

Batch Bayesian optimization of attosecond betatron pulses from laser wakefield acceleration

Corresponding Author: Dr Dominika Maslarova

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The manuscript titled "Batch Bayesian optimization of attosecond betatron pulses from laser wakefield acceleration" presents an optimisation of betatron radiated power through tailoring of a density spike added to increase the amplitude of electron oscillations with an LWFA. The density profiles are determined by three parameters with their effect on the betatron beam assessed using 3D PIC simulations.

The work builds on previous publications by the group (references 5&6) which described the regime of operation. However, the work is novel due to the exploration of optimizing the x-ray source while maintaining a sub-femtosecond pulse length. In my opinion, the results are of moderate importance to the sub-field, containing a small advance on previous results and exploration of a novel betatron oscillation enhancement mechanism.

However, I am not convinced that the paper is fully technically sound or that it provides strong evidence for its conclusions. Therefore, I do not recommend publication of this article in its current form.

My main concerns are:

1) The paper frequently refers to an increase in "photon number", "photon yield" and "photon gain". However, no photon numbers are given, and the optimisation is of the on-axis radiated power per unit solid angle (which is incorrectly called 'power spectrum').

2) The conclusion states "We conclude that BBO successfully identified the region of betatron enhancement with respect to photon gain and located a region in parameter space where the photon gain increased more than threefold compared to the reference value. The analysis of the PIC simulations revealed that the enhancement was driven by a boost in the electron oscillation amplitude occurring behind the bubble, a mechanism that has not been previously investigated in this context.". This is in-fact only shown for the case of an 80-micron wide density peak, but this is not the case found by the BBO (5 micron). Are the mechanisms the same? Why is there no enhancement for intermediate scales (20 microns)? The conclusion does not reflect the evidence.

3) The authors communicate an over-confidence that the BBO has found the true global optimum. The failure of an algorithm to improve on a previous optimum in the 4th iteration is taken to mean 'convergence' but this is not strong evidence of having found a global optimum. This is a general point on multi-dimensional optimisation. Supplementary note 2, figure 1, shows 2D slices through the 3D model predictions. These plots are also described as evidence of global optimisation as there is only one region of high performance, but these can never reveal the full picture. The author's statement is even contradicted by their own figure (2b in the main manuscript) which shows a region of increased performance which is not seen in the 2D slices. Overall, the BBO was able to optimise performance compared to the reference case but any claim that this is a global optimum is not supportable.

4) In the description of the betatron radiation it is observed that the oscillation amplitude typically decays as the electron beam accelerates. This is correct; however, I infer from the choice of references that this effect is incorrectly attributed to radiative damping. The scaling given is a purely adiabatic effect due to the changing oscillating frequency as the electrons, as described for example in A. G. R. Thomas; "Scalings for radiation from plasma bubbles." Phys. Plasmas 1 May 2010; 17 (5): 056708. Was radiative dampening included in the PIC simulations? It does not seem to be relevant to these results.

5) The title is inaccurate as 100 attoseconds is hardly an 'attosecond pulse'. This form of title is also used in preceding publications, but it is clearly misleading.

In addition to these comments, I have some further minor comments on the paper.

6) The paper states "This configuration was used also by Ferri et al.⁶, to demonstrate the generation of an attosecond betatron pulse from a down-ramp injection in LWFA.". Please make it clear that these results are numerical and not an experimental demonstration.

7) "This approach also enabled close monitoring of the optimization process. This is particularly relevant for strongly nonlinear processes, where it is difficult to a priori know that limiting the search space in a specific way will not lead to solutions of significantly different character, which are physically uninteresting."
The meaning of this sentence is unclear – please revise.

8) The meaning of L_g is not defined in the text, but from the numbers I take it as the position of the density peak for injection. Could the authors clarify this in the text or the figure.

9) The abstract should state that the 350% increase in radiated power is relative to a constant density case.

10) Betatron radiation calculations – how are macroparticle contributions combined? Each electron presumably radiates incoherently and therefore they should be weighted as the square root of macroparticle charge. Please clarify.

11) The Bayesian optimisation methodology is not clearly explained in the main paper. To aid understanding of one of the main components of the paper this should be more clearly explained in the main text. In particular: what kernels and acquisition functions were used? How were the batch of points selected to ensure they are not too close to each other? How are the N_{ran} points used in the optimisation? Does that just refer to the 0th iteration? The supplementary note 1 makes some of this clearer but the main text is confusing. Also, adding in additional random points after the initial set does not make sense to me. As they do not appear to depend on the GP, they could also be run in the 0th iteration and then the model would have more information in the early iterations.

Reviewer #2

(Remarks to the Author)

Review of "Batch Bayesian optimization of attosecond betatron pulses from laser wakefield acceleration"

Dear Editor,

Thank you for the invitation to review the manuscript titled "Batch Bayesian optimization of attosecond betatron pulses from laser wakefield acceleration". I would like to clarify that my review focuses on the machine learning and Bayesian optimization aspects of the manuscript. While I have experience in these areas, I do not have a background in plasma physics. Therefore, I am not able to evaluate the correctness or novelty of plasma physics modeling and experimental components. My comments are limited to the methodology and application of Bayesian optimization presented in the paper. Manuscript presents an application of batch Bayesian optimization (BBO) for the enhancement of attosecond betatron radiation in a laser wakefield acceleration (LWFA). The authors aim to simultaneously maximize the on-axis spectral peak power PS_{peak} and minimize the pulse duration $\tau\beta$ through cost function: $C(PS_{\text{peak}}, \tau\beta) = -PS_{\text{peak}}/\tau\beta$. BBO implementation was done using a Gaussian Process (GP) surrogate model with the Matern 5/2 kernel and standard acquisition strategies including Expected Improvement (EI), Local Penalization (LP), and the Fantasizer method. Exploration mechanism was enhanced with Sobol and uniform random sampling to cover the space more effectively and to prevent surrogate overfitting. Description of strategy of point sampling was described very nicely in Table 1.

Overall, the topic and outcomes of this work will be of interest to researchers working at the intersection of laser-plasma interactions and machine learning-based optimization. The BBO methodology is described in a detailed and structured manner, demonstrating interesting implementation choices; however, there are still some suggestions that could improve the quality of this paper:

1. Uncertainty Discussion: since BBO depends on using uncertainty estimates to guide exploration vs exploitation, the manuscript lacks discussion or visualization of the GP model's posterior uncertainty estimates. This is important for using acquisition function and the reliability of the surrogate model.
2. Kernel choice: Although Matern 5/2 kernel is popular choice in BO for modeling physical systems the authors could briefly justify its selection.
3. Batch Size: Although strategy of point sampling was described in Table 1. no criteria for batch size are provided. How was the batch size decided across iterations?
4. Surrogate model training: There is no description of how GP surrogate is trained and how hyperparameters are optimized. Was there any cross validation performed?
5. Cost function: What is the advantage of using $C(PS_{\text{peak}}, \tau\beta) = -PS_{\text{peak}}/\tau\beta$ as a cost function instead of using Multi-

Objective BBO (MOBO)? Was MOBO considered?

6. Comparison to Sequential BO: Would be nice to include comparison of BBO with a sequential Bayesian optimization concerning convergence speed, final performance, or computational efficiency.

7. Treatment of Noise: Manuscript does not discuss the effect of noise in the simulation outputs and how it is handled in the GP model.

I hope these comments are helpful in further improving the manuscript.

Reviewer #3

(Remarks to the Author)

The manuscript "Batch Bayesian optimization of attosecond betatron pulses from laser wakefield acceleration" by Dominika Maslarova, et al., describes PIC simulation of laser electron acceleration with post-processing to estimate the betatron radiation in the keV spectral region. The simulation setup included spike in the plasma density with three parameters: the spike position, its peak density, and width. These three parameters were optimized using random search and batch Bayesian optimization. The latter is similar to Bayesian optimization, but selection of next parameter set is performed in batches, namely, at each step, several parameter sets to be tried is determined, after which the corresponding simulations run in parallel. The authors set the cost function as the ratio of the post-processed (calculated) betatron radiation power and Full Width at Half Maximum (FWHM) duration of the main pulse, thus trying to make the betatron pulses powerful and keep them short.

As a measure of the betatron pulse duration, the authors use FWHM of the strongest pulse. This measure is not appropriate for the complicated train of complex-shaped betatron pulses shown in Fig. 3b. (Also, the way the pulse train is shown is misleading, so that it makes an impression of a single pulse.) Further, the simulation box and duration are artificially limited, so the actual betatron pulse train may be even more complicated: to clarify if this is so, a longer simulation in a larger box is required.

Finally, the simulations and post-processing (pulse train calculation) procedures need to be documented in more detail. The particular comments and questions are listed below.

1. Page 3: "We set the total plasma length in the x direction to the dephasing length L_d , which is the distance at which the electrons outrun the accelerating wakefield region." The electron bunch entering the laser field can also radiate bright x-rays, which can create an additional atto- or femtosecond-duration pulse(s). Please perform a simulation with the optimum parameters found by the authors with a longer simulation box to check if this effect changes the temporal shape of the x-ray pulse.

2. Fig. 3b shows pulse shapes which are not consistent with the abstract and text, where the authors write "duration around 100 attoseconds" and "keeping durations around $\tau_{\beta} < 100$ as." This is because the authors use the Full Width at Half Maximum (FWHM) duration, which has no meaning for the complicated multi-peak pulse shapes shown in Fig. 3b. A possible reasonable change would be to use the effective pulse width, i.e. the area under the normalized pulse shape curve.

3. Fig. 3b: the number of short pulses shown is three (one "main" pulse around 50 fs, and two weaker but comparable pulses at ~75 and ~105 fs. This possibly indicates that the number of pulses in experiments will be larger, considering that the plasma density cannot abruptly drop to zero, and that the number of periods in the wake wave can be quite large. Please perform at least one simulation where the plasma density gradually drops to zero with a reasonable (practically feasible) linear gradient or another kind of smooth dependence, and where the simulation duration is increased to include a sufficient number of betatron pulses, so that their decrease in time and thus the total pulse train duration can be estimated.

4. Following the previous comment, please provide more details on the sentence on page 5: "These pulses come from electrons accelerated in the second and third wake period and they have significantly lower amplitude than the one coming from the bubble." However, as far as I can see, the amplitudes of the post-pulses is quite similar to the "main" pulse, at least for $d_s = 110$ μm . Please show a figure without the time-break where this relative amplitude can be easily visible (possibly, a set of figures where curve overlapping does not hinder the presentation).

5. Fig 3a: do the spectra shown include only the "main" betatron pulse around 50 fs, or all three pulses shown in Fig. 3b?

6. The "optimum" spike density, $3.9n_0$, is close to the limit set by the authors, $4n_0$. This may indicate that the real optimum spike density could be larger, and $3.9n_0$ was found by the optimization procedure as an approximate value limited by the region of interest. Please clarify this point.

