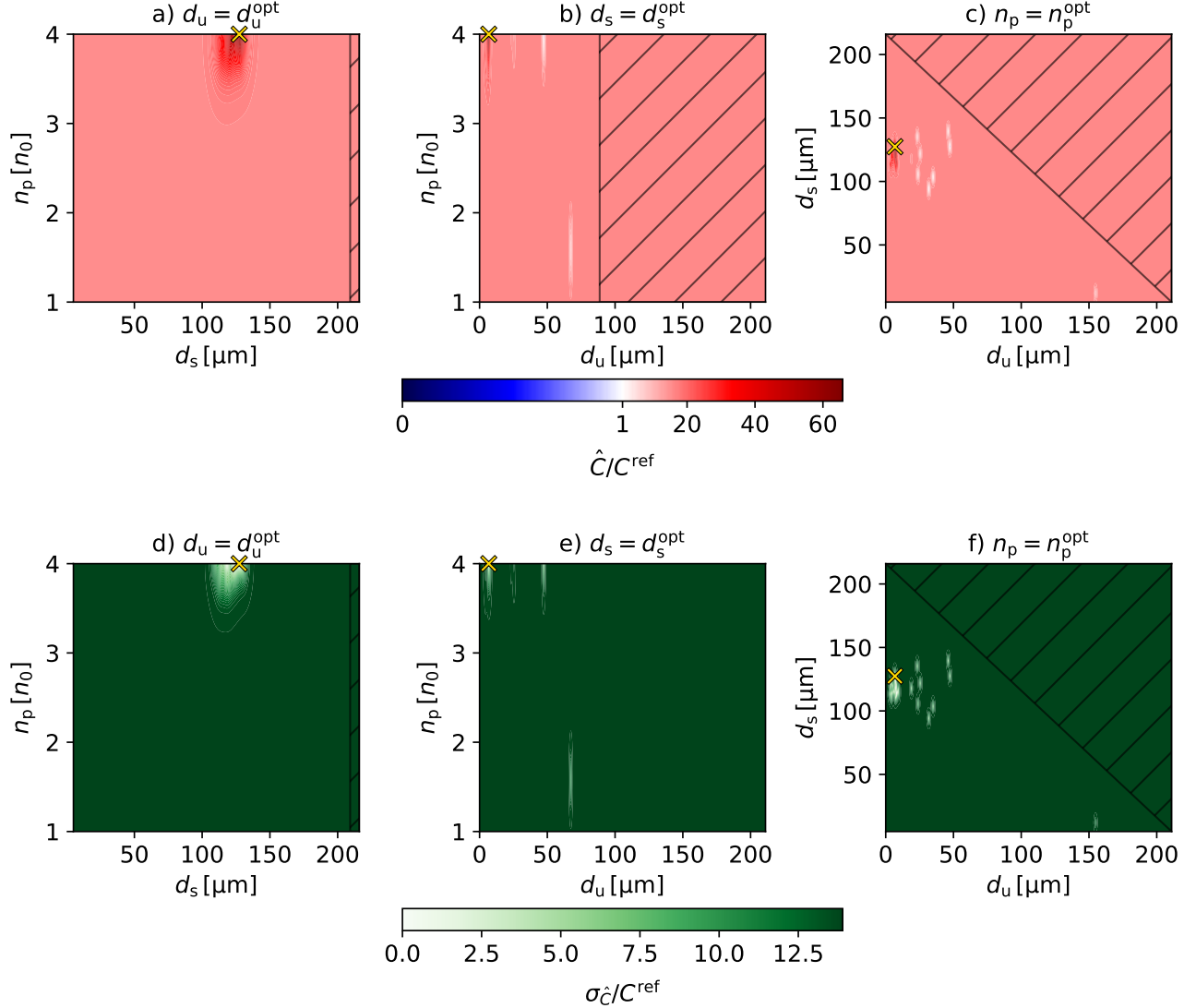


Figure S1. 2D maps of the surrogate model at the best-performing point. a-c) Estimated values from the cost function model \hat{C} : a) d_u^{opt} , b) d_s^{opt} , c) n_p^{opt} . Red/blue shades represent an improvement/reduction of performance compared to the reference case C^{ref} . d-f): Values of predicted uncertainty $\sigma_{\hat{C}}$ corresponding to a)-c): d) d_u^{opt} , e) d_s^{opt} , f) n_p^{opt} . In the hashed areas, the search was forbidden due to the restriction in Eq. (5) in the main text. The yellow cross indicates the best-performing point.

Supplementary Note 2: On the properties of the electron beam and betatron radiation

In this note, we show properties of the electron beams and the electromagnetic radiation that were not included in the main text, specifically electron energy and charge, and angular spectrum of the radiation.

Energy spectrum and charge evolution of the electron beam

To investigate how the evolution of the electron properties influences the final radiation, we present the energy spectrum at different simulation times for both the reference case and the best-performing (optimized) case, from the 7th iteration of BBO with $N = 8$. The times correspond to the density evolution of the optimized case in Fig. 5 of the main manuscript. Fig. S2 depicts the energy spectrum of electrons participating in the radiation calculation (electrons with energies $E_e \geq 10$ MeV) at different simulation times. Corresponding evolution of their total charge is summarized in Tab. S1.

