

MIR laser CEP estimation using machine learning concepts in bulk high harmonic generation

BALÁZS NAGYILLÉS^{1,2,†}, GERGELY N. NAGY^{1,2,†}, BÁLINT KISS¹,
ERIC CORMIER^{1,3}, PÉTER FÖLDI^{1,2}, KATALIN VARJÚ^{1,2}, SUBHENDU
KAHALY¹, MOUSUMI UPADHYAY KAHALY^{1,2,*} AND ZSOLT DIVEKI^{1,**}

¹ELI ALPS, ELI-HU Non-Profit Ltd., Wolfgang Sandner utca 3., Szeged 6728, Hungary

²Institute of Physics, University of Szeged, Dóm tér 9, H-6720 Szeged, Hungary

³Laboratoire Photonique et Nanosciences (LP2N), UMR 5298, CNRS-IOSG-Université
Bordeaux, 33400 Talence, France

[†]Authors contributed equally to this work

*mousumi.upadhyaykahaly@eli-alps.hu

**zsolt.diveki@eli-alps.hu

Abstract: Monitoring the carrier-envelope phase (CEP) is of paramount importance for experiments involving few cycle intense laser fields. Common measurement techniques include f-2f interferometry or stereo-ATI setups. These approaches are adequate, but are challenging to implement on demand, at different locations as additional metrology tools, in intense few cycle laser-matter interaction experiments, such as those prevalent in sophisticated user beamlines. In addition there are inherent difficulties for CEP measured at non-conventional laser wavelengths (like e.g. mid infrared) and measurements above 10 kHz laser repetition rates, on single shot basis. Here we demonstrate both by simulations and by experiments a machine learning (ML) driven method for CEP estimation in the mid infrared, which is readily generalizable for any laser wavelength and possibly up to MHz repetition rates. The concept relies on the observation of the spectrum of high harmonic generation (HHG) in bulk material and the use of ML techniques to estimate the CEP of the laser. Once the ML model is trained, the method provides a way for cheap and compact real-time CEP tagging. This technique can complement the otherwise sophisticated monitoring of CEP, and is able to capture the complex correlation between the CEP and the observable HHG spectra.

1. Introduction

High harmonic generation (HHG) relies on the highly non-linear interaction between ultrashort intense laser pulses and matter [1], and has been demonstrated with a wide range of driving laser wavelengths [2–4]. Mid-infrared (MIR) driving lasers in particular have two appealing features. On the one hand, because of the wavelength scaling of the ponderomotive energy [2] $U_p \propto I_L \lambda^2$, the cut-off energy of the generated harmonics in gas can be extended to the keV spectral regime [2, 4, 5], providing X-rays with unmatched temporal and spatial qualities. On the other hand, because of its low linear excitation rate, ultrashort MIR lasers can perform HHG in transparent solid state media [6, 7], like semiconductors or dielectrics, well below the damage threshold of the material [8, 9].

In the few cycle regime, the carrier-envelope phase (CEP) of an ultrashort laser, which indicates the phase of the carrier wave with respect to the peak of the intensity envelope of the pulse, has significant impact on the strong field interaction and the HHG process. Therefore accurate measurement/monitoring and proper control of the CEP is of paramount importance, and is essential for multiple applications in attoscience relevant to chemistry [10], atomic and molecular physics [11] and lightwave electronics [12], to mention a few.

There are different ways to measure the CEP of few cycle pulses, utilizing different types of interactions, for example, like observation of half cycle cut off in HHG spectra from gas [13] or quantum interference in semiconductors [14, 15]. Nonetheless, two other techniques have become

predominant. The first, one uses f-2f interferometry where a large bandwidth fundamental spectrum overlaps with its second harmonic signal and the appearance of the spectral fringes reveals the relative CEP of the laser [16]. Single-shot CEP measurement above 10 kHz is very challenging with f-2f, although several methods allow one to reach MHz repetition rates [17–19] and the overlap between the fundamental and the second harmonic’s spectral region might lie outside some common detectors. The second, technique is based on the measurement of stereographic above threshold ionization (Stereo-ATI) signal [20], which is capable of performing single-shot absolute measurements at high repetition rate [21]. But, in spite of its advantages, Stereo-ATI needs sophisticated and expensive instrumentation and is an in-vacuum CEP metrology tool. The last point makes it difficult to integrate into and permanently keep in place as a metrology tool inside the sophisticated existing high repetition rate attosecond beamlines like those existing at ELI-ALPS [22].

Recently new techniques relying on solid HHG were proposed to measure the (relative) CEP of the laser [23] where the overlap between adjacent harmonic orders was exploited. The technique could be scaled to longer driving wavelengths, however the appearance of the interference pattern between adjacent harmonics is not a general condition. If the latter condition is not fulfilled an alternative approach was proposed providing the CEP stability shot-to-shot [24] or time domain based electric field reconstruction in solids using a delayed replica of the driving laser to perturb the HHG process [25].

