RESEARCH ARTICLE

Prospects for statistical tests of strong-field quantum electrodynamics with high-intensity lasers

C. Olofsson [ORCID] and A. Gonoskov

Department of Physics, University of Gothenburg, Gothenburg, Sweden
(Received 8 March 2023; revised 21 June 2023; accepted 27 July 2023)

Abstract

Exploiting high-energy electron beams colliding into high-intensity laser pulses brings an opportunity to reach high values of the dimensionless rest-frame acceleration χ and thereby invoke processes described by strong-field quantum electrodynamics (SFQED). Measuring deviations from the results of Furry-picture perturbation theory in SFQED at high χ can be valuable for testing existing predictions, as well as for guiding further theoretical developments. Nevertheless, such experimental measurements are challenging due to the probabilistic nature of the interaction processes, dominating signals of low-χ interactions and limited capabilities to control and measure the alignment and synchronization in such collision experiments. Here we elaborate a methodology of using approximate Bayesian computations for drawing statistical inferences based on the results of many repeated experiments despite partially unknown collision parameters that vary between experiments. As a proof-of-principle, we consider the problem of inferring the effective mass change due to coupling with the strong-field environment.

1. Introduction

Although fundamental principles of quantum electrodynamics (QED) are known for their precise experimental validations, the implications they purport for sufficiently strong electromagnetic fields remain theoretically intricate and lack experimental data. Colliding accelerated electrons with high-intensity laser pulses can be seen as a newly emerging pathway to such experimental data[1–4]. The local interaction is characterized by the dimensionless ratio of the electron acceleration in its rest frame to the acceleration that would be caused by the Schwinger field E_{crit}:

$$\chi = \frac{\gamma_e}{E_{\mathrm{crit}}}\sqrt{\left(\vec{E}+(\vec{v}/c)\times\vec{B}\right)^2 - \left(\vec{E}\cdot\vec{v}/c\right)^2}, \quad (1)$$

where $\vec{v}, \gamma_e$ are the velocity and Lorentz factor of the electron, respectively, whereas $\vec{E}, \vec{B}$ are the electromagnetic field vectors. Here, $E_{\mathrm{crit}} = m_e^2 c^3/q_e\hbar \approx 10^{18}$ V/m, where $\hbar$ is the reduced Planck constant, c is the speed of light and m_e and q_e are the mass and charge of the electron, respectively.

At $\chi \ll 1$ the electrons are subject to classical emission and the corresponding radiation reaction[5]. Emission of photons and corresponding recoils at $\chi \sim 1$ are described by nonlinear Compton scattering and have been experimentally observed in several experiments[6–9]. Measuring quantitative properties of the photon emission (e.g., energy, angular or polarization distribution) at $\chi \sim 1$ can be perceived as a logical next step, while results for $\chi \gg 1$ can potentially facilitate theoretical developments or even lead to fundamental discoveries (see Ref. [10] and references therein).

A severe obstacle for the outlined efforts is the interaction complexity. The value of χ for each electron in the beam varies in time and overall depends on the electron position relative to the laser pulse location, which can also vary from experiment to experiment due to spatio-temporal mismatches. In addition, for contemporary laser pulse durations, many electrons can lose a significant part of their initial energy prior to reaching the strong-field region, where they have a chance to emit at high χ. Furthermore, due to the Breit–Wheeler process the emitted photons can decay into electron–positron pairs, which can lead to the onset of an electromagnetic cascade. In combination, this can make the measurable post-collision distributions of photons, electrons and positrons be predominantly determined by low-χ emissions, giving no direct information about emissions at high-χ, even if they had been invoked.

Correspondence to: C. Olofsson, Department of Physics, University of Gothenburg, SE-41296 Gothenburg, Sweden. Email: christof-fer.olofsson@physics.gu.se

One known way of dealing with such difficulties is Bayesian binary hypothesis testing, which is based on comparing experimental results with the outcomes computed on the basis of each of two competing theories. However, even in the absence of a distinct hypothesis to be tested, one can use a similar technique to determine parameters that quantify deviations from the approximate theory (sometimes referred to as the parameter calibration procedure[11–13]), which in our case can be the theory on nonlinear Compton scattering that is valid for moderate χ values. One practicality of this approach is the possibility to gain statistically rigorous knowledge from many experiments even in case of low repeatability. For example, the inference about high-χ events is feasible even if the collision between the laser pulse and the electron bunch happens with a spatial offset that varies from collision to collision without being measured.

In this paper we consider the possibility of using the technique of approximate Bayesian computation (ABC) in the forthcoming experiments[11,14,15]. Since the application of ABC is known to require problem-specific developments and analysis, we consider some essential elements using a proof-of-principle problem incorporating this method for measuring the constant that quantifies the effective mass shift[10,16–18]. We assess the use of the ABC technique in the context of possible experimental conditions and analyze the main requirements, difficulties and opportunities for improvements. The paper is arranged as follows. In Section 2 we demonstrate a proof-of-principle approach to infer the effective mass change, assessing the difficulties and limitations. In Section 3 we motivate the use of likelihood-free inference and state the ABC algorithm. Section 4 provides the numerical aspects in simulating the experiment and gives the prospects of the outlined methodology. We make conclusions in Section 5.

2. Problem statement

Our goal is to analyze the ABC method for quantitative studies of the processes described by strong-field quantum electrodynamics (SFQED). For this purpose, we choose to consider the problem of detecting and measuring the extent of the effective mass shift for the electron due to its coupling with the strong-field environment[10,16–18]. In this section, we first give a rather general description of a potential experiment and then motivate a simplified problem formulation that is sought to serve as a proof-of-principle case. In the following sections we elaborate the application of ABCs and describe how the developed routine can be generalized to deal with realistic experimental conditions.

The presence of a strong background electromagnetic field is conjectured to drive the expansion parameter of QED to $\alpha_f \chi^{2/3}$, where $\alpha_f \approx 1/137$ is the fine-structure constant[17,19]. When $\alpha_f \chi^{2/3} \gtrsim 1$, the theory is rendered nonperturbative.