At 171 fs (Fig. S2a), the spectrum corresponds to the beam from the first injection process, and the reference and optimized cases are the same. Later, at 307 fs (Fig. S2b), the charge per energy rapidly decreases in the optimized case, while it grows as the electrons are accelerating in the reference case. The reduction in charge is a direct consequence of the bubble contraction induced by the upward density gradient within the spike, which displaces electrons from the accelerating phase and stops the acceleration process.

At 495 fs (Fig. S2c), a strong second injection occurs in the optimized case as the laser propagates through the density down-ramp. At the end of the simulation (Fig. S2d), the charge distribution is reshaped. For the optimized case, no significant additional acceleration is observed, as the main laser-driven wakefield has weakened and the wake induced by the injected electron beam dominates. The charge per energy of the reference case is significantly lower. As also shown in Tab. S1, at the end of the simulation the final charge in the best-performing case is more than $10\times$ higher, which resulted in a rapid increase of the radiation.

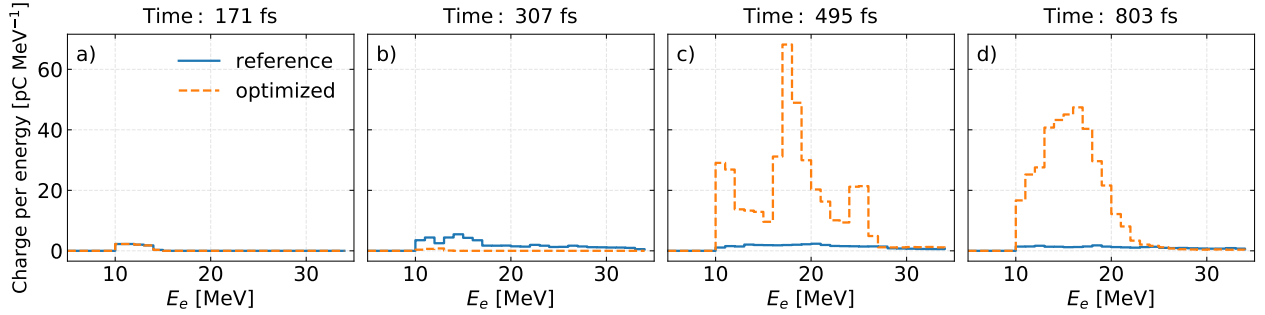


Figure S2. Energy spectra depicted at different simulation times a) 171 fs, b) 307 fs, c) 495 fs, d) 803 fs for the reference (blue solid) and the best-performing optimized (orange dashed) cases. The times shown above the panels are measured from when the pulse centre reaches the simulation box entrance.