7. Fig. 4 caption: "The colorbar is a semi-logarithmic scale with boundaries F_y belongs to $[-5.5 \text{ N}, 5.5 \text{ N}]$." It is not clear how the force is transformed to values from -1.5 (dark blue) to +1.5 (dark red), please provide a formula.

8. Fig 4: please add the on-axis longitudinal field (E_x) dependence on x_{mw} .

9. Page 7: "Choosing BBO in particular is well motivated if queuing and running the simulations is time consuming, or when time intensive manual processing is required." However, each individual task with a single parameter set (not a batch) can

be completed faster and can use smaller number of CPUs, which in practice means shorter waiting time in the queue, as there are different queues for different task complexities. Thus, it is not obvious that the batch optimization provides an advantage. Please formulate quantitative conditions under which the presented approach is advantageous in terms of the total time (task setting + queuing + simulation).

10. Methods: it is known that when the moving window velocity is different from c , artifacts appear at the simulation box borders. How do the authors deal with this problem?

11. Methods: please comment on any convergence tests performed before the simulation parameters were fixed. Please also provide the values of energy and particle losses (in percents) by the end of the simulations.

12. Methods: "Particle trajectories were extracted from the output of the Smilei code, sampled every $25\Delta t_{SMILEI}$, where electron positions and momenta were tracked for each electron macroparticle with energy greater than 10 MeV." It is not clear if this 10 MeV threshold is significant. Please show electron spectra for the investigated cases (i.e. 6 cases shown in Fig. 3). How does the betatron pulse shape change if the threshold is set at 1 MeV?

13. Please also show the angular patterns of the betatron radiation for the six presented cases.

14. Methods: "first, the temporal resolution was refined from the original $25\Delta t_{SMILEI}$ to $\Delta t_{FIKA} = 25\Delta t_{SMILEI}/(2\gamma_{emax}^2)$, where γ_{emax} is the maximum energy reached along the electron trajectory for each electron. The signal, originally captured every $25\Delta t_{SMILEI}$, is then interpolated to this finer resolution. Subsequently, the Fourier transform was performed on the interpolated signal, obtaining the betatron frequency spectrum." This procedure may miss significant part of the electron dynamics and betatron spectrum, considering that $\gamma_{emax} \sim 100$, and $2\gamma_{emax}^2$ is $\sim 2 \times 10^4$. Certainly, this depends on the smoothness or sharpness of the electron trajectories. Please compare the presented betatron spectra with another run, where the trajectories are saved with the time step of Δt_{SMILEI} (i.e. 25 times finer time step).

15. Fig. S1a: the caption reads "In the hashed areas, the search was forbidden due the restriction in Eq. (5) in the main text," however, points 2, 3, 6, and 7, which seems to be used in the simulations, are within the hashed area. Please clarify.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The revised manuscript is much stronger than the original and my comments have all been addressed satisfactorily. I am happy to support publication.

Reviewer #2

(Remarks to the Author)

Thank you for your detailed revisions. The authors have clearly addressed all of my previous comments related to the Bayesian optimization component of the work, and I am satisfied with the changes. I have no further comments from my side.

Reviewer #3

(Remarks to the Author)

The authors corrected many mistakes and answered many questions in the revised version. I now see only one serious problem, namely, the question of coherent/incoherent radiation summation in the post-processing of PIC results. From the present description I am still not sure how do the authors treat it: from their reply to Referee 1 (Question 10), it seems that the author's strategy is "to use the standard code". However, the code the authors use is not a "standard", although in their group it may be considered so. The best possible solution (for now, until a real breakthrough) can be a more detailed description in the text (not in the reply letter), so that potential readers can understand what was done and how the results were obtained. I provide more detailed comments on what description is necessary below. Certainly, the contradictory statements should be removed/revised, please see my comments 4 and 5.

1. Abstract: the term "electron-rich beam" is not commonly used, please re-word.

2. Supplemental Fig. S3: please add lineouts (in the linear scale), horizontal and vertical.

3. Reply to Q10 of Referee 1: In my opinion, the reply is too brief and should be significantly extended, as this directly relates to the coherent/incoherent summation problem. The authors should mention this problem explicitly, and answer what is done in their code. In particular, is constructive/destructive interference of the radiation produced by post-processing code possible? Certainly, this discussion should be given in the text, not only in the reply letter.

4. Supplementary Note 3: In the second paragraph, the authors write "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)." From this, I conclude

that the betatron radiation was calculated for the $ppc = 8$ and $ppc = 27$ simulations, and compared to the reference $ppc = 1$ case. Please show comparisons (without any normalization) of the betatron radiation spectra and betatron attosecond pulse shapes in Supplemental Fig. S4, as two additional frames.

5. Supplementary Note 3, page 4: "Because radiation with a higher number of macroparticles was difficult to process with our computational resources, we verified only the spectrum of the electron beam, which consequently influences the radiation in a similar manner." This sentence contradicts the previous one and should be removed or revised.

6. Fig. 2 from Response to Referee 3 (with 7 betatron pulses) should be shown in the main text or supplementary.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Response to Referee 1

D. Maslarova, A. Hansson, M. Luo, V. Horný, J. Ferri,
I. Pusztai and T. Fülöp

November 30, 2025

Batch Bayesian optimization of attosecond betatron pulses from laser wakefield acceleration

We thank the referee for their detailed comments. We are particularly grateful for recognizing that the "work is novel due to the exploration of optimizing the x-ray source while maintaining a sub-femtosecond pulse length". We have carefully considered all points raised and made substantial revisions to the manuscript accordingly.

Thus, the current version of the manuscript underwent a major revision, including re-running the whole optimization process with a newly defined cost function. Previously, the cost function was defined based on the peak value of the spectrum and the corresponding full-width-half-maximum duration, which are numerically less robust parameters that depend on the evaluation at particular points. Now we rely on a more robust metric by calculating the energy of the pulse around the area of this peak (in particular, the energy corresponding to 50% of the total pulse energy). We believe that this aspect also considerably improved the overall approach of the paper.

Moreover, in the previous definition, we limited the temporal region where the pulse appeared in the reference case (without a density spike) and then evaluated the improvement within this range. For the current definition, we did not apply this restriction and instead, in each new simulation, the algorithm first searched for the region with the highest peak. After that, it proceeded to calculate the cost function. This approach also resulted in more than an order-of-magnitude enhancement of the peak on-axis radiated energy and led to the discovery of a different enhancement mechanism, which is now described in the manuscript.

Below, we provide a list of responses to each of the comments.

1. *The paper frequently refers to an increase in "photon number", "photon yield" and "photon gain". However, no photon numbers are given, and the optimisation is of the on-axis radiated power per unit solid angle (which is incorrectly called 'power spectrum').*

We thank the referee for pointing out this inconsistency. We now refer to $\frac{d^2W}{dt d\Omega}|_{\text{on-axis}}$ as "on-axis radiated energy per time per solid angle" and $\frac{d^2W}{d\omega d\Omega}|_{\text{on-axis}}$ as "on-axis radiated energy per frequency per solid angle" consistently throughout the manuscript.

2. *The conclusion states “We conclude that BBO successfully identified the region of betatron enhancement with respect to photon gain and located a region in parameter space where the photon gain increased more than threefold compared to the reference value. The analysis of the PIC simulations revealed that the enhancement was driven by a boost in the electron oscillation amplitude occurring behind the bubble, a mechanism that has not been previously investigated in this context.”. This is in-fact only shown for the case of an 80-micron wide density peak, but this is not the case found by the BBO (5 micron). Are the mechanisms the same? Why is there no enhancement for intermediate scales (20 microns)? The conclusion does not reflect the evidence.*

We agree with the referee that the origin of the enhancement differs for short spikes (~ 5 microns) and long spikes (~ 80 microns), which explains the reduced improvement observed at intermediate scales (~ 20 microns). In both cases, the electron oscillation amplitude increases because the transverse force acting on an electron reverses sign as a consequence of the density spike. However, for the 5 μm spike, the bubble responds non-adiabatically because the density rise occurs over a distance comparable to the bubble-size scale, i.e., approximately one plasma wavelength ($\lambda_p = 7.5 \mu\text{m}$ for the uniform density n_0 and $\lambda_p = 3.8 \mu\text{m}$ for the peak spike density $3.9n_0$). The enhancement was therefore caused by a transient change in the local fields (see Figs. 1a,c).

The efficiency of this mechanism decreases at intermediate spike lengths, where the spike exceeds $\sim \lambda_p$. In contrast, for longer spikes, with a length of a few λ_p , the bubble evolves adiabatically, and its size adjusts gradually to the increasing density over several plasma periods (see Figs. 1b,d). Consequently, the electrons slip behind the bubble and remain there for a distance longer than λ_p . This results in an oscillation boost of a significant portion of the electron beam, as it was described in the original version of the manuscript.

Note that in the revised version of the manuscript, with the new definition of the cost function, we discovered a much better performing enhancement regime, and therefore we do not include this discussion there.

3. *The authors communicate an over-confidence that the BBO has found the true global optimum. The failure of an algorithm to improve on a previous optimum in the 4th iteration is taken to mean ‘convergence’ but this is not strong evidence of having found a global optimum. This is a general point on multi-dimensional optimisation. Supplementary note 2, figure 1, shows 2D slices through the 3D model predictions. These plots are also described as evidence of global optimisation as there is only one region of high performance, but these can never reveal the full picture. The author’s statement is even contradicted by their own figure (2b in the main manuscript) which shows a region of increased performance which is not seen in the 2D slices. Overall, the BBO was able to optimize performance compared to the reference case but any claim that this is a global optimum is not supportable.*

We agree that fundamentally, independently of the optimization algorithm and the number of samples taken, it is not possible to prove whether the global optimum has been found. Bayesian optimization can only identify regions of high performance and provide evidence of convergence within the explored domain. We therefore did not claim specifically that we found the global optimum but we agree that some phrasing we used might have sounded overconfident. In the last paragraph of the main manuscript, we have now added the following sentence, to explicitly highlight this:

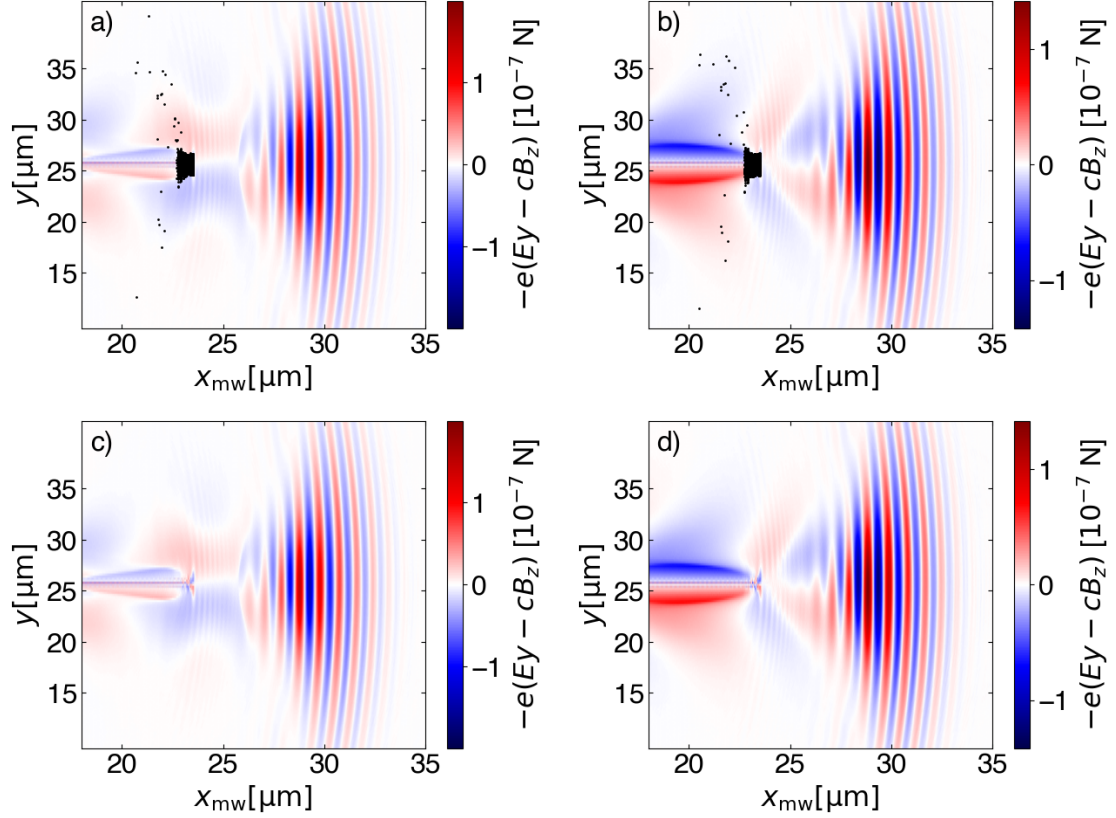


Figure 1: Snapshot of transverse force $-e(E_y - v_x B_z)$ acting on the electron during the laser passage through a spike, for spike lengths a),c) $d_s = 5 \mu\text{m}$, b),d) $d_s = 80 \mu\text{m}$. Here, e is the electron charge, E_y is the transverse electric field in the y direction, the longitudinal electron velocity is approximated with the speed of light $v_x \approx c$ and B_z is the magnetic field acting in the z direction, x_{mw} is an x -coordinate co-moving with the simulation window. Electrons with energies ≥ 10 MeV are shown as black dots in a) and b). c) and d) show the profiles without the black dots, for the visibility of the underlying force structure. The 2D force slices correspond to the center of the z -axis.

”It should be emphasized that although the optimization converged within the explored domain, global optimality cannot be proven in principle for such multi-dimensional problems that may contain several local extrema. This limitation is inherent to such complex optimization landscapes.”

As the description of the batch Bayesian optimization (BBO) process has now been revised to reflect the new results, we also avoided any phrasing that could imply overconfidence in identifying the global optimum. We use the phrasing ’best-performing point’ to indicate the optimum found by BBO (not the global one).

4. In the description of the betatron radiation it is observed that the oscillation amplitude typically decays as the electron beam accelerates. This is correct; however, I infer from the choice of references that this effect is incorrectly attributed to radiative damping. The scaling given is a purely adiabatic effect due to the changing oscillating frequency as the electrons, as described for example in A. G. R. Thomas; “Scalings for radiation from plasma bubbles.” Phys. Plasmas 1 May 2010; 17 (5): 056708. Was radiative dampening included in the PIC simulations? It does not seem to be relevant to these results.

We agree with the referee that this statement was not correct in this context. As the attenuation of the betatron amplitude is not important in the current manuscript version, we deleted this sentence.

5. The title is inaccurate as 100 attoseconds is hardly an “attosecond pulse”. This form of title is also used in preceding publications, but it is clearly misleading.

We agree with the referee that this word choice might be somewhat misleading. However, we would prefer to keep this title as it is the standard usage within “attosecond science” (a field pioneered by the physics Nobel laureates of 2023); namely sub-femtosecond pulses are being referred to as attosecond pulses, see e.g. Refs. [1, 2, 3, 4]. To the best of our knowledge, the shortest electromagnetic pulse produced to date is 43 as [5] - a timescale comparable to those appearing in the manuscript.

6. The paper states “This configuration was used also by Ferri et al., to demonstrate the generation of an attosecond betatron pulse from a down-ramp injection in LWFA.”. Please make it clear that these results are numerical and not an experimental demonstration.

The wording has been now changed to reflect this:

This configuration was used also by Ferri et al. [6], to numerically demonstrate the generation of an attosecond betatron pulse from a down-ramp injection in LWFA.

7. “This approach also enabled close monitoring of the optimization process. This is particularly relevant for strongly nonlinear processes, where it is difficult to a priori know that limiting the search space in a specific way will not lead to solutions of significantly different character, which are physically uninteresting.” The meaning of this sentence is unclear – please revise.

We have expanded on what is meant by this by extending the sentences as
... of significantly different character – posing different resolution requirements, being

physically uninteresting or poorly quantified by the cost function.

As an example (taken from a different field), during a Bayesian optimization of disruption mitigation in fusion reactors by material injection, in part of the parameter space of injected material quantities the disruption remains incomplete. The figures of merit used in the cost function are then not representative of the problems posed by such scenarios. In that case the problematic part of parameter space can be excluded from the optimization [7]. Note though that it is not known, a priori, what parameter regions are problematic, requiring a monitoring of the process; especially if the individual simulations are expensive.

8. The meaning of L_g is not defined in the text, but from the numbers I take it as the position of the density peak for injection. Could the authors clarify this in the text or the figure.

We thank the referee for pointing this out. L_g is the total length of the plasma density gradient that is used to trigger the injection. We have now rephrased the part of the *Attosecond pulse generation* subsection in the main text to describe L_g properly:

”At the entrance of the plasma, a longitudinal density gradient of total length $L_g = L_{g1} + L_{g2} = 40 \text{ } \mu\text{m}$ is introduced to trigger the initial electron injection. The plasma density rises linearly from zero to $n_p = 1.2n_0$, with $n_0 = 2 \times 10^{19} \text{ cm}^{-3}$, over the first $L_{g1} = 30 \text{ } \mu\text{m}$, and then decreases back to n_0 over the following $L_{g2} = 10 \text{ } \mu\text{m}$.”

The L_g notation is now also used in Fig. 1 of the manuscript.

9. The abstract should state that the 350% increase in radiated power is relative to a constant density case.

We changed the formulation of this sentence, as the definition of the cost function was changed. In the new version we use this statement:

This leads to an improvement of over one order of magnitude in the energy contained within the central 50% of the on-axis radiated-energy distribution compared to the reference case without the density spike.

We slightly prefer this formulation (‘reference case without the density spike’) over ‘constant density case’, since the reference plasma profile is not strictly constant: it contains an injection-triggering density gradient at the entrance.

10. Betatron radiation calculations – how are macroparticle contributions combined? Each electron presumable radiates incoherently and therefore they should be weighted as the square root of macroparticle charge. Please clarify.

As described, for instance, in Eqs. 9-12 of Ref. [8], $\frac{d^2W}{d\omega d\Omega}|_{\text{on-axis}} \sim F^2(\omega)$, where the incoherent part of the form factor $F^2(\omega)$ is equal to the number of particles N representing one macroparticle, regardless of their shape. The form factor is obtained by averaging the phase contributions of all electrons over an N -dimensional probability distribution of their positions. Therefore, radiation from each macroparticle was multiplied by the weight of the specific macroparticle, representing N . These contributions from each macroparticle were then summed together.

11. *The Bayesian optimisation methodology is not clearly explained in the main paper. To aid understanding of one of the main components of the paper this should be more clearly explained in the main text. In particular: what kernels and acquisition functions were used? How were the batch of points selected to ensure they are not too close to each other? How are the N_{ran} points used in the optimisation? Does that just refer to the 0th iteration? The supplementary note 1 makes some of this clearer but the main text is confusing. Also, adding in additional random points after the initial set does not make sense to me. As they do not appear to depend on the GP, they could also be run in the 0th iteration and then the model would have more information in the early iterations.*

We changed the sentences in the main text so the specific kernel and acquisition usage are stated explicitly:

In a GP model, correlations between input points are specified by a kernel function. Here, we used the Matérn 5/2 kernel, (defined in Supplementary Note 1), which corresponds to the prior assumption that the cost function is twice mean-square differentiable. Matérn kernels enable more relaxed assumptions on the smoothness of the cost function, which allows to capture also sharper variations that might locally occur, compared to, for instance, the squared exponential (RBF) kernel, which assumes infinitely smooth functions. The acquisition function guides the search by balancing exploration and exploitation to determine where the next point should be evaluated. Here, we used the Expected Improvement (EI) acquisition function [9, 10], which estimates the expected gain over the current best observed value.

Next, the diversity of the points within the batch was assured by the Fantasizer strategy [11]. Fantasizer predicts the potential outcomes of evaluation points in parallel with the optimization process, generating "fantasy" evaluations to estimate how different points might perform. This method guides the optimization towards more promising regions of the search space, allowing for better exploration without needing to evaluate every point directly. We added the following sentence to *Methods* in the main text:

Batch points were selected using the Fantasizer approach [11], which sequentially maximizes EI while updating the GP with "fantasy" observations to avoid selecting points that are too close together.

We agree with the referee that the additional random points were redundant in the optimization process. In the original manuscript, they were included to enhance exploration, but this role is already sufficiently covered by the acquisition function. In the revised version, we therefore excluded N_{ran} , except for the pseudorandom Sobol points used in the initial 0th iteration before the Bayesian optimization began.

We thank the referee again for the constructive feedback that guided the revisions, it has helped us enhance the quality and consistency of the manuscript. We believe that the revisions have now considerably improved the manuscript and we hope that it can be recommended for publication.