In this domain, electrons and positrons can be thought to acquire an effective mass as a result of radiative corrections. Specifically, one can show that the effective mass of the electron $\tilde{m}_e$ can be estimated to be as follows[16]:

$$\tilde{m}_e^2 = m_e^2 + \delta m_e^2 = m_e^2 \left(1 + 0.84\alpha_f \chi^{2/3}\right), \quad (2)$$

which implies an effective value of χ (the mass enters Equation (1) through E_{crit}):

$$\tilde{\chi}^{2/3} = \frac{\chi^{2/3}}{1 + 0.84\alpha_f \chi^{2/3}}. \quad (3)$$

To benchmark this effect and determine its extent, one can consider the value of $\theta^{\text{true}} = 0.84$ as a model parameter θ to be inferred from the data of repeated experiments:

$$\tilde{\chi}^{2/3} = \frac{\chi^{2/3}}{1 + \theta\alpha_f \chi^{2/3}}. \quad (4)$$

In general, a chosen parameterization is not necessarily unique and several possibilities can exist. Here, another option is to replace $\chi^{2/3} \rightarrow \chi^\theta$ to test the conjectured power-law behavior that causes QED to become nonperturbative.

In replacing the quantities $m_e \rightarrow \tilde{m}_e, \chi \rightarrow \tilde{\chi}$, the rates of photon emission and pair formation computed from QED being valid at moderate χ become modified. As for the former, the rate can be written as follows[20,21]:

$$\frac{\partial I}{\partial \omega} = \frac{\sqrt{3}\tilde{m}_e q_e^2 c \tilde{\chi} (1 - \delta)}{2\pi \gamma_e \hbar} \left(F_1(\zeta) + \frac{3}{2}\delta\tilde{\chi}\zeta F_2(\zeta)\right), \quad (5)$$

where $\zeta = \frac{2}{3\tilde{\chi}}\frac{\delta}{1-\delta}$, $\delta = \frac{\hbar\omega}{\tilde{m}_e c^2 \gamma_e}$ represents the photon energy with frequency ω normalized to the emitting electron energy and $F_1(x), F_2(x)$ denote the first and second Synchrotron functions defined by the following:

$$F_1(y) = y\int_y^\infty K_{5/3}(y)\mathrm{d}y, \quad F_2(y) = yK_{2/3}(y), \quad (6)$$

in which $K_\nu(y)$ are the modified Bessel functions of the second kind.

Equation (4) indicates that the dependency of the emission rate (Equation (5)) on θ becomes weaker with the decrease of χ and we would ideally need to reach values of the order of $\chi \sim \alpha_f^{-3/2} \approx 1600$ to have a prominent dependency. Since such values are currently unattainable, it is clear that achieving as large values of χ as possible can be crucial. The collision of tightly focused laser pulses with accelerated electron bunching provides a promising layout for these experimental efforts. The use of optimal focusing, provided by the so-called bi-dipole wave, can yield values of χ as large as $5.25\left(\varepsilon/\left(1\,\text{GeV}\right)\right)\left(P/\left(1\,\text{PW}\right)\right)^{1/2}\left(\left(1\,\mu\text{m}\right)/\lambda\right)$, where ε is the energy of electrons, whereas P and λ are the laser power and wavelength, respectively[22]. Nevertheless,

bi-dipole focusing as well as any tight focusing implies a spatial localization of the strong-field region to a size of approximately λ that can be comparable or smaller than the typical shot-to-shot variation of the relative electron beam location provided by current alignment capabilities. This leads to a reasonable question: to what extent can information about θ be retrieved despite a large variation of experimental outcomes due to the varied and unmeasured offset between the laser pulse and electron bunch inherent in current experiments? This difficulty is further complicated by the probabilistic nature of emissions and the dominance of the signal produced by emissions at low values of χ. We try to assess the prospects of overcoming these difficulties by the use of ABCs and to do so we consider the problem of retrieving information about θ from a sequence of measured experimental outcomes assuming that there is an unknown offset for each measurement.

To focus on the outlined difficulty, we make a number of simplifications, while a more realistic problem formulation is to be considered in future works. Firstly, we neglect the generation of pairs, which in practice can be suppressed by the use of short laser pulses[23] and/or the high energy of electrons. We also consider a circularly polarized laser pulse and the use of the angular-energy distribution of produced photons, because in this case high-energy photons emitted toward the directions that deviate the most from the collision direction are predominantly produced at large values of χ[22]. Although the deviation angle is used for our diagnostics, we still assume that the high energy of electrons makes this angle sufficiently small so that the instantaneous transverse deviations of electrons are much less than λ and thus do not affect the observed field amplitude. Moreover, we assume that the laser pulse is sufficiently short not to cause multiple emissions, which means that we assume all electrons to have their initial energy throughout the interaction. Finally, we need to account for the presence of an unknown spatio-temporal offset between the electron beam and focused laser field for each collision. Since such a mismatch leads to a reduced field amplitude observed by the electrons, we choose to model these variations by assuming that the electron beam propagates through a 1D laser pulse with an unknown amplitude that varies from collision to collision. In what follows, we detail this model of hypothetical experiments.

In defining the model of the experiment, we simulate a single electron of momentum $p_z = -m_{\mathrm{e}}c\gamma$ to interact with a plane wave laser pulse having circular polarization. The corresponding vector potential and electric field amplitudes are given by the following:

$$
A_x(\xi) = (1-d)\,A_0 \cdot \frac{1}{4}\left(\frac{-\cos\left(\xi\left(k+2\pi/L\right)\right)}{2\pi+kL} \right.
$$

$$
\left. -\frac{2\cos(k\xi)}{kL} + \frac{\cos\left(\xi\left(k-2\pi/L\right)\right)}{2\pi-kL} \right)\Pi\left(\frac{\xi}{L}\right), \tag{7}
$$

$$
E_x(\xi) = -\frac{\partial A_x(\xi)}{\partial t}, \tag{8}
$$

where $\xi = z - ct$ is the moving coordinate, A_0 is the peak vector potential amplitude, k and L are the wavenumber and pulse length of the laser, respectively, and $\Pi(x)$ is defined as a function equating to unity when $|x| < 1/2$ and zero otherwise. The electromagnetic field amplitude is measured in dimensionless units $a_0 = \frac{q_{\mathrm{e}}A_0}{m_{\mathrm{e}}ckL}$ and $0 \le d \le 1$ expresses the uncontrollable misalignment mentioned previously and is referred to as a latent parameter.