	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.

Angular spectrum of the radiation

In addition to the increase in radiation energy, the angular spread of the betatron beam for the best-performing optimized case in BBO (batch size $N = 8$, iteration 7) shows a wider angular spread compared to the reference, no-spike case. The root-mean-square angular spread values in both directions increased as follows: from $\sigma_{\theta_y} = 10.6$ mrad to $\sigma_{\theta_y} = 15.2$ mrad, and from $\sigma_{\theta_z} = 11.0$ mrad to $\sigma_{\theta_z} = 14.4$ mrad. The corresponding angular profiles are shown in Fig. S3.

Influence of multiple wakefield periods and a density down-ramp at the plasma exit on the radiation profile

In the main text, we focused on the first dominant betatron pulse emitted from the plasma bubble. In principle, additional electrons can be trapped in subsequent wakefield periods and contribute to the overall betatron radiation. To investigate the contribution of these later wakefield periods, we performed an additional simulation of the reference case using an extended simulation window that captures eight wakefield periods. Specifically, the longitudinal size of the simulation domain was increased from $L_x = 40.96 \mu\text{m}$ to $L_x = 81.92 \mu\text{m}$.

Additionally, to better approximate experimentally realistic plasma conditions, we performed an additional simulation using an extended longitudinal window that includes a plasma exit region. In contrast to the main-text simulations, where the initial