Sincerely,

The Authors

References

- [1] Sansone, G. *et al.* Isolated single-cycle attosecond pulses. *Science* **314**, 443–446 (2006). URL <https://www.science.org/doi/abs/10.1126/science.1132838>. <https://www.science.org/doi/pdf/10.1126/science.1132838>.
- [2] López-Martens, R. *et al.* Amplitude and phase control of attosecond light pulses. *Phys. Rev. Lett.* **94**, 033001 (2005). URL <https://link.aps.org/doi/10.1103/PhysRevLett.94.033001>.
- [3] Mauritsson, J. *et al.* Attosecond pulse trains generated using two color laser fields. *Phys. Rev. Lett.* **97**, 013001 (2006). URL <https://link.aps.org/doi/10.1103/PhysRevLett.97.013001>.
- [4] Remetter, T. *et al.* Attosecond electron wave packet interferometry. *Nature Physics* **2**, 323–326 (2006). URL <https://doi.org/10.1038/nphys290>.
- [5] Gaumnitz, T. *et al.* Streaking of 43-attosecond soft-x-ray pulses generated by a passively cep-stable mid-infrared driver. *Optics express* **25**, 27506–27518 (2017).
- [6] Ferri, J., Horný, V. & Fülöp, T. Generation of attosecond electron bunches and x-ray pulses from few-cycle femtosecond laser pulses. *Plasma Physics and Controlled Fusion* **63**, 045019 (2021). URL <https://iopscience.iop.org/article/10.1088/1361-6587/abe885>.
- [7] Ekmark, I. *et al.* Runaway electron generation in disruptions mitigated by deuterium and noble gas injection in SPARC. *Journal of Plasma Physics* **91**, E82 (2025).
- [8] Pausch, R. *et al.* Quantitatively consistent computation of coherent and incoherent radiation in particle-in-cell codes—a general form factor formalism for macro-particles. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **909**, 419–422 (2018).
- [9] Moćkus, J. On Bayesian methods for seeking the extremum. In *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974* 6, 400–404 (Springer, 1975). URL https://link.springer.com/chapter/10.1007/3-540-07165-2_55.
- [10] Jones, D. R., Schonlau, M. & Welch, W. J. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization* **13**, 455–492 (1998). URL <https://link.springer.com/article/10.1023/A:1008306431147>.
- [11] Ginsbourger, D., Le Riche, R. & Carraro, L. Kriging is well-suited to parallelize optimization. In *Computational Intelligence in Expensive Optimization Problems*, 131–162 (Springer, 2010). URL https://link.springer.com/chapter/10.1007/978-3-642-10701-6_6.

Response to Referee 2

D. Maslarova, A. Hansson, M. Luo, V. Horný, J. Ferri,
I. Pusztai and T. Fülöp

November 30, 2025

Batch Bayesian optimization of attosecond betatron pulses from laser wakefield acceleration

Thank you for your thorough reading of our manuscript. We appreciate your insightful suggestions regarding the Bayesian optimization, which helped to improve the reliability of our modeling. Please note that, in the revised manuscript, we now use a more stable cost function: the energy contained in the main radiation pulse, quantified through the 50% energy window around the peak. This quantity is naturally more robust due to its integral basis and provides a more reliable target for Bayesian optimization. We believe that, in practice, this approach substantially improved the robustness and interpretability of the optimization process.

This approach also identified a new optimum compared to the previous definition. This is because, in the previous version, we limited the temporal window where the to-be-enhanced pulse appeared in the reference case, and we subsequently checked for improvement only within this window. In the current definition, we did not apply this restriction. Instead, in each new simulation, the algorithm first searched for the region with the highest peak and then calculated the cost function. As a consequence, a better-performing configuration was found outside the original time range, as the enhanced radiation appeared at a different time. This change substantially affected the final results.

Please find our responses to your comments below. We incorporated most of the suggestions during the new optimization runs. We believe that these updates have significantly improved the methodology from the optimization perspective.

1. Uncertainty Discussion: since BBO depends on using uncertainty estimates to guide exploration vs exploitation, the manuscript lacks discussion or visualization of the GP model's posterior uncertainty estimates. This is important for using acquisition function and the reliability of the surrogate model.

We thank the referee for highlighting the importance of discussing the uncertainty of the surrogate model. In *Supplementary Note 1*, we now present uncertainty 2D maps around the best-performing points from the optimization process (Fig. 1d–f), in addition to the

mean value estimates (Fig. 1a–c). These plots reveal that the model is very confident in the vicinity of the best-performing point and in narrow regions around previously evaluated samples. In contrast, the uncertainty remains large across most of the unexplored parameter space.

This, in principle, indicates that the exploration was insufficient and that the algorithm tended to explore mainly around the best-performing area, which we also observed during the optimization. We could possibly tune the kernel hyperparameters and try different acquisition functions to achieve more exploration. However, we decided not to proceed to explore the full domain further within the scope of this work, as the physical results themselves have very good relevance and a new, experimentally feasible mechanism was discovered due to this optimization. We also show the uncertainty estimates in the main manuscript (Fig. 3) for the points around the best-performing point, mostly to indicate for which experimental parameters this technique is more/less robust. To transparently state the limitations of our model, we say this in the manuscript:

Note that, as shown in Supplementary Note 1, Fig. 1, the uncertainty remains large outside this vicinity, around $\sigma_{\hat{C}} \approx 14$. Therefore, it is generally possible that additional enhancement regions could have been found outside the best-performing region identified by our model if the exploration had been boosted by slightly modifying the optimization strategy. Exploring such possibilities lies beyond the scope of the present work.

2. Kernel choice: Although Matérn 5/2 kernel is popular choice in BO for modeling physical systems the authors could briefly justify its selection.

We added the following discussion to the *Methods* section of the main text, explaining our choice:

In a GP model, correlations between input points are specified by a kernel function. Here, we used the Matérn 5/2 kernel, (defined in Supplementary Note 1), which corresponds to the prior assumption that the cost function is twice mean-square differentiable. Matérn kernels enable more relaxed assumptions on the smoothness of the cost function, which allows to capture also sharper variations that might locally occur, compared to, for instance, the squared exponential (RBF) kernel, which assumes infinitely smooth functions.

The definition of the kernel is also kept in *Supplementary Note 1*.

3. Batch Size: Although strategy of point sampling was described in Table 1. no criteria for batch size are provided. How was the batch size decided across iterations?

We have had to remove the original Table 1 as part of the editing related to trying different batch sizes.

Previously, we directly opted for the batch size $N = 8$ and did not explore its specific performance in detail. We have now also run the optimization for $N = 1$ (corresponding to sequential Bayesian optimization) and for $N = 4$. It is shown that within 8 iterations, $N = 8$ found the best-performing case, but $N = 4$ appears to be the most beneficial

option among the explored ones regarding performance and computational wall-time. More details on this are now shown in Fig. 2 of the main manuscript, and discussed within the *Optimization of betatron radiation* subsection.

Note that we explored the cases $N = 1$ and $N = 4$ only for a limited number of simulations (8 and 32, respectively), and the performance was compared by wall-time rather than total CPU hours. Therefore, for the same number of simulations as performed for $N = 8$ (64), $N = 1$ might have eventually performed better. We did not pursue this further, as the overall optimization time would be prolonged immensely for the $N = 1$ case. However, the performance after the first 8 simulations for each N can still be compared, as we state in the manuscript:

In the first iteration, all batch sizes identified the same best-performing point, providing an improvement of more than 30× compared to the reference. Since only one simulation was required, $N = 1$ was the most efficient at this stage in terms of computational resources. However, after the next seven iterations, the $N = 1$ case reached values comparable to the second-iteration results of $N = 4$ and $N = 8$, which means that for similar number of simulations, all batch-sizes gave similar results.

Also it is worth mentioning, that the 0th iteration, where pseudorandom points were used, was run in a batch of $N_0 = 8$ for each N . The reason for this was to start the optimization from the same initial conditions for each N .

4. Surrogate model training: There is no description of how GP surrogate is trained and how hyperparameters are optimized. Was there any cross validation performed?

At each batch iteration, all accumulated data was used to retrain the surrogate. The kernel hyperparameters of the Gaussian process were optimized by maximizing the Gaussian process log marginal likelihood. This optimization is handled automatically by the Trieste framework [1] during model updates. Regarding this, we added the following statement to *Supplementary Note 1*:

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.

No explicit cross-validation was performed because the dataset is small and produced by expensive simulations. To make best use of all available information, we trained the surrogate on all evaluated points.

5. Cost function: What is the advantage of using $C(\text{PS}_{\text{peak}}, \tau_\beta) = -\text{PS}_{\text{peak}}/\tau_\beta$ as a cost function instead of using Multi-Objective BBO (MOBO)? Was MOBO considered?

We chose $-\text{PS}_{\text{peak}}/\tau_\beta$ because it captures a reasonable trade-off between the peak power value and duration that is easy to interpret. In the current, revised version of the manuscript, we opted for the very similar but more robust function $C(W_{50}, \tau_{50}) = -\frac{W_{50}}{\tau_{50}}$, where W_{50} is the energy contained within the central 50% of the betatron pulse, and τ_{50} is the time interval over which this energy is emitted. Using W_{50} avoids relying on a single spectral peak and offers a more physically representative metric, as described in the introduction of this response letter for the referee.

Our main goal was to increase the radiation gain while avoiding a significant increase of the radiation pulse duration, rather than to explore the full space of trade-offs between objectives with the Pareto front. Therefore, the chosen cost function naturally aligns with our optimization goals and remains within the scope of this study. However, we completely agree that MOBO can provide a deeper insight into the physical behavior, we thus added a following sentence to the Conclusion:

Future work could also benefit from using multi-objective Bayesian optimization techniques to better capture trade-offs across the explored parameter space.

6. Comparison to Sequential BO: Would be nice to include comparison of BBO with a sequential Bayesian optimization concerning convergence speed, final performance, or computational efficiency.

The sequential approach was now performed in the same way as the batch approach, with the specific choice of the batch size of $N = 1$. As discussed in the answer to question 3, we were able to compare the convergence speed of the sequential approach to the batch approaches with $N = 4$ and $N = 8$ for a similar wall-time (i.e., the same number of iterations). Regarding efficiency in terms of computational resources—meaning the total CPU hours (number of simulations)—the performance for the first 8 simulations can be compared. This corresponds to the 8th iteration for $N = 1$, the 2nd iteration for $N = 4$, and the 1st iteration for $N = 8$. We observed comparable performance. The relevant information is shown in Fig. 2 of the main manuscript and discussed in the main text.

7. Treatment of Noise: Manuscript does not discuss the effect of noise in the simulation outputs and how it is handled in the GP model.

Typically, particle positions and velocities are randomly chosen in a simulation, which – in different realizations of the same macroscopic initial condition – yields a statistical variation in the results. However, in our particle-in-cell simulations, we used regular initialization of particles instead of random initialization, with no thermal spread (zero initial momentum). That means that particles are always initialized in the same space-momentum position, which should effectively eliminate statistical noise. Therefore, we assume noise-free simulation output. We add the following statement to Methods: *We assumed a noise-free simulation outputs from Smilei and FIKA. In the PIC Smilei simulations, macroparticles were initialized with regular positions and cold momentum, avoiding randomized initialization. We neglected potential statistical variability between runs, as all simulations were performed with the same number of particles per cell, identical resolution, and on the same computational system (16 nodes, 128 cores each), ensuring reproducible outputs, with any remaining variability expected to be negligible. The GP surrogate model was therefore trained assuming a noise-free problem, with a very small likelihood (noise) variance of 10^{-7} included for numerical stability.*

We thank the reviewer again for their helpful feedback, which, we believe, has contributed to strengthening the manuscript.

Sincerely,

The Authors

References

- [1] SecondMind Labs. Batch Bayesian optimization (2024). URL https://secondmind-labs.github.io/trieste/4.2.2/notebooks/batch_optimization.html. Accessed: 2024-12-19.