During the last decade the applications of machine learning techniques have slowly entered not just into the every day life but into various fields of science, too. At XFELs [26] it was successfully applied to improve and accelerate the metrology of the emitted radiation. Convolutional neural networks have been applied to reconstruct the temporal shape of femtosecond laser pulses [27]. Interestingly, their model was robust enough to retrieve the spectral amplitude and phase from experimental second harmonic generation-frequency resolved optical gating spectrograms while being trained only with simulated spectrograms. Recently, independently from our study, there was a theoretical proposition [28] to reconstruct the band structure of a crystal and at the same time characterize the driving few-cycle laser pulse in the solid HHG process, including both their chirp and CEP, relying on the training of deep neural network models. In another theoretical study [29], the CEP dependence of the solid HHG was combined with deep learning models to retrieve the band structure of MgO crystal.

In this report, we combine the benefits of solid HHG and machine learning in order to propose a concept that helps tagging the driving laser’s relative CEP in a simple setup that is instrumentally not demanding while still offering high repetition rate tagging. First, we demonstrate the feasibility of such tagging relying on simulated harmonic spectra from thin ZnO crystal based on a simple 1D model for bulk ZnO [30], that correctly reproduces its semiconducting features, while proving useful for accurate retrieval of the spectral phases of the incident few-cycle pulse, despite some amplitude noise and phase jitter. Then we show that even without large training data the machine learning model can still achieve good relative CEP estimations. Next, relying on our estimations on the number of needed training data we experimentally demonstrate that solid HHG spectra is indeed a good indicator of the relative CEP of the driving laser.

2. Formulation of our approach

In most of the ultrafast light matter interactions, including solid HHG as well, the temporal/spectral profile of the driving laser exhibit a deterministic influence on the outcome. In order to enable exploitation of this feature and utilize it for laser CEP estimations, two prerequisites need to be fulfilled. Firstly, there should be a direct, but not necessarily obvious or even explicit, correlation between the harmonic spectrum and the laser CEP in solid HHG. Secondly, this correlation should manifest as a one-to-one mapping, ensuring that the high harmonic spectrum of a CEP scan exhibits a periodicity of 2π , otherwise CEP prediction would be limited for only on a

fraction of the full range.

In the case of few cycle lasers, where the amplitude of the electric field changes substantially from one half cycle to the other, interband harmonics are generated at different times in each cycle [31, 32]. This causes a phase shift (attochirp) between the same harmonics originating from different half cycles. By changing the CEP of the laser, the timing of the emission of given harmonics can be directly controlled, causing the emergence of CEP dependent patterns in the harmonic spectrum [31–33]. Therefore, the first requirement, the presence of a CEP-dependent feature in the spectrum, is fulfilled. While CEP dependence of harmonic generation through intraband processes in mid-infrared regime is claimed to be insignificant [34, 35], with increasing driving wavelength, signatures of CEP dependence of intraband harmonics become more prominent [33, 36]. The symmetry properties of the target material also have influence on the CEP response of the harmonics. When the interatomic structure is randomly distributed in the crystal on a scale smaller than the laser wavelength, the harmonic spectrum will exhibit π periodicity versus the CEP change, because each half-cycle will experience the same average response, like in the case of fused silica [34]. Crystalline quartz possesses a non-centro-symmetric crystal structure, which results in an absence of inversion symmetry, thereby leading to a nonzero nonlinear susceptibility tensor. The broken inversion symmetry also ensures that the consecutive laser half-cycles experience a different collective response, leading to 2π CEP dependence.

By assuring interband contributions to the harmonic generation in a broken inversion symmetry material one paves the way for mapping one-to-one the laser CEP to the generated high harmonic spectrum. This objective is achieved through the utilization of a representative theoretical model for bulk ZnO that incorporates interband [36] contributions (driving with a MIR laser) to harmonic generation and possesses C_{6v} symmetry group, through a 2-band model. Then, a machine learning algorithm can be trained with known CEP and solid HHG spectrum pairs to estimate the CEP of a new laser pulse based on the generated harmonic spectrum, as depicted in Figure 1. First, we test and validate this assumption by simulating solid HHG spectra generated from ZnO crystal, applying theoretically constructed laser pulses with known absolute CEP parameter. Subsequently, we demonstrate that the effectiveness of this approach does not depend on an extensive dataset, thereby highlighting its practicality.

In the present work, we select three machine learning algorithms: Linear Regression, Extremely Randomized Trees (ExtraTree) and Gradient Boosting. The linear regressor, the simplest one of all three, works by assuming a linear relationship between the predictor values X_j and the predicted value Y , expressed as $Y = \beta_0 + \sum_j \beta_j X_j$. Major advantages of this algorithm are simplicity and fast training. The simple dependence on the β_j coefficients enables a direct estimation of the significance of certain features. However, the model only performs well if there is a direct linear correspondence between X_j and Y . [37, 38]

The ExtraTree and Gradient Boosting algorithms are both decision tree based ensemble models. In a conventional decision tree model, a tree is formed by segmenting the dataset into branches. The model makes decisions at each node based on feature values, eventually arriving to a prediction at the end-points of the branches. In the ExtraTree model, multiple decision trees are built and used concurrently, and the prediction is decided by averaging the prediction of the independent trees. Different trees are grown from different parts of the training dataset by employing bagging, which