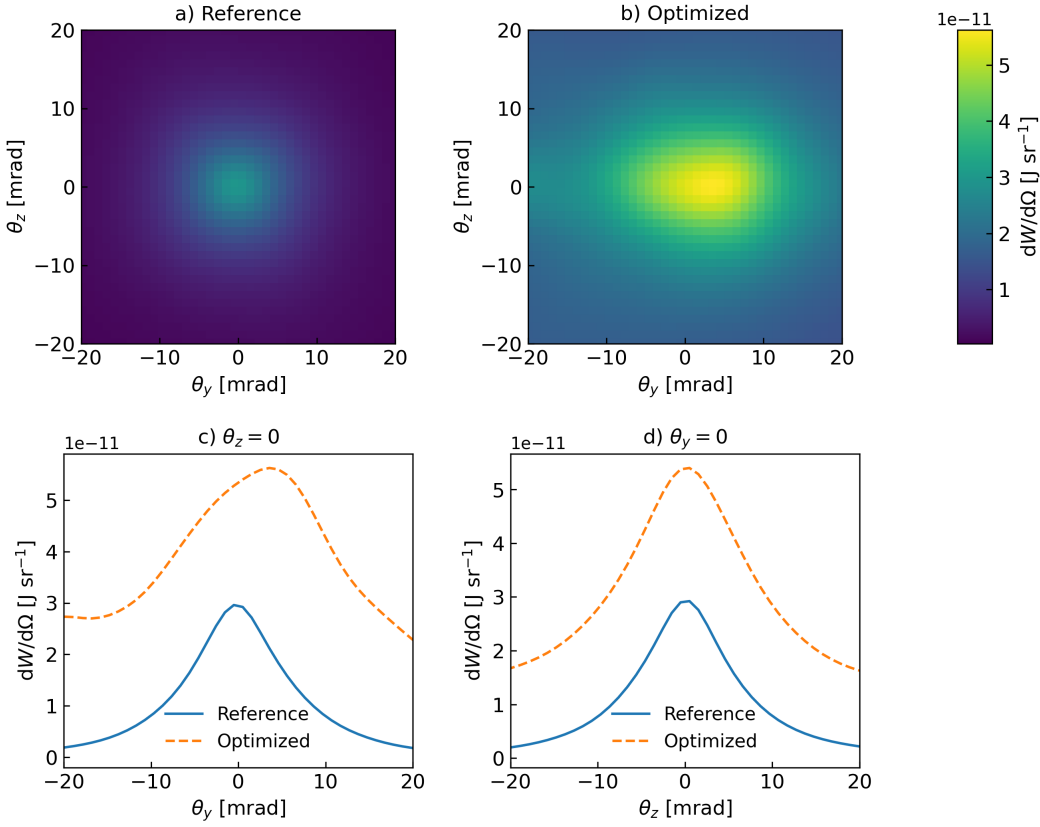


Figure S3. Angular spectra of the betatron radiation. Spectra are shown for a) the reference case b) best-performing optimized case. θ_y and θ_z represent the divergences in y-direction and z-direction, respectively. $dW/d\Omega$ is the radiation energy W per solid angle Ω . On-axis lineouts of the spectra are shown at c) $\theta_z = 0$ and d) $\theta_y = 0$ for the reference case (blue solid) and the optimized case (orange dashed).

plasma injection gradient is followed by a 216 μm -long uniform density plateau, the plasma profile here consists of the same injection gradient followed by a 176 μm plateau and a subsequent 40 μm linear down-ramp, over which the density decreases from n_0 to zero.

The corresponding results are shown in Fig. S4. As shown in Fig. S4c, electrons originating from each wakefield period generate distinct radiation bursts, leading to an overall increase in the energy spectrum presented in Fig. S4a. While extending the simulation window to include additional wakefield periods beyond eight would likely result in further radiation emission, the contribution of successive pulses is observed to decrease progressively, with the exception of the third pulse. This trend can be expected to continue for later periods. Including the density down-ramp slightly modifies the energy distribution of the emitted radiation, primarily shifting photons toward higher energies, while the overall spectral and temporal behavior remains largely unchanged.

Supplementary Note 3: On numerical convergence

We verified the numerical convergence of Smilei particle-in-cell simulations by varying the longitudinal spatial resolution (45 versus 50 cells per laser wavelength), the transverse spatial resolution (6 versus 8 cells per laser wavelength), and the number of macroparticles per cell ($ppc = 1, 8, 27$). Increasing the longitudinal resolution changed the resulting spectrum negligibly and the cost function increased only by 0.3%. Increasing the transverse resolution slightly shifted the electron beam position and the timing of the betatron pulse due to a small change in the injection dynamics. However, the corresponding difference in the cost function still remained as low as 1.1%.