The inclusion of the latent parameter d becomes the primary reason to require the angular-energy spectra of photons as opposed to the energy distribution alone. Indeed, from Equations (1) and (4) one can see that the same value of χ (which completely determines radiation) can be achieved by various combinations of d and θ. For instance, the case of ($\theta = \theta^{\mathrm{true}}$; $d = 0$) yields the same value for χ as the combination of $\theta = 0$ and the following:

$$
d = 1 - \left(1 + \theta^{\mathrm{true}}\alpha_{\mathrm{f}}\chi^{2/3}\right)^{-3/2}. \tag{9}
$$

Because of this degeneracy, inferring the value of θ from the energy spectrum alone is impossible.

Let us now analyze the implications of including the angular part of the spectrum. Firstly, consider the transverse motion of an electron in a plane electromagnetic wave[24]:

$$
\vec{p}_\perp = -q_{\mathrm{e}}\vec{A}_\perp, \tag{10}
$$

where $\vec{p}_\perp$ and $\vec{A}_\perp$ denote the transverse components of the electron momentum and vector potential, respectively. Consequentially, at each instance of time, the electron propagates toward the direction that deviates from the initial direction by an angle α:

$$
\alpha = \arctan\left(\frac{|\vec{p}_\perp|}{|\vec{p}_z|}\right), \tag{11}
$$

where we assume that the motion remains highly relativistic. Evidently, emitted photons retain this angle and, given the circular polarization of the wave, this becomes correlated to the value of χ[22]. We assume that in the case of highly relativistic motion with $\alpha \ll 1$, the change of effective mass ($\theta \ne 0$) does not lead to the change of $\vec{p}_z$ before the emission, while $\vec{p}_\perp$ is totally defined by the vector potential according to Equation (10). Following Equation (11), this means that the value of θ does not affect α. On the contrary, the increase of d implies a diminished electromagnetic field amplitude perceived by the electrons, which leads to smaller angles α. Note that our idea here is to break the described degeneracy, which does not necessarily require α being independent of θ (other cases can be analyzed in a similar way).

Having discussed the necessity for learning from both the energy and angle of emitted photons we define the measured data of our model experiment as a fractional energy distribution $\partial^2 E/\partial\delta\partial\alpha$, which is a function of normalized photon energy δ and deviation angle α. For the computational implementation this function is represented by a matrix x_{obs}, of which the elements $x_{\text{obs}}^{(k)(l)}$ can be computed from the result of a simulation by counting the total energy of photons belonging to each cell: $k\Delta\delta < \delta \leq (k+1)\Delta\delta$; $l\Delta\alpha < \alpha \leq (l+1)\Delta\alpha$, where the choice of cell sizes $\Delta\delta$ and $\Delta\alpha$ is a matter of a trade-off between the level of noise and resolution. The problem formulation is then to infer the value of θ from a sequence of measured x_{obs} despite unknown d in each case.

3. Methodology

We now establish the methodology of ABC sampling using a general problem formulation in which Bayesian statistics is employed for making inferences based on the results from repeated experiments. Starting from Bayes' theorem, the problem of intractable likelihoods is discussed and the appeal for 'likelihood-free' methods is introduced by considering standard rejection sampling. By reformulating the rejection step by imposing an exact agreement between simulated data and that of experiments, one obtains a likelihood-free technique. We shall see that this suffers from the curse of dimensionality and through dimensionality reduction techniques we will arrive at ABC sampling. Lastly, we consider two additional improvements for ABC sampling with the aim to accelerate convergence.

Let us start from considering the task of characterizing a probabilistic process (for instance, nonlinear Compton scattering with an effective electron mass shift) by carrying out experiments. Each experiment yields measurement data x_{obs}. We have a model $M(\theta, d)$ that gives predictions $x = M(\theta, d)$ for this data for any given value of a model parameter θ and a latent parameter d. Here, θ is a fundamental parameter that quantifies the process itself and thus its unique value is of interest, whereas d denotes an unmeasured parameter that can vary from experiment to experiment and determines the outcome x in accordance with model M. We assume that there exists a value of θ for which the model describes (to some extent) observations given an appropriate value of d for each experiment. Our task is to infer the probability distribution for the value of θ from a series of repeated experimental measurements. Put differently, the objective is to infer the most probable range for θ given the observed data x_{obs}. Bayesian statistics provides a framework for the outlined problem. The probability distribution to be determined is referred to as a posterior distribution $p(\theta|x_{\text{obs}})$, which explicitly indicates the data x_{obs} used for making the inference. Let us start from the case of no latent parameter. The posterior can then be calculated using Bayes' theorem:

$$p(\theta|x_{\text{obs}}) = \frac{p(x_{\text{obs}}|\theta) \cdot p(\theta)}{p(x_{\text{obs}})}, \qquad (12)$$

where $p(\theta)$ quantifies the prior knowledge about possible values of θ, the likelihood $p(x_{\text{obs}}|\theta)$ conveys how likely a measurement yielding x_{obs} is for a given θ and $p(x_{\text{obs}}) = \int p(x_{\text{obs}}|\theta)p(\theta)\,\partial\theta$ appears as a normalizing factor. To incorporate the dependence on the latent parameter we integrate over all its possible values, denoting $p(x_{\text{obs}}|\theta, d)$ as the corresponding joint likelihood:

$$p(\theta|x_{\text{obs}}) = \frac{\int p(x_{\text{obs}}|\theta, d)p(d)\partial d \cdot p(\theta)}{\iint p(x_{\text{obs}}|\theta, d)p(d)p(\theta)\,\partial d\partial\theta}, \qquad (13)$$

where $p(d)$ specifies prior knowledge related to values of the latent parameter d. Now we can sequentially account for all observations, each time using the obtained posterior as the prior for processing the next observation. Note that we do not update the prior for d because its value is assumed to be different in all the experiments.