Response to Referee 3

D. Maslarova, A. Hansson, M. Luo, V. Horný, J. Ferri,
I. Pusztai and T. Fülöp

November 30, 2025

Batch Bayesian optimization of attosecond betatron pulses from laser wakefield acceleration

We thank the referee for their thorough reading of our manuscript. We appreciate all the suggestions, especially those on the detailed analysis that have improved the overall reliability of the results. The current version of the manuscript underwent a major revision, including re-running the whole optimization process. We changed the cost function, which influenced the final outcome, and the mechanism behind the betatron enhancement is now different. More on this choice is explained in the answer to Question 2 below. We hope these changes have sufficiently addressed all the comments raised by the referee. Please find our responses to the comments below.

1. Page 3: “We set the total plasma length in the x direction to the dephasing length L_d , which is the distance at which the electrons outrun the accelerating wakefield region.” The electron bunch entering the laser field can also radiate bright x-rays, which can create an additional atto- or femtosecond-duration pulse(s). Please perform a simulation with the optimum parameters found by the authors with a longer simulation box to check if this effect changes the temporal shape of the x-ray pulse.

We agree with the referee that the electron beam can enter the laser pulse area. This occurs during the final ~ 150 fs of our simulation as the pulse evolves. Fig. 1 (below) shows the electric field of the laser pulse just before the end of the acceleration.

Note that for this pulse type the depletion is significant, so the electric field strength decreases quickly. As can be seen in Fig. 5d of the main manuscript, the new electron beam fully overtakes the plasma wave creation. It is possible that the residual laser field contributes to the transverse momentum of some electrons and slightly enhances the final radiation peak (but it does not generate a new x-ray pulse).

Since the interaction is visible in the current simulation window, performing longer simulation is not necessary. We added a brief clarification in the manuscript:

A small overlap of the electron beam and laser tail was present during final tens of femtoseconds and might have contributed to the additional radiation gain.

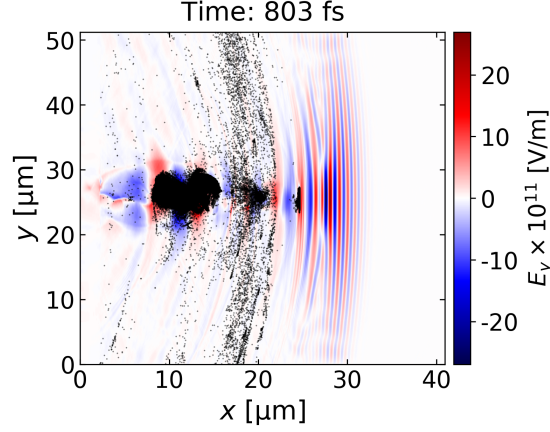


Figure 1: Electric field E_y (red-blue colormap) and the position of electron macroparticles with longitudinal momentum $\geq 10\text{MeV}/c$ (black) shortly before the end of the acceleration. A short electron beam overlapping with the laser field is located at $x \approx 24 \mu\text{m}$, and is followed by a long tail of energetic electrons.

2. *Fig. 3b shows pulse shapes which are not consistent with the abstract and text, where the authors write “duration around 100 attoseconds” and “keeping durations around $\tau_\beta < 100\text{as}$.” This is because the authors use the Full Width at Half Maximum (FWHM) duration, which has no meaning for the complicated multi-peak pulse shapes shown in Fig. 3b. A possible reasonable change would be to use the effective pulse width, i.e. the area under the normalized pulse shape curve.*

We agree with the referee that using the pure FWHM duration for a multi-peaked pulse shape can be misleading. Our initial motivation was that FWHM is a commonly reported experimental measure of the pulse duration, e.g., in Refs. [1, 2]. However, during the review process we decided to reconsider the definition of the cost function, as the original form of the cost function, especially the peak value of the intensity, can be sensitive to the shape fluctuation. The effective pulse width would have a similar problem and be similarly sensitive to the peak fluctuations. For this reason, we decided to adopt a more robust cost-function definition in the revised version of the manuscript. The new cost function is defined as

$$C(W_{50}, \tau_{50}) = -\frac{W_{50}}{\tau_{50}}, \quad (1)$$

where W_{50} is the energy contained within the central 50% of the betatron pulse energy distribution, i.e., between the 25% and 75% of the cumulative energy curve, and τ_{50} is the time interval over which this energy is emitted. The “first pulse” here is defined as the part of the signal that ends when, after the main peak, the intensity stays below 1% of the peak value for a certain minimum time (5 as).

This updated definition again focuses on the first betatron pulse in the pulse train, as it is the brightest one, which may not entirely resolve the referee’s concern. Nevertheless, as we discussed in the manuscript, the choice of the cost function is not absolute and is essentially given by the specific requirements. We note that several reasonable definitions of the cost function can be used, for example, whether one prioritizes total photon yield,

pulse duration, or angular collimation. If, instead, for example, we used the total radiated energy divided by the duration containing half of that total energy instead of Eq. (1), the optimization would favor higher overall photon yield. Our chosen cost function neither explicitly penalizes nor rewards the presence of additional, weaker pulses.

In summary, the current definition of the cost function represents a practical compromise between physical relevance and numerical stability. It captures the most intense and temporally confined part of the radiation, which is generally the quantity of interest in experiments, and remains stable against small distortions of the pulse shape. In the table below, for overview, we show the values of the current metric τ_{50} and the effective pulse duration for the reference and optimal case.

	Reference	Optimum
τ_{50}	437 as	44 as
Effective width	200 as	73 as

We would like to also explain the fact that the new cost function definition led to a different optimal configuration than the one obtained with our previous cost function. Previously, the evaluation was restricted to a fixed temporal interval, the same interval in which the enhanced pulse appeared in the reference case. Improvements during BBO were then investigated only within that window. In the revised formulation, we removed this constraint. For each simulation, the algorithm first located the temporal region containing the strongest peak and then evaluated the cost function based on that location. As a result, the best-performing solution changed, since the enhanced radiation in the improved configuration occurred at a different time than in the original case.

3. Fig. 3b: the number of short pulses shown is three (one “main” pulse around 50 fs, and two weaker but comparable pulses at ~ 75 and ~ 105 fs. This possibly indicates that the number of pulses in experiments will be larger, considering that the plasma density cannot abruptly drop to zero, and that the number of periods in the wake wave can be quite large. Please perform at least one simulation where the plasma density gradually drops to zero with a reasonable (practically feasible) linear gradient or another kind of smooth dependence, and where the simulation duration is increased to include a sufficient number of betatron pulses, so that their decrease in time and thus the total pulse train duration can be estimated.

We performed an additional simulation of the reference case using a longer simulation box to capture eight wakefield periods. In this simulation, instead of 216- μm -long density plateau as in the original reference case, we used a 176 μm plateau followed by a 40 μm gradient that linearly drops from n_0 to 0. The results are shown in Fig. 2. As seen in Fig. 2c, electrons from each period generate a radiation burst, contributing to the increase in the energy spectrum in Fig. 2a. It is very likely that more periods (above eight) would radiate even more. However, the signal (except for the third pulse) gradually decreases, and this trend is expected to continue for later periods.

We did not analyse the subsequent pulses further, as our study focuses on the brightest peak. Moreover, compared to the optimized case (Fig. 4a of the main manuscript), their

energy contribution is negligible. In general, the choice of which pulse to optimize depends on the requirements of the particular experiment, and such trade-offs can be incorporated into the cost function. One could, in principle, investigate the enhancement of later pulses using alternative cost functions that favour or penalize their energy, but this lies outside the scope of the present work. To comment on this, we added the following statement to the main text, subsection *Optimization of betatron radiation* in *Results*:

Note that in the reference case additional, weaker betatron pulses appear at later observation times. In our simulation setup we observe two such secondary peaks. To verify their physical origin and relative contribution, we performed an extended reference simulation covering eight wakefield periods, in which betatron radiation is generated during each plasma period. In the later periods, however, the accelerating fields are significantly weaker, resulting in electron beams with lower charge and energy. Consequently, the combined radiation from the extra seven periods amounts to only around 14% of the total betatron on-axis energy. The cost function in Eq. (1) is defined to focus on improving the main first, strongest betatron pulse, which is the pulse of primary interest in this study. As a consequence, the secondary pulses are neither rewarded nor penalized. Optimizing these later pulses would require a different metric depending on a specific application aim, and is beyond the scope of the present work.

4. Following the previous comment, please provide more details on the sentence on page 5: “These pulses come from electrons accelerated in the second and third wake period and they have significantly lower amplitude than the one coming from the bubble.” However, as far as I can see, the amplitudes of the post-pulses is quite similar to the “main” pulse, at least for $d_s = 110 \mu\text{m}$. Please show a figure without the time-break where this relative amplitude can be easily visible (possibly, a set of figures where curve overlapping does not hinder the presentation).

Fig. 4a of the main manuscript now presents the full radiation signal for the reference and best-performing optimized cases. The relative amplitude is now therefore well visible. The main first peaks for both cases are shown as insets, to also present the details of the pulses. Note that the noisy tail in the optimized case has almost the same peak value of on-axis radiated energy per time per solid angle as the peak value of the reference, which is significantly lower compared to the main signal (by $\sim 25 \times$). Therefore, we did not focus on more detailed analysis of it.

5. Fig 3a: do the spectra shown include only the “main” betatron pulse around 50 fs, or all three pulses shown in Fig. 3b?

The spectra in the original Fig. 3a, now Fig. 4d, contain the whole radiation signal from the simulation box. We added a comment to the figure caption to clarify that:

The energy spectra are calculated from the whole temporal profile depicted in a).

6. The “optimum” spike density, $3.9n_0$, is close to the limit set by the authors, $4n_0$. This may indicate that the real optimum spike density could be larger, and $3.9n_0$ was found by the optimization procedure as an approximate value limited by the region of interest. Please clarify this point.