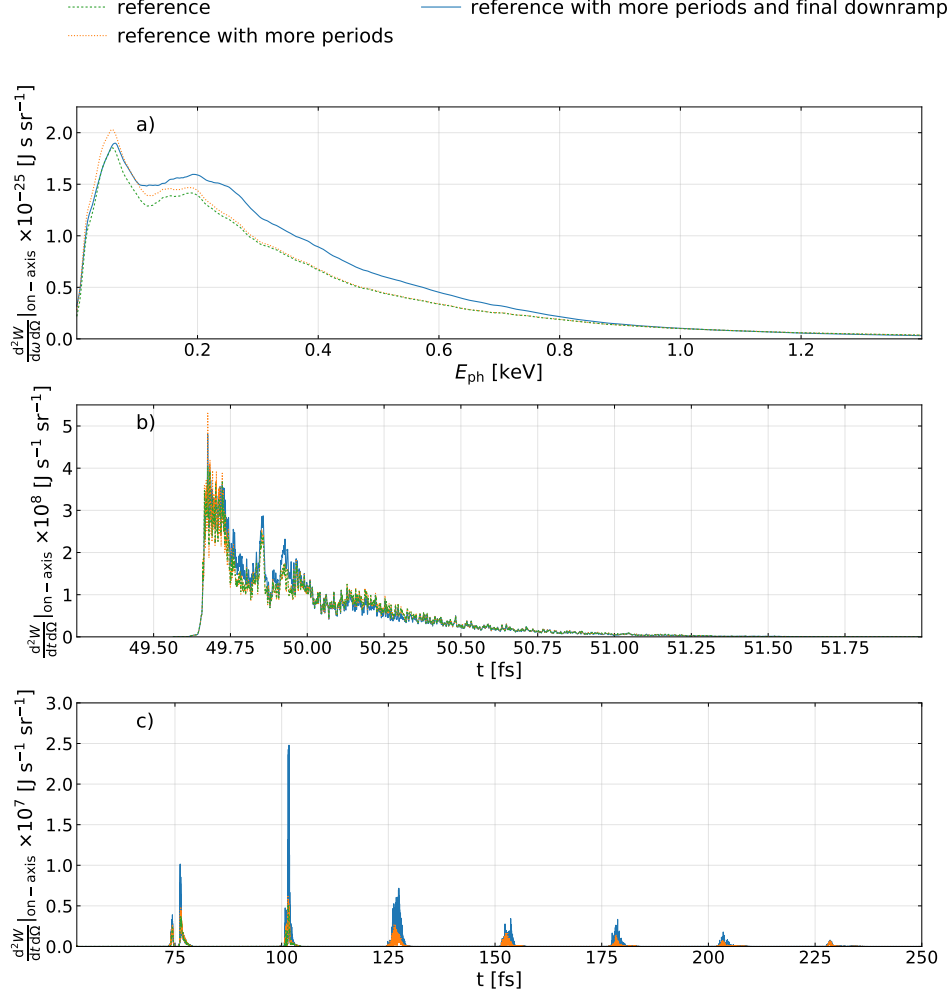


Figure S4. a) Dependence of the radiated energy W per photon angular frequency ω per solid angle Ω on photon energy E_{ph} . b) Temporal dependence of W per observer time t per Ω in the observer time window $t \in [49.25 \text{ fs}, 52 \text{ fs}]$. c) Temporal dependence of W per observer time t per Ω in the observer time window $t \in [52 \text{ fs}, 250 \text{ fs}]$. The spectra are shown for the reference case without the density spike with the simulation setup from the main text (green dashed), for the reference case with the longer particle-in-cell simulation window (orange dotted), and for the reference case with the longer particle-in-cell simulation window and additional density gradient at the end of the acceleration (blue solid).

Increasing the number of macroparticles from $ppc = 1$ to $ppc = 8$ and $ppc = 27$ produced some reductions in the cost function (1.2% and 1.7%, respectively). The corresponding comparison of the spectra is shown in Fig. S5. Note that, in order to eliminate stochastic effects in the BBO process, we used a regular macroparticle distribution, which restricts the number of macroparticles to the third power of integers (corresponding to the three simulation dimensions). During initial BBO runs, some simulations with higher ppc led to several million macroparticles being tracked (when applying an energy threshold of $E_e \geq 10 \text{ MeV}/c$ for tracking). The processing of radiation from these macroparticles was extremely demanding on available RAM and runtime increased dramatically (up to $17\times$ longer for Smilei and $> 30\times$ longer for FIKA). Since the cost function integrates over the full spectrum and is robust against noise, and because all tested spatial resolutions produced nearly identical relative performance, we selected 45 cells per laser wavelength longitudinally, 6 cells per laser wavelength transversely, and $ppc = 1$ as the optimal balance between numerical stability, computational feasibility, and reproducibility in our BBO runs.

Additional tests also confirmed minimal sensitivity to the transverse window size. Increasing the transverse window size from $51.2 \mu\text{m}$ to $64 \mu\text{m}$ produced no visible changes in the electron dynamics or radiation emission, and the cost function changed by only 0.02%, indicating that any numerical aperturing did not influence the results.