A closed form of the posterior rarely exists and numerical approaches are often used. A common strategy is to approximate the posterior by collecting a finite number of samples from it. Methods such as importance sampling, Markov chain Monte Carlo (MCMC) and sequential Monte Carlo (SMC)[25–27] are prevalent choices. However, all of the above will require direct evaluation of the likelihood, which can be computationally prohibitive for highly dimensional datasets[28]. If the model M is implicitly defined through a computer simulation, its concomitant likelihood can be intractable[14]. A remedy is offered by the rapidly developing field of simulation-based inference[29] in which the direct calculation of the likelihood is averted. To motivate its use we adopt and develop the discussion made in Ref. [28].

Consider the standard rejection sampling algorithm with the goal of sampling a target density $T(\theta)$ provided some auxiliary sampling density $A(\theta)$ with the requirement $A(\theta) > 0$ if $T(\theta) > 0$. Then, the algorithm reads as follows.

Algorithm 1. Standard rejection sampling algorithm.

1: Sample a proposal $\theta^* \sim A(\theta)$.

2: Admit the proposal with a probability of $\frac{T(\theta^*)}{CA(\theta^*)}$, where $C \geq \text{argmax}\left(\frac{T(\theta)}{A(\theta)}\right)$.

3: If θ^* was not admitted, discard the proposal and repeat steps (1) and (2) as many times as necessary.

After N trials, a collection of samples from $T(\theta)$ is obtained. The connection to Bayesian statistics is made by selecting $T(\theta) = p(\theta|x_{\text{obs}})$ and $A(\theta) = p(\theta)$. Then, Equation (12) states that the acceptance rate appearing in Algorithm 1 becomes proportional to the likelihood

$\frac{p(\theta^*|x_{\text{obs}})}{p(\theta^*)} \propto p(x_{\text{obs}}|\theta^*)$, which is incalculable by our premise. Still, it is possible to determine whether or not to accept proposals without explicit computation of the likelihood. To demonstrate this, we first note that the model $M(\theta, d)$ is capable of generating samples of observations $x \sim p(x_{\text{obs}}|\theta, d)$ provided values of θ and d. Then, the key aspect is to recognize that the probability to produce $x = x_{\text{obs}}$ coincides with $p(x_{\text{obs}}|\theta, d)$. Put differently, calculating the acceptance rate from the likelihood can be replaced by enforcing an exact match between the datasets, meaning that a likelihood-free approach has been found. Therefore, the rejection step of Algorithm 1 is reformulated to read as follows.

Algorithm 2. Likelihood-free rejection sampling.

1: Sample proposals $\theta^* \sim p(\theta)$, $d^* \sim p(d)$.

2: Generate data $x^* = M(\theta^*, d^*)$ from the model.

3: If $x^* = x_{\text{obs}}$, the proposal is admitted; if not, it is discarded.

4: Repeat steps (1)–(3) as many times as necessary.

While avoiding direct computation of the likelihood, demanding $x^* = x_{\text{obs}}$ introduces a notable impediment. To illustrate it, consider the binning of data from an experiment into $\dim(x_{\text{obs}}) = B$ bins such that the following are obtained:

$$x_{\text{obs}} = [c_1, c_2, c_3, \ldots, c_B], \tag{14}$$

$$x = [c'_1, c'_2, c'_3, \ldots, c'_B], \tag{15}$$

where $c_b, c'_b \in \mathbb{Z}$ denote integer counts belonging to the bth bin. Then, denote p_b as the probability to coincide $c_b = c'_b$ at bin b, assuming that this is independent between bins. Then, the probability to accept a proposal θ^* becomes as follows:

$$p(x = x_{\text{obs}}) = \prod_{b=1}^{b=B} p_b, \tag{16}$$

which approaches zero in the limit of highly dimensional datasets, $B \to \infty$. The acceptance rate becomes even lower or infeasible for continuous data in which $c_b, c'_b \in \mathbb{R}$ are real numbers. Hence, the appeal for a precise match has to be relieved in making the sampling efficiency practical. Realizing that this rate becomes significantly higher by admitting samples if $x \approx x_{\text{obs}}$ prompts us to define a rule when data are sufficiently close:

$$\|x - x_{\text{obs}}\| \le \epsilon, \tag{17}$$

where $\|\cdot\|$ is a suitable distance metric and ϵ is a threshold. Accepted samples in accordance with Equation (17) are

inevitably drawn from an approximate posterior $p(\theta|x_{\text{obs}})$ and its accuracy is solely dictated by ϵ, which also affect the sampling efficiency. Rejection based on Equation (17) provides a notable improvement, but high dimensionality remains an issue. Consider the aforementioned example with an Euclidean distance metric so that Equation (17) reads as follows:

$$\left(\sum_{b=1}^{b=B} (c_b - c'_b)^2 \right)^{1/2} \le \epsilon, \tag{18}$$

and examine the favorable case in which $c_b - c'_b \sim \Delta_c \ll 1$ varies negligibly between bins. We can then naively state Equation (18) as follows:

$$\dim(x_{\text{obs}}) \le (\epsilon / \Delta_c)^2. \tag{19}$$

Evidently, Equation (19) states that the dimension of x_{obs} is bounded from above by the threshold ϵ and the error Δ_c. However, for the quality of inference $\epsilon \to 0$ is desired, which puts a stringent limit on the dimensionality of x_{obs}. The solution, being the central component of ABCs, is to introduce so-called summary statistics:

$$S : \mathbb{R}^B \mapsto \mathbb{R}^\beta, \tag{20}$$

being a function that transforms data of a potentially noisy nature into a vector of indicative characteristics sought to unambiguously characterize the data with respect to all possible θ. Clearly, the dimensionality β of the space of such vectors can be much less than the number of cells B. Moreover, the function of summary statistics can even be defined in an agnostic way with respect to the binning choice. As an example, one could construct a vector containing the sample mean μ and variance σ^2 of x_{obs}: $S(x_{\text{obs}}) = (\mu, \sigma^2)$.