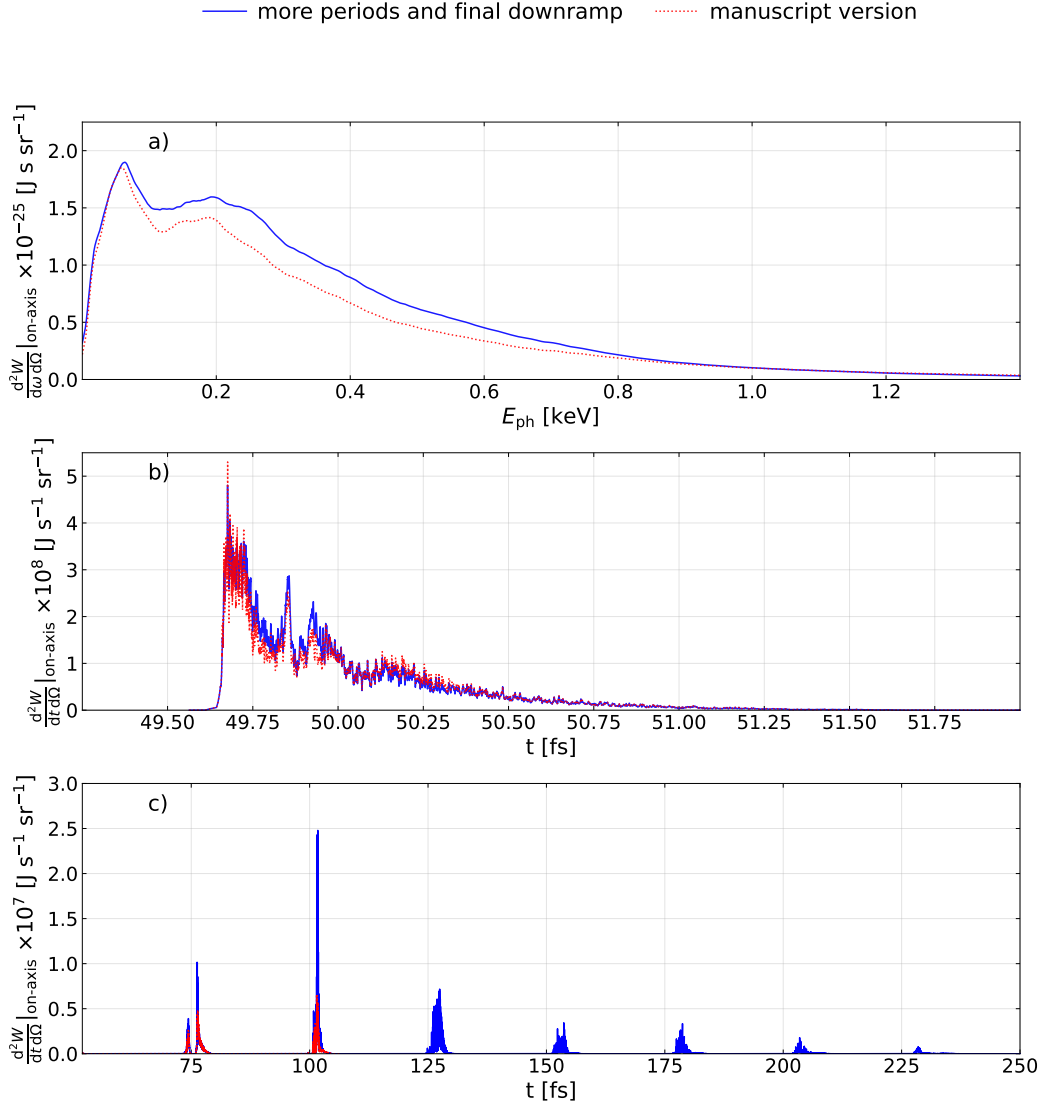


Figure 2: a) Dependence of the radiated energy W per photon angular frequency ω per solid angle Ω on photon energy E_{ph} . b) Dependence of W per observer time t per Ω for the signal time range $t \in [49.25 \text{ fs}, 52 \text{ fs}]$. c) Dependence of W per observer time t per Ω on the observer time for the signal time range $t \in [52 \text{ fs}, 250 \text{ fs}]$.

We agree with the referee that the improvement might be better beyond the explored region. However, as the boundaries of the search area can be potentially extended in various directions, we decided to limit the investigations within this manuscript and we kept the fixed the bounds, as decided a priori. We now clarify this point in the subsection *Optimization of betatron radiation*:

Because the search space allows for many reasonable choices of parameter bounds, we fixed the bounds a priori and did not optimize them in advance. The results therefore reflect improvements within this specified domain. Alternative or expanded limits could

potentially lead to other enhancement schemes - for example, higher plasma density for the accelerator-radiator arrangement [3] or longer plasma favoring additional interaction of the electrons with the laser pulse [4, 5, 6, 7]. While such alternative configurations might yield additional improvements, analyzing these dependencies lies beyond the present scope of this work.

7. Fig. 4 caption: The colorbar is a semi-logarithmic scale with boundaries F_y belongs to $[-5.5N, 5.5N]$. It is not clear how the force is transformed to values from -1.5 (dark blue) to $+1.5$ (dark red), please provide a formula.

In the original version of the manuscript, values close to zero were displayed on a linear scale to preserve small-force features, while values above 1 and below -1 were displayed on a logarithmic scale to compress large variations. In the current version of the manuscript, we did not include F_y as it was not relevant for the new findings. However, we use a power-law rule for the density profiles in the current version (Fig. 5), which we explain with a formula: *For better visualization, n/n_0 is transformed using a power-law function with exponent 0.35, $((n/n_0)/(n/n_0)_{\max})^{0.35}$, where $(n/n_0)_{\max}$ corresponds to the maximum value of the normalised density n/n_0 .*

8. Fig 4: please add the on-axis longitudinal field (E_x) dependence on x_{mw} .

Thank you for this comment. The on-axis E_x is now depicted and added onto the density profile in Fig. 5 with a blue line. (Note that in the current version of the manuscript Fig. 5 corresponds to the same discussion as Fig. 4 in the previous version.)

9. Page 7: “Choosing BBO in particular is well motivated if queuing and running the simulations is time consuming, or when time intensive manual processing is required.” However, each individual task with a single parameter set (not a batch) can be completed faster and can use smaller number of CPUs, which in practice means shorter waiting time in the queue, as there are different queues for different task complexities. Thus, it is not obvious that the batch optimization provides an advantage. Please formulate quantitative conditions under which the presented approach is advantageous in terms of the total time (task setting + queuing + simulation).

We now comment on this aspect more extensively in the subsection *Optimization of betatron radiation*. A comparison of the convergence of the sequential Bayesian optimization (batch size $N = 1$) and batch Bayesian optimization (for $N = 4, 8$) is shown in Fig. 2 of the manuscript. We also compare the wall time and the queueing process and comment on the usage of more CPUs, e.g.:

In general, however, most variations in wall-clock time across iterations and batch sizes were driven by fluctuating queue delays on the supercomputing system rather than by differences among the simulations themselves. This shows that the performance of BBO depends on available parallelization. Note that increasing the batch size leads to higher total CPU usage, since more simulations run simultaneously.

In this way, we provide a discussion of the advantages and disadvantages of using the batch approach.

10. *Methods: it is known that when the moving window velocity is different from c , artifacts appear at the simulation box borders. How do the authors deal with this problem?*

Regarding the fact that in our original manuscript we used the group velocity of the laser pulse as the velocity of the moving window $v_{\text{mw}} = v_{L_g}$ instead of the speed of light $v_{\text{mw}} = c$, here, we simulate both cases and compare the radiation. The results are shown in Fig. 3. It can be seen that the differences in the spectra are negligible. Also the difference in the current cost function makes 0.04% for the current simulation setup. We now keep $v_{\text{mw}} = c$ in all the simulations in the new version of the manuscript, for simplicity.

11. *Methods: please comment on any convergence tests performed before the simulation parameters were fixed. Please also provide the values of energy and particle losses (in percents) by the end of the simulations.*

Before we decided to rerun the optimization with the updated cost function, we performed a new comprehensive set of numerical tests, which is now described in *Supplementary Note 3: On numerical convergence*. We performed several tests for both the particle-in-cell simulations and the radiation output, which also address other referees' comments discussed below.

Regarding the evolution of the energy and charge of the electron beam, we discuss that now in *Supplementary Note 2: On the properties of the electron beam and betatron radiation*. Supplementary Fig. 2 shows spectra at four simulation times for the reference case and best-performing optimized case. Supplementary Tab. 1 shows the corresponding evolution of the charge.

Regarding the numerical energy loss and particle losses in the particle-in-cell simulation, we also included it in *Supplementary Note 3: On numerical convergence*:

Finally, we evaluated the energy reduction and particle losses across the boundaries of the PIC domain. The total energy within the particle-in-cell 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.

12. *Methods: "Particle trajectories were extracted from the output of the Smilei code, sampled every $25\Delta t_{\text{SMILEI}}$, where electron positions and momenta were tracked for each electron macroparticle with energy greater than 10 MeV." It is not clear if this 10 MeV threshold is significant. Please show electron spectra for the investigated cases (i.e. 6 cases shown in Fig. 3). How does the betatron pulse shape changes if the threshold is set at 1 MeV?*

The 10 MeV threshold is not a physical cutoff but a technical parameter that determines which electrons are recorded for trajectory tracking in *Smilei* and subsequently passed to *FIKA* for radiation calculation. Fig. 4 depicts a convergence test. Note that we now also switched to longitudinal momentum threshold p_x^{thresh} instead of energy in our simulations (i.e. 10 MeV/c instead of 10 MeV), which simplifies the setup and express the

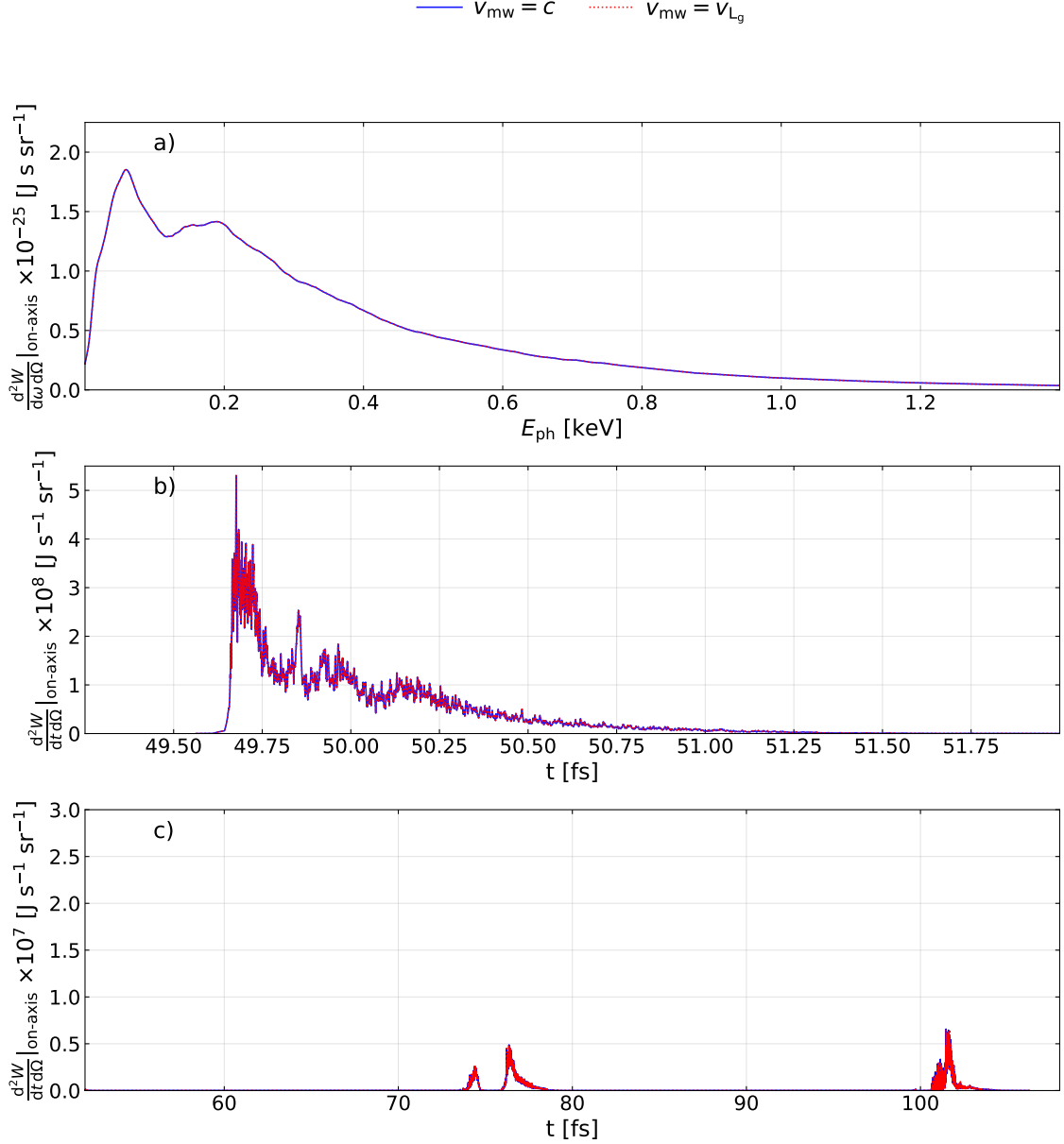


Figure 3: Comparison of the radiation for moving window velocities $v_{mw} = v_{L_g}$ and $v_{mw} = c$: a) Energy spectrum, b) temporal profile of the first pulse, c) temporal profile of the second and third betatron pulse.