As an additional numerical consistency check, over the total simulation time, the total energy within the particle-in-cell

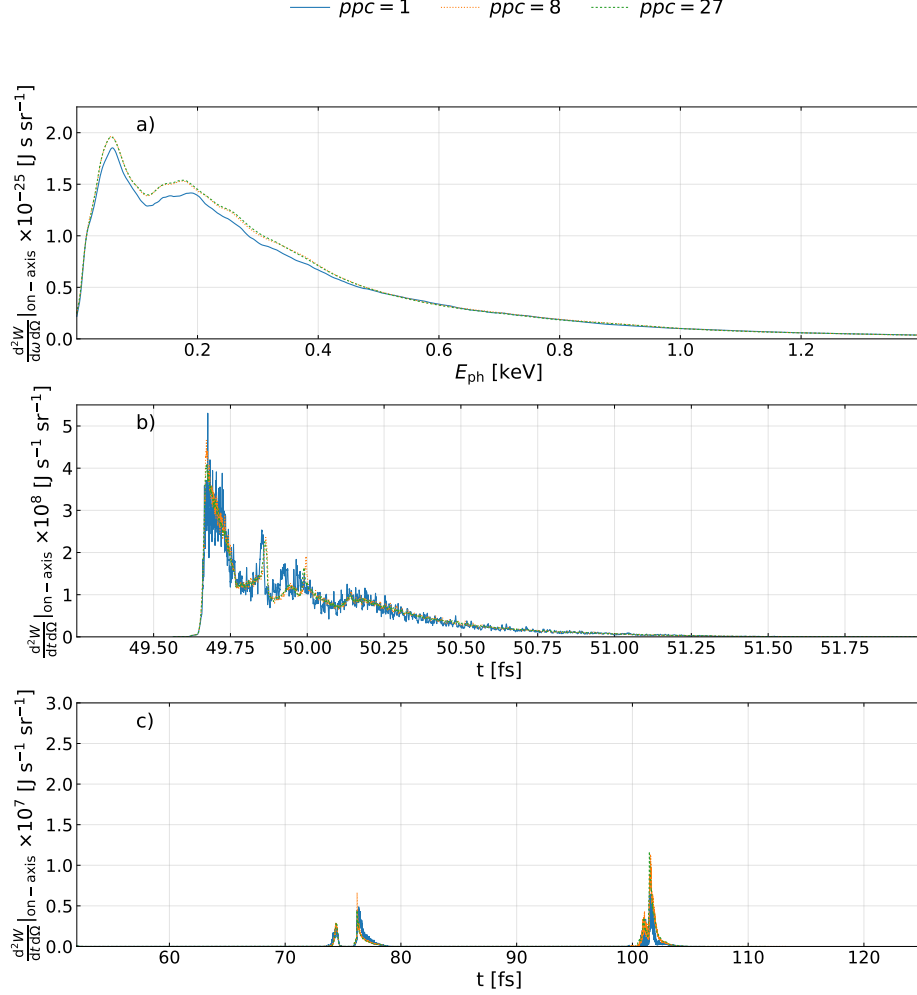


Figure S5. a) Dependence of the radiated energy W per photon angular frequency ω per solid angle Ω on photon energy E_{ph} . b) Temporal dependence of W per observer time t per Ω in the observer time window $t \in [49.25 \text{ fs}, 52 \text{ fs}]$. c) Temporal dependence of W per observer time t per Ω in the observer time window $t \in [52 \text{ fs}, 125 \text{ fs}]$. The spectra are shown for the reference case without the density spike for different numbers of macroparticles per cell: a) $ppc = 1$ (blue solid), b) $ppc = 8$ (orange dotted), c) $ppc = 27$ (green dashed).

simulation domain decreased by 14.1% for the reference case without the spike, and by 39.8% for the best-performing optimization case, consistent with energy carried out through the open boundaries. The number of electrons in the box decreased by 3.4% for the reference case and by 71.6% for the best-performing case, corresponding to particles leaving the computational domain throughout the interaction.