By converting $x_{\text{obs}} \to S(x_{\text{obs}})$, the third step of Algorithm 2 can be reformulated to accept samples if the following applies:

$$\|S(x) - S(x_{\text{obs}})\| \le \epsilon. \tag{21}$$

Note that the mapping $x_{\text{obs}} \to S(x_{\text{obs}})$ into a vector of summary statistics is always possible in terms of reducing the data dimension. The question is how well the chosen vector of summary statistics retains the information contained in x_{obs}, which is discussed later in Section 4.

We now have a methodologically accurate and in some cases practically feasible routine for sampling the posterior. However, we shall examine two more improvements for accelerating the sampling convergence. Firstly, note that Equation (21) implies an acceptance probability of either zero or one without accounting for how close the match is. To enhance the contribution of the cases yielding a more accurate agreement relative to the ones giving a marginal

agreement, one can use a so-called kernel function:

$$K_\epsilon : \mathbb{R}^\beta \mapsto \mathbb{R} = K_\epsilon \left(\frac{\|S(x) - S(x_{\text{obs}})\|}{\epsilon} \right), \qquad (22)$$

which defines a probability transition from one in case of a perfect match ($K_\epsilon(0) = 1$) to zero in cases of deviation by the summary-statistics distance of order ϵ and greater.

The second improvement concerns the fact that Algorithm 1 either accepts or rejects cases, which means that many accepted cases are needed to mitigate the noise related to this additional probabilistic element in the algorithm. Effectively, this means that we marginally benefit from cases of low acceptance probability. To avoid this, one can instead interpret the acceptance probability as the weight of samples, thereby accounting for all the proposals that yield non-zero acceptance probability.

We can now return back to the inclusion of the latent variable d. In this case, we can generate several proposals $d^* \sim p(d)$ based on our prior knowledge of it and again accept the cases of good enough matches based on the outlined procedure. Effectively, we try to guess d using as many attempts as needed. Finally, we note that we can sequentially update our posterior using each x_{obs} in a sequence of measurements. To do so, we can compute the posterior for each new measurement using the previous posterior as the prior. The algorithm for processing the ith observation ($i = 1$ denotes the first measurement in the sequence) of x_{obs}^i for computing the posterior $p\left(\theta \mid x_{\text{obs}}^i, x_{\text{obs}}^{i-1}, \ldots, x_{\text{obs}}^1\right)$ from the previous $p\left(\theta | x_{\text{obs}}^{i-1}, \ldots, x_{\text{obs}}^1\right)$ then takes the following form.

Algorithm 3. ABC sampling with a latent variable.

1: Sample proposals $\theta^* \sim p\left(\theta | x_{\text{obs}}^{i-1}, \ldots, x_{\text{obs}}^1\right)$, $d^* \sim p(d)$.

2: Perform a simulation, retrieve $x^* = M\left(\theta^*, d^*\right)$ and compute the weight:

$$w^* = \frac{K_\varepsilon \left(\| S\left(x_{\text{obs}}^i\right) - S(x^*) \| / \epsilon \right)}{p\left(\theta^* | x_{\text{obs}}^{i-1}, \ldots, x_{\text{obs}}^1\right) p(z^*)}. \qquad (23)$$

3: If $w^* > 0$, accept the proposal with the computed weight.

4: Repeat steps (1)–(3) as many times as needed to approximate the posterior $p\left(\theta | x_{\text{obs}}^i, \ldots, x_{\text{obs}}^1\right)$.

In practice, one central difficulty of the ABC routine is choosing valid summary statistics, that is, summary statistics that differentiate all the cases in terms of θ and d. This means that summary statistics do not yield close states for any two different pairs of θ and d. Clearly, if this is not the case the procedure admits the acceptance of cases of wrong θ^* when d^* provides a compensation to make $S(x^*(\theta^*, d^*))$
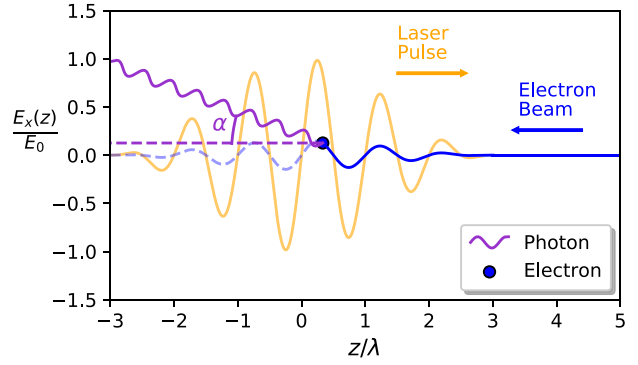

Figure 1. Representation of the numerical implementation of the experiment in which E_0 denotes the peak electric field amplitude (the deviation angle is exaggerated).

close to $S\left(x_{\text{obs}}\left(\theta^{\text{true}}, d^{\text{true}}\right)\right)$. This can totally preclude the convergence of the ABC sampling procedure. Finding robust summary statistics is known to be a problem-dependent task that requires analysis of possible cases. In the next section we determine valid summary statistics for the proof-of-principle problem as well as outlining the simulation routines.

4. Analysis

With respect to the simulation, the plane wave pulse is designated by a wavelength of $\lambda = 0.8$ μm, pulse length $L = 6\lambda$ and peak amplitude $a_0 = 100$ (excluding the factor of $1 - d$). Electrons are assigned an initial energy of 170 GeV ($\gamma_e \sim 10^5$) situated a distance $z_s = 5\lambda$ from the origin (the numerical layout can be seen in Figure 1). Both the electron and pulse are allowed to counter propagate for $N_t = 100$ time steps $\Delta t = \frac{(L+z_s/2)c^{-1}}{N_t}$. Here, $x(\delta, \delta + \Delta\delta, \alpha, \alpha + \Delta\alpha)$ is discretized into a 100×100 grid of cells $x(m\Delta\delta, n\Delta\alpha)$ each with size $\Delta\delta \times \Delta\alpha$ and $m, n = 0, 1, 2, \ldots, 99$. At each time step q, Equations (10) and (11) are used to estimate the index $n \approx \alpha/\Delta\alpha$. Then, for each m we accumulate the following:

$$x_{q\Delta t} = x_{(q-1)\Delta t} + \Delta\alpha \Delta\omega \Delta t \frac{\partial I}{\partial \omega}(m\Delta\delta, \theta), \qquad (24)$$

where we have suppressed the arguments of x for readability and the subscripts denote the time step. In this analysis, we perform blind tests of the simulated cases against an 'experiment' $x_{\text{obs}} = M\left(\theta^{\text{true}}, d\right)$ that serves as a ground truth. Here, the θ value is fixed to $\theta^{\text{true}} = 0.84$ but the latent variable d varies randomly between experiments.