gain in the direction of the acceleration. The transverse momentum is low compared to the longitudinal one and this approach does not change the convergence test perspective.

Testing thresholds below $3 \text{ MeV}c^{-1}$ (such as $1 \text{ MeV}c^{-1}$) was not feasible with our available storage and memory, because tracking of many low-energy electrons produces extremely large trajectory datasets, however, we tested the performance for $p_{xe}^{\text{thresh}} = 3, 5, 10 \text{ MeV}c^{-1}$ and $15 \text{ MeV}c^{-1}$ in simulations and examined how it influences the spec-

trum. The comparison is depicted in Fig. 4. The spectral shape does not vary much for all the cases. The values of the cost function difference show very little difference (maximum of 0.5%) for choices of $p_{xe}^{\text{thresh}} = 3, 5 \text{ MeV}/c$ and $15 \text{ MeV}/c$, as shown in Tab. 1. Therefore, we picked $10 \text{ MeV}/c$ as a reasonable convergence value.

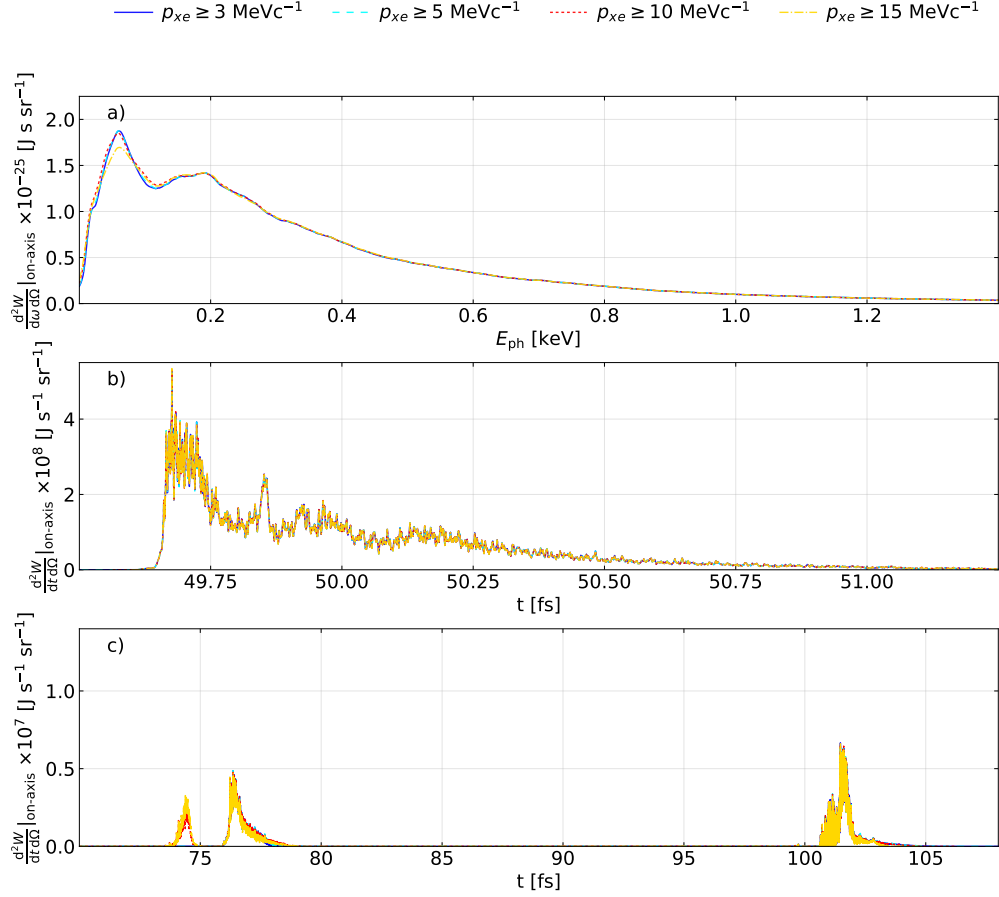


Figure 4: **Radiation spectra for different momentum thresholds.** Each case represents different $p_{xe} > p_{xe}^{\text{thresh}}$, where $p_{xe}^{\text{thresh}} = 3, 5, 10$, and $15 \text{ MeV}/c$. a) Dependence of the radiated energy W per photon angular frequency ω per solid angle Ω on photon energy E_{ph} . b) Dependence of W per observer time t per Ω for the signal time range $t \in [49.5 \text{ fs}, 51.25 \text{ fs}]$. c) Dependence of W per observer time t per Ω for the signal time range $t \in [70 \text{ fs}, 108 \text{ fs}]$.

We also commented on this in *Supplementary Note 3*: *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.*

13. Please also show the angular patterns of the betatron radiation for the six presented cases.

$p_{x_e}^{\text{thresh}}$	Cost function difference
3 MeVc ⁻¹	0.5%
5 MeVc ⁻¹	0.0%
15 MeVc ⁻¹	0.1%

Table 1: Percentage differences of the cost function for different threshold of electron momentum $p_{x_e} \geq p_{x_e}^{\text{thresh}}$ compared to the reference case 10 MeVc⁻¹. The values in the table were calculated as (case – reference)/reference · 100%.

Angular spreads are now also included and discussed in *Supplementary Note 2*, in subsection *Angular spectrum of the radiation*. For clarity and conciseness, considering the changes in the main manuscript, we decided to show only two most relevant cases, the reference and the best performing case. By this approach, we believe that the most important aspects of the radiation modification regarding the angular patterns are captured.

14. Methods: “first, the temporal resolution was refined from the original $25\Delta t_{\text{SMILEI}}$ to $\Delta t_{\text{FIKA}} = 25\Delta t_{\text{SMILEI}}/(2\gamma_{e\text{max}}^2)$, where $\gamma_{e\text{max}}$ is the maximum energy reached along the electron trajectory for each electron. The signal, originally captured every $25\Delta t_{\text{SMILEI}}$, is then interpolated to this finer resolution. Subsequently, the Fourier transform was performed on the interpolated signal, obtaining the betatron frequency spectrum.” This procedure may miss significant part of the electron dynamics and betatron spectrum, considering that $\gamma_{e\text{max}}$ is ~ 100 , and $2\gamma_{e\text{max}}^2$ is $\sim 2 \times 10^4$. Certainly, this depends on the smoothness or sharpness of the electron trajectories. Please compare the presented betatron spectra with another run, where the trajectories are saved with the time step of $1\Delta t_{\text{SMILEI}}$ (i.e. 25 times finer time step).

We performed a convergence test, which we also comment on in *Supplementary Note 3: On numerical convergence*.

To elaborate, we systematically varied the *FIKA* timestep as $\Delta t_{\text{FIKA}} = 25\Delta t_{\text{SMILEI}}$, $\Delta t_{\text{FIKA}} = 15\Delta t_{\text{SMILEI}}$ and $\Delta t_{\text{FIKA}} = 5\Delta t_{\text{SMILEI}}$. The spectra are visually indistinguishable, as can be seen in Fig. 5 below. Tab. 2 summarizes the corresponding variations in the cost function and computational resources used. The simulations exhibited good convergence, with the *FIKA* runtime being the dominant contributor to the total computational cost. The difference in the cost function for $\Delta t_{\text{FIKA}} = 15\Delta t_{\text{SMILEI}}$ compared to the $\Delta t_{\text{FIKA}} = 5\Delta t_{\text{SMILEI}}$ case was below 1%, while the total computational time did not increase significantly compared to $\Delta t_{\text{FIKA}} = 25\Delta t_{\text{SMILEI}}$.

Based on this convergence study, we now selected $\Delta t_{\text{FIKA}} = 15\Delta t_{\text{SMILEI}}$ for all Bayesian runs, as it provides a great balance between accuracy and computational cost. Performing *Smilei* simulations with output corresponding to $t_{\text{FIKA}} = 1\Delta t_{\text{SMILEI}}$ would require extensive computational responses and given the observed convergence, we believe that exploring $\Delta t_{\text{FIKA}} = 1\Delta t_{\text{SMILEI}}$ would provide minimal additional insight in this context.