As a next step, we also varied the longitudinal-momentum threshold used to select electrons from Smilei for the radiation calculation in FIKA. It was found that thresholds of 3 and 5 MeV/ c have cost-function differences of $\leq 0.5\%$ compared to 10 MeV/ c . This is because low-energy particles contribute only marginally to the spectrum gain. However, they consume a lot of computational storage. Therefore, we picked only macroparticles with $\geq 10 \text{ MeV}/c$ from Smilei for the radiation calculation in FIKA.

For the FIKA radiation calculation, we varied the temporal step as $\Delta t_{\text{FIKA}} = 25, 15$, and $5 \Delta t_{\text{SMILEI}}$, where Δt_{SMILEI} is the Smilei timestep. All spectra were visually indistinguishable. The computational cost increased strongly for the smallest timestep: the $\Delta t_{\text{FIKA}} = 5 \Delta t_{\text{SMILEI}}$ case required 233% more FIKA runtime and 36% more Smilei processing time compared to $\Delta t_{\text{FIKA}} = 15 \Delta t_{\text{SMILEI}}$, while providing no significant improvement in the spectral output (cost function changed only by 0.6%). In contrast, increasing the timestep to $25 \Delta t_{\text{SMILEI}}$ reduced the computational time only slightly (7% for Smilei, $< 1\%$ change for FIKA) compared to $\Delta t_{\text{FIKA}} = 15 \Delta t_{\text{SMILEI}}$. The cost function decreased only by 0.4%. Based on these results, we selected

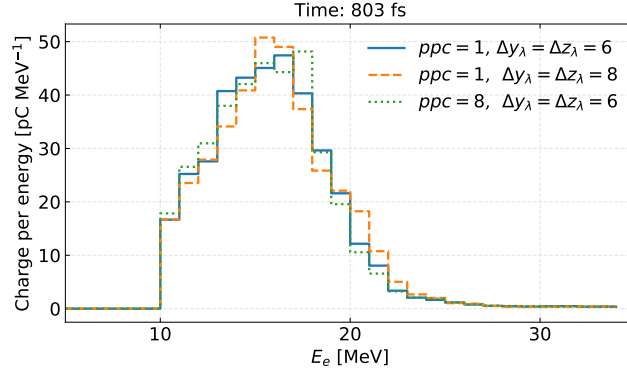


Figure S6. Energy spectra comparison for different numerical setups. The spectra correspond to the best-performing case at the time of 803 fs. Following cases are depicted: $ppc = 1$ with the transverse resolution (Δy_λ in the y direction and Δz_λ in the z direction) of 6 cells per laser wavelength (blue solid), $ppc = 1$ with the transverse resolution of 8 cells per laser wavelength (orange dashed), and $ppc = 8$ with a transverse resolution of 6 cells per laser wavelength (green dotted).

$\Delta t_{\text{FIKA}} = 15 \Delta t_{\text{SMILEI}}$ as the optimal compromise between accuracy and efficiency for BBO runs: it achieves convergence within $< 1\%$ compared to $\Delta t_{\text{FIKA}} = 5 \Delta t_{\text{SMILEI}}$, while avoiding the large computational load.

As the number of macroparticles per cell was quite low in our optimization, we additionally verified the numerical convergence of the final best-performing run. The electron energy spectrum in Fig. S6 corresponds to a later simulation time shown in Fig. S2d. In one of the studied cases, we increased the transverse resolution to 8 cells per laser wavelength, and in the second case, we increased the number of macroparticles per cell to $ppc = 8$, while keeping the other parameters unchanged. Fig. S6 shows that there are only slight differences in the energy distribution functions. The total charge of the depicted spectra is as follows: 370 pC for the main-text case ($ppc = 1$, transverse resolution of 6 cells per laser wavelength), and 372 pC for both other cases, $ppc = 1$ with 8 cells per laser wavelength, and $ppc = 8$ with 6 cells per laser wavelength, which indicates sufficient numerical convergence.

Supplementary References

1. Matérn, B. *Spatial Variation* (Springer, New York, 1960).