It was concluded in Section 2 that the fractional energy distribution x_{obs} needs to be measured in order to infer the effect of the effective mass shift. For this to serve as input to ABC sampling, corresponding summary statistics must be chosen to map x_{obs} onto a smaller subspace while still retaining the information necessary to separate the effects of θ and d. There might exist many configurations that achieve
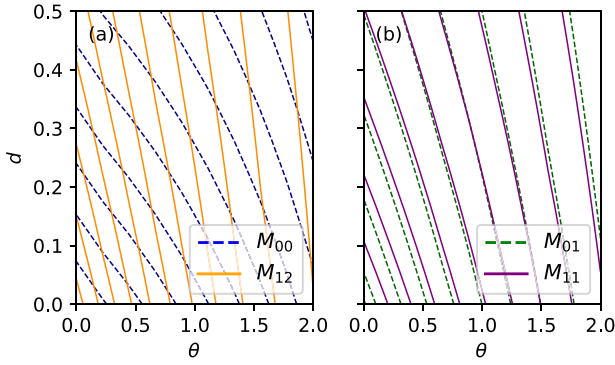
Figure 2. Contours of M_{ij} as a function of θ and d where (a) compares M_{00} and M_{12} and (b) compares M_{01} and M_{11}.

this given the broad definition of Equation (20). However, it is also beneficial to select a mapping that is tolerant (insensitive) to the probabilistic random variations of the experimental outcome. To identify some robust and simple enough options we consider moments of the 2D data x_{obs} to orders i and j:

$$M_{ij} = \sum_k \sum_l x_{\text{obs}}^{(k)(l)} (k\Delta\delta)^i (l\delta\alpha)^j \Delta\delta\Delta\alpha. \qquad (25)$$

Now, let us try to select a set of moments such that any combination (θ, d) maps to a unique value of this set. To select a valid option we analyze the dependency of several of the lowest distribution moments on θ and d. Figure 2 illustrates contours of four distinct moments M_{ij} in the parameter space of θ and d. The set of moments in Figure 2(a) is a practical choice as the contours are not parallel anywhere, suggesting a unique pair for every θ and d. In contrast, Figure 2(b) depicts a scenario when the contours become parallel at several points in the parameter space, meaning that the values of the plotted moments do not unambiguously indicate a single pair of θ and d. For instance, the contour lines that intersect the x-axis at $\theta \approx 1.25$ imply that the values for (M_{01}, M_{11}) are almost identical along this line, which holds for many pairs (θ, d). We conclude that selecting $S(x_{\text{obs}}) = (M_{00}, M_{12})$ is a valid choice to proceed with ABC sampling.

Note that for a larger number of model parameters (representing both theory inquiries and latent parameters of collisions) one would need to inspect the dependency of summary statistics in a multidimensional space, which can be unfeasible. In this case one might need physics insights to make an educated strategy to disentangle specific types of degeneracy. In the considered case we can interpret the identified choice through the following logic. The value of M_{00} indicates the overall rate of emission that does not differentiate the effect of θ and d. Nevertheless, the value of M_{12} reflects the change of the spectrum shape (mean energy) with angle α. This is exactly what is needed because, due to circular polarization, the emissions with large α happen

exclusively at large values of χ (for details, see Ref. [22]) and have a shape altered due to the modification of χ (see Equation (4)) to the extent defined by θ.

Let us now continue the analysis for the chosen summary statistics. We make the following choice of priors over θ and d:

$$p(\theta) = \mathcal{U}(0, 150), \ p(d) = \mathcal{U}(0, 0.5), \qquad (26)$$

where $\mathcal{U}(a, b)$ denotes the uniform distribution with lower and upper bounds a and b, respectively. Although there is no prior knowledge apart from $\theta \geq 0$ and $0 \leq d \leq 1$, we argue that the given simulation parameters yield $\chi_0 \approx 100$ and a value of $\theta = 150$ would reduce the value of $\tilde{\chi}_0$ below one, approaching a classical description. For each round of sampling, the prior is replaced by the obtained posterior as described in Algorithm 3. For instance, the new prior obtained from the first observation (and those following it) is modeled as a normal distribution $\pi(\theta) \rightarrow p(\theta|x_{\text{obs}}^1) \approx \mathcal{N}(\mu_\theta, \sigma_\theta)$, where μ_θ and σ_θ are the weighted mean and standard deviation of the accepted samples, respectively. As for the prior over d, one could construct a prior from empirical values obtained in a real experiment. Lacking this option, we assume that the amplitude can vary at most by 50%. Allowing d to vary as widely as 100% will not affect the final result. Instead, the sampling efficiency will be reduced as larger values of d are allowed, potentially affecting the dataset to the extent that proposals are rejected more often.

During sampling, the following distance is calculated to discriminate between observations:

$$\|S(x) - S(x_{\text{obs}})\| = \sqrt{d_{00}^2 + d_{12}^2}, \qquad (27)$$

where $d_{ij} = |1 - \frac{M_{ij}^{\text{sim}}}{M_{ij}}|$ (not to be confused with the latent parameter) in which the superscript labels moments evaluated from simulated data. A uniform kernel $K_\varepsilon(\cdot) = \Pi(\cdot)$ is chosen as well as a threshold that is progressively decreased from $\epsilon = 1.1$ to $\epsilon = 0.065$ over the course of the processed observations x_{obs} included in the analysis. Here, the choice of ϵ is manually selected to yield a narrower posterior for each processed observation x_{obs} while maintaining a feasible sampling duration. For every 50th proposal θ^* we generate new observed data x_{obs} as to not bias the result toward the existing value of $d^* \sim p(d)$.