We also comment on this within the frame of numerical tests in *Supplementary Note 3: On numerical convergence*:

*For the FIKA radiation calculation, we varied the input 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:*

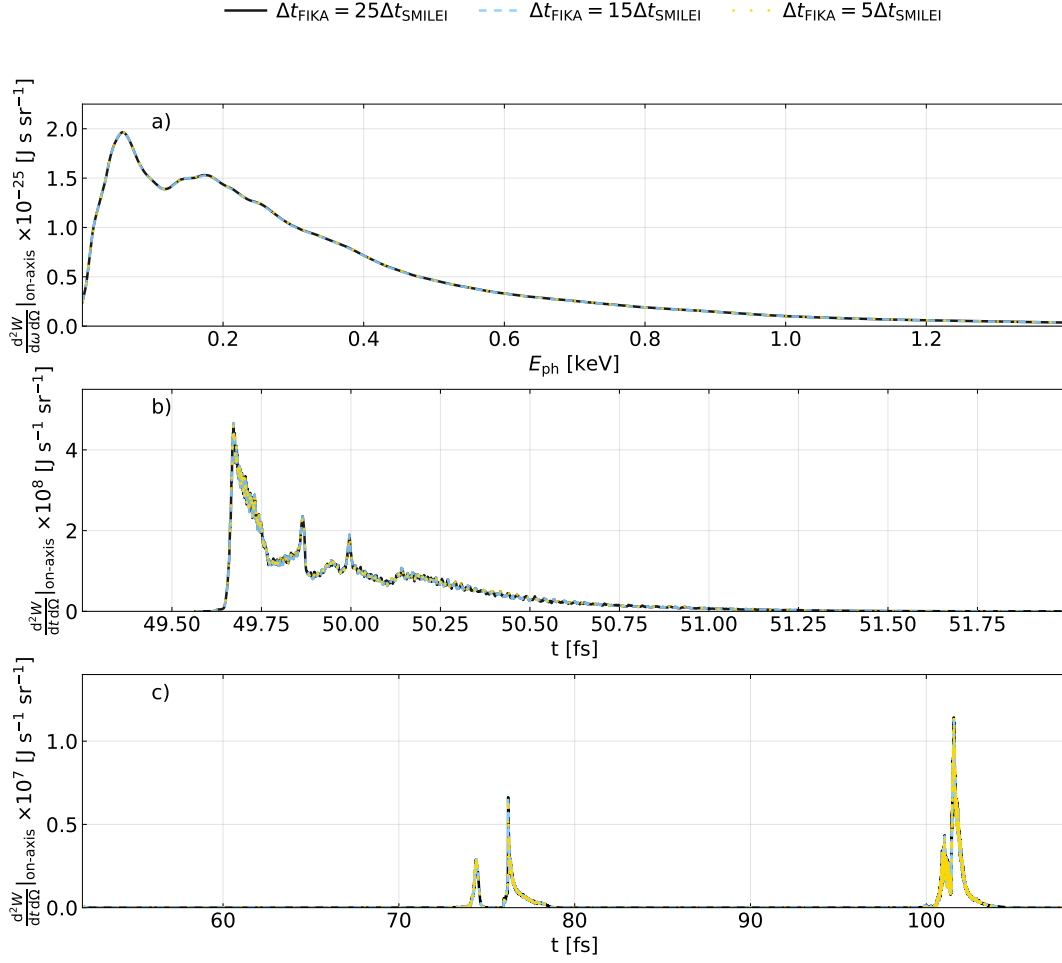


Figure 5: Comparison of the radiation for different input timesteps to the *FIKA* code: a) Energy spectrum, b) temporal profile of the first pulse, c) temporal profile of the second and third betatron pulses.

Numerical case	Cost function	Smilei time	FIKA time
$\Delta t_{\text{FIKA}} = 25\Delta t_{\text{SMILEI}}$	-0.4%	-7.1%	0%
$\Delta t_{\text{FIKA}} = 5\Delta t_{\text{SMILEI}}$	0.6 %	35.7%	233.3%

Table 2: **Percentage differences of the cost function and computational time for different time resolution input to FIKA.** The cases are compared to the reference case of $t_{\text{FIKA}} = 15\Delta t_{\text{SMILEI}}$, which was used as a final pick in our simulations. For this reference case, the value of the cost function is $-1.128 \times 10^8 \text{ Wsr}^{-1}$, the Smilei computational time is 2902 core hours (wall-time of 1.4 hours) and the Fika computational wall-time is 0.3 hours. The Smilei computations were all performed with 2048 cores (16 nodes, 128 cores each) and FIKA computations were performed in a serial manner, always on one computational node. All the simulations were run on the same computational system. The values in the table were calculated as $(\text{case} - \text{reference})/\text{reference} \cdot 100\%$.

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 $\Delta t_{\text{FIKA}} = 15 \Delta t_{\text{SMILEI}}$ 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.

15. Fig. S1a: the caption reads “In the hashed areas, the search was forbidden due the restriction in Eq. (5) in the main text,” however, points 2, 3, 6, and 7, which seems to be used in the simulations, are within the hashed area. Please clarify.

In the original manuscript, the additional parametric scan included parameter values extending outside the region explored by the optimization algorithm. To keep the analysis consistent, all investigations in the revised version remain within the parameter range explored during the optimization. The parametric scan shown in Fig. 3 of the main manuscript is restricted to the parameter range used in the BBO search (Eqs. 2-5 of the manuscript). This means that it excludes the regions outside the hashed area in Supplementary Fig. 1.

Please note that not all of the detailed information raised in the questions is included explicitly within the main manuscript or the Supplementary Notes, due to the limited scope and length constraints. Nevertheless, we have made every effort to address the main points throughout the text and to clarify the underlying reasoning. Thank you again for your constructive and insightful comments. We believe that they have significantly improved the quality of our work, and we greatly appreciate the time invested in the review process.

Sincerely,

The Authors

References

- [1] Hammond, T., Brown, G. G., Kim, K. T., Villeneuve, D. & Corkum, P. Attosecond pulses measured from the attosecond lighthouse. *Nature Photonics* **10**, 171–175 (2016).
- [2] Xu, J. *et al.* Terawatt-scale optical half-cycle attosecond pulses. *Scientific Reports* **8**, 2669 (2018).
- [3] Ferri, J. *et al.* High-brilliance betatron γ -ray source powered by laser-accelerated electrons. *Physical Review Letters* **120**, 254802 (2018).

- [4] Nemeth, K. *et al.* Laser-driven coherent betatron oscillation in a laser-wakefield cavity. *Physical Review Letters* **100**, 095002 (2008). URL <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.100.095002>.
- [5] Cipiccia, S. *et al.* Gamma-rays from harmonically resonant betatron oscillations in a plasma wake. *Nature Physics* **7**, 867–871 (2011). URL <https://www.nature.com/articles/nphys2090>.
- [6] Huang, K. *et al.* Resonantly enhanced betatron hard x-rays from ionization injected electrons in a laser plasma accelerator. *Scientific Reports* **6**, 27633 (2016). URL <https://doi.org/10.1038/srep27633>.
- [7] Mangles, S. P. *et al.* Laser-wakefield acceleration of monoenergetic electron beams in the first plasma-wave period. *Physical Review Letters* **96**, 215001 (2006). URL <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.96.215001>.

Response to Referee 3

D. Maslarova, A. Hansson, M. Luo, V. Horný, J. Ferri,
I. Pusztai and T. Fülöp

January 18, 2026

Batch Bayesian optimization of attosecond betatron pulses from laser wakefield acceleration

We thank the referee for their positive feedback about our changes. We appreciate the additional suggestions to further improve our manuscript and to address the ambiguity about the radiation code we use. The discussion regarding the main concern about coherent/incoherent contribution is included in the response to Question 3. We also respond to the other comments below.

1. Abstract: the term "electron-rich beam" is not commonly used, please re-word.

We changed the wording from *electron-rich beam* to *high-charge electron beam*.

2. Supplemental Fig. S3: please add lineouts (in the linear scale), horizontal and vertical.

The lineouts of the spectra are now added to Supplemental Fig. 3 in separate panels c) and d) (for better visibility), with the corresponding caption:

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).

3. Reply to Q10 of Referee 1: In my opinion, the reply is too brief and should be significantly extended, as this directly relates to the coherent/incoherent summation problem. The authors should mention this problem explicitly, and answer what is done in their code. In particular, is constructive/destructive interference of the radiation produced by post-processing code possible? Certainly, this discussion should be given in the text, not only in the reply letter.

Thank you for pointing out this lack of clarity regarding the code. *FIKA* treats only incoherent radiation and does not include a module for coherent radiation, where the macroparticle shape would also need to be taken into account. This approach is sufficient for the calculation of betatron radiation, which is incoherently summed [1]. We have

now added the following text to *Methods* of the main manuscript, subsection *Betatron radiation calculation*:

In FIKA, only the incoherent part of the radiation is calculated, neglecting any coherence effects in the radiation. This approach is sufficient, as electrons emit radiation incoherently as betatron radiation from relativistic electron beams [1], and collective coherence effects can be neglected. Each PIC macroparticle is treated as a statistically independent radiation source, with no interference between different macroparticles. The final radiation is then obtained through a weighted summation of the radiation contributions from each PIC macroparticle [2], with macroparticle statistical weights incorporated into the calculation.

4. *Supplementary Note 3: In the second paragraph, the authors write "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)." From this, I conclude that the betatron radiation was calculated for the $ppc = 8$ and $ppc = 27$ simulations, and compared to the reference $ppc = 1$ case. Please show comparisons (without any normalization) of the betatron radiation spectra and betatron attosecond pulse shapes in Supplemental Fig. S4, as two additional frames.*

We added it as a separate figure in the Supplementary Material, currently numbered as Fig. 5. We refer to it in body of the Supplementary Note as follows:

The corresponding comparison of the spectra is shown in Fig. 5.

5. *Supplementary Note 3, page 4: "Because radiation with a higher number of macroparticles was difficult to process with our computational resources, we verified only the spectrum of the electron beam, which consequently influences the radiation in a similar manner." This sentence contradicts the previous one and should be removed or revised.*

We agree with the referee, and removed the sentence.

6. *Fig. 2 from Response to Referee 3 (with 7 betatron pulses) should be shown in the main text or supplementary.*

We added this figure to the Supplementary Material as Supplemental Fig. 4. In addition, we included results for simulations with multiple betatron pulses but without the final density down-ramp at the end of the simulation window. This was done in order to distinguish the respective contributions of additional wakefield periods and the plasma density down-ramp to the radiation spectrum. Because these results required more detailed discussion, we added a new subsection *Influence of multiple wakefield periods and a density down-ramp at the plasma exit on the radiation profile* to Supplementary Note 2. We refer to it in the main text as:

More details on the electron and betatron beam properties, including electron energy, charge, the angular radiation spectrum, as well as the effects of multiple wakefield periods and a plasma density down-ramp at the plasma exit, are described in Supplementary Note 2.

Note that, in addition to the changes suggested by the referee, we noticed that the description of the cost function improvement in the abstract could be misleading. We therefore changed it to:

This results in an improvement of more than one order of magnitude in the on-axis time-averaged power within the central time containing half of the radiated energy, compared to the reference case without the density spike.

In addition, we added the following statement about increased energy W_{50} , for consistency with τ_{50} :

The value of τ_{50} dropped to $\tau_{50}^{\text{opt}} = 43.50$ as.

was changed to:

The value of W_{50} increased to $W_{50}^{\text{opt}} = 305.0$ nJ sr⁻¹ and the value of τ_{50} dropped to $\tau_{50}^{\text{opt}} = 43.50$ as.

The changes in the main text and supplementary material are highlighted in the edited documents with blue text color. In addition, we have added a new reference (Ref. 16) to the main manuscript addressing electron beam durations in the sub-femtosecond regime.

We thank the referee again for their careful review and valuable comments.

Sincerely,

The Authors

References

- [1] Corde, S. *et al.* Femtosecond x rays from laser-plasma accelerators. *Reviews of Modern Physics* **85**, 1–48 (2013). URL <https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.85.1>.
- [2] Pausch, R. *et al.* Quantitatively consistent computation of coherent and incoherent radiation in particle-in-cell codes—a general form factor formalism for macro-particles. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **909**, 419–422 (2018).