In Figure 3 we present the result of sampling the posterior based on the described ABC routine after processing $i = 1, 5, 8$ observations x_{obs}. Each observation corresponds to the simulated outcome of a single collision experiment with unknown value of d and fixed $\theta = \theta^{\text{true}}$. The fact that the accepted samples are distributed around the actually selected value of $\theta^{\text{true}} = 0.84$ indicates the claimed capability of the
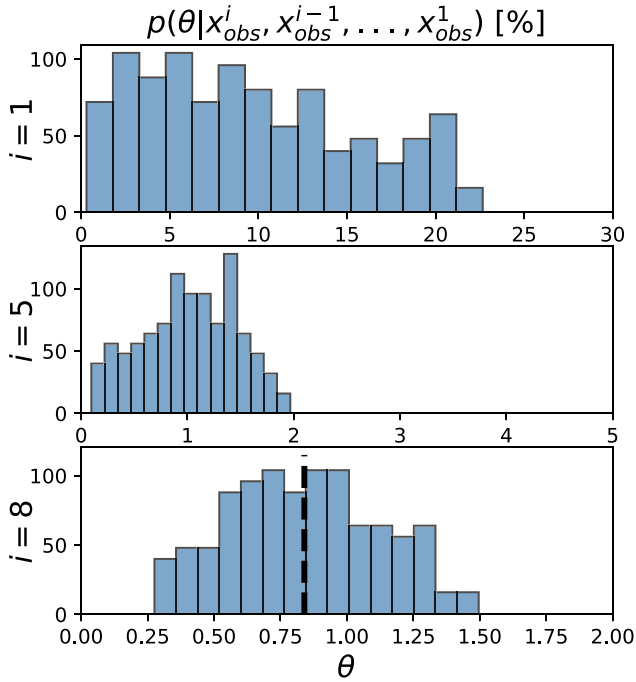
Figure 3. The approximate posterior after processing $i = 1, 5, 8$ observations in line with Algorithm 3, where in the bottom panel the dashed line demarcates the sample mean. All distributions have $N_\theta = 1000$ accepted proposals.

method. Attaining a narrower posterior can be achieved by reducing the threshold ϵ.

Having discussed the use of ABCs via the selected proof-of-principle case, we can outline further developments that could help to deal with real experiments. The first direction of developments concerns the elaboration of an appropriate model of experiments, which implies a relevant 3D geometry of interaction with an account of unmeasured variations via an extended list of latent parameters, such as all spatio-temporal offsets, as well as parameters quantifying both the electron bunch and the focused laser pulse. The second direction of developments concerns computational methods to deal with the chosen geometry of the experiment, as well as accounting for all physical processes of relevance, such as the generation of electron–positron pairs and the interaction between particles and photons. Efficient use of supercomputer resources can be crucial to compensate for the increase of computational demands in relation to both the complexity of simulations and the necessity to obtain a meaningfully narrow posterior using a large number of collisions. The third direction of developments concerns improvements related to the use of the ABC itself. This may include a better choice of summary statistics, a modification of diagnostics or the experiment layout, as well as reducing the rejection rate by employing machine learning methods for an informed generation of proposals. Finally, the fourth direction of developments concerns the very formulation of theoretical questions, which takes the form of determining models and parameters to be inferred from experiments.

5. Conclusions

We have considered prospects for an experiment capable of inferring a parameter θ that signifies deviations from nonlinear Compton scattering via the notion of effective mass in the regime $\chi \gg 1$. The results propel the strategies necessary to incorporate ABC sampling in analogous experiments, scalable to the inclusion of several model parameters (θ), as well as latent parameters that are sought to account for unmeasured shot-to-shot variations in the conditions of collisions. An improved implementation of the interaction will be needed for designing future experiments. This can be done by, for example, simulating a realistically focused laser pulse, devising a more comprehensive description via latent parameters and accounting for electromagnetic cascades. Carrying it out might pose an increased computational load that can be mitigated by further developments. The convergence can be accelerated by further investigating additional summary statistics, nonuniform kernels and the use of machine learning to suggest better proposals. In addition, the use of high-performance computing to recruit many ABC samplers in parallel can alleviate both impairments.

References

1. J. M. Cole, K. T. Behm, E. Gerstmayr, T. G. Blackburn, J. C. Wood, C. D. Baird, M. J. Duff, C. Harvey, A. Ilderton, A. S. Joglekar, K. Krushelnick, S. Kuschel, M. Marklund, P. McKenna, C. D. Murphy, K. Poder, C. P. Ridgers, G. M. Samarin, G. Sarri, D. R. Symes, A. G. R. Thomas, J. Warwick, M. Zepf, Z. Najmudin, and S. P. D. Mangles, Phys. Rev. X **8**, 011020 (2018).
2. K. Poder, M. Tamburini, G. Sarri, A. Di Piazza, S. Kuschel, C. D. Baird, K. Behm, S. Bohlen, J. M. Cole, D. J. Corvan, M. Duff, E. Gerstmayr, C. H. Keitel, K. Krushelnick, S. P. D. Mangles, P. McKenna, C. D. Murphy, Z. Najmudin, C. P. Ridgers, G. M. Samarin, D. R. Symes, A. G. R. Thomas, J. Warwick, and M. Zepf, Phys. Rev. X **8**, 031004 (2018).
3. H. Abramowicz, U. Acosta, M. Altarelli, R. Aßmann, Z. Bai, T. Behnke, Y. Benhammou, T. Blackburn, S. Boogert, O. Borysov, M. Borysova, R. Brinkmann, M. Bruschi, F. Burkart, K. Büßer, N. Cavanagh, O. Davidi, W. Decking, U. Dosselli, N. Elkina, A. Fedotov, M. Firlej, T. Fiutowski, K. Fleck, M. Gostkin, C. Grojean, J. Hallford, H. Harsh, A. Hartin, B. Heinemann, T. Heinzl, L. Helary, M. Hoffmann, S. Huang, X. Huang, M. Idzik, A. Ilderton, R. Jacobs, B. Kämpfer, B. King, H. Lahno, A. Levanon, A. Levy, I. Levy, J. List,

W. Lohmann, T. Ma, A. J. Macleod, V. Malka, F. Meloni, A. Mironov, M. Morandin, J. Moron, E. Negodin, G. Perez, I. Pomerantz, R. Pöschl, R. Prasad, F. Quéré, A. Ringwald, C. Rödel, S. Rykovanov, F. Salgado, A. Santra, G. Sarri, A. Sävert, A. Sbrizzi, S. Schmitt, U. Schramm, S. Schuwalow, D. Seipt, L. Shaimerdenova, M. Shchedrolosiev, M. Skakunov, Y. Soreq, M. Streeter, K. Swientek, N. T. Hod, S. Tang, T. Teter, D. Thoden, A. I. Titov, O. Tolbanov, G. Torgrimsson, A. Tyazhev, M. Wing, M. Zanetti, A. Zarubin, K. Zeil, M. Zepf, and A. Zhemchukov, Eur. Phys. J. Spec. Top. **230**, 2445 (2021).

4. V. Yakimenko, L. Alsberg, E. Bong, G. Bouchard, C. Clarke, C. Emma, S. Green, C. Hast, M. J. Hogan, J. Seabury, N. Lipkowitz, B. O'Shea, D. Storey, G. White, and G. Yocky, Phys. Rev. Accel. Beams **22**, 101301 (2019).

5. A. Gonoskov, T. Blackburn, M. Marklund, and S. Bulanov, Rev. Mod. Phys. **94**, 045001 (2022).

6. C. Bula, K. T. McDonald, E. J. Prebys, C. Bamber, S. Boege, T. Kotseroglou, A. C. Melissinos, D. D. Meyerhofer, W. Ragg, D. L. Burke, R. C. Field, G. Horton-Smith, A. C. Odian, J. E. Spencer, D. Walz, S. C. Berridge, W. M. Bugg, K. Shmakov, and A. W. Weidemann, Phys. Rev. Lett. **76**, 3116 (1996).

7. M. Iinuma, K. Matsukado, I. Endo, M. Hashida, K. Hayashi, A. Kohara, F. Matsumoto, Y. Nakanishi, S. Sakabe, S. Shimizu, T. Tauchi, K. Yamamoto, and T. Takahashi, Phys. Lett. A **346**, 255 (2005).

8. T. Kumita, Y. Kamiya, M. Babzien, I. Ben-Zvi, K. Kusche, I. V. Pavlishin, I. V. Pogorelsky, D. P. Siddons, V. Yakimenko, T. Hirose, T. Omori, J. Urakawa, K. Yokoya, D. Cline, and F. Zhou, Laser Phys. **16**, 267 (2006).

9. T. Englert and E. Rinehart, Phys. Rev. A **28**, 1539 (1983).

10. A. Fedotov, A. Ilderton, F. Karbstein, B. King, D. Seipt, H. Taya, and G. Torgrimsson, arXiv:2203.00019 (2022).

11. T. Ritto, S. Beregi, and D. Barton, Mech. Syst. Sig. Process. **181**, 109485 (2022).

12. M. C. Kennedy and A. O'Hagan, J. Roy. Stat. Soc. Ser. B **63**, 425 (2001).

13. D. N. DeJong, B. F. Ingram, and C. H. Whiteman, J. Business Econ. Stat. **14**, 1 (1996).

14. J. Brehmer, G. Louppe, J. Pavez, and K. Cranmer, Proc. Natl. Acad. Sci. U. S. A. **117**, 5242 (2020).

15. J. Akeret, A. Refregier, A. Amara, S. Seehars, and C. Hasner, J. Cosmol. Astropart. Phys. **2015**, 043 (2015).

16. V. Yakimenko, S. Meuren, F. Del Gaudio, C. Baumann, A. Fedotov, F. Fiuza, T. Grismayer, M. J. Hogan, A. Pukhov, L. O. Silva, and G. White, Phys. Rev. Lett. **122**, 190404 (2019).

17. V. Ritus, Sov. Phys. JETP **30**, 052805 (1970).

18. S. Meuren and A. Di Piazza, Phys. Rev. Lett. **107**, 260401 (2011).

19. N. B. Narozhny, Phys. Rev. D **21**, 1176 (1980).

20. V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Quantum Electrodynamics*, Vol. 4 (Butterworth-Heinemann, 1982).

21. V. Baier and V. Katkov, Phys. Lett. A **25**, 492 (1967).

22. C. Olofsson and A. Gonoskov, Phys. Rev. A **106**, 063512 (2022).

23. D. E. Rivas, A. Borot, D. E. Cardenas, G. Marcus, X. Gu, D. Herrmann, J. Xu, J. Tan, D. Kormin, G. Ma, W. Dallari, G. D. Tsakiris, I. B. Földes, S. W. Chou, M. Weidman, B. Bergues, T. Wittmann, H. Schröder, P. Tzallas, D. Charalambidis, O. Razskazovskaya, V. Pervak, F. Krausz, and L. Veisz, Sci. Rep. **7**, 5224 (2017).

24. L. D. Landau, *The Classical Theory of Fields*, Vol. 2 (Elsevier, 2013).

25. S. T. Tokdar and R. E. Kass, Wiley Interdisc. Rev. Comput. Stat. **2**, 54 (2010).

26. A. Doucet, N. D. Freitas, and N. Gordon, in *Sequential Monte Carlo Methods in Practice* (Springer, 2001), pp. 3–14.

27. S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo* (CRC Press, 2011).

28. S. A. Sisson, Y. Fan, and M. Beaumont, *Handbook of Approximate Bayesian Computation* (CRC Press, 2018).

29. K. Cranmer, J. Brehmer, and G. Louppe, Proc. Natl. Acad. Sci. U. S. A. **117**, 30055 (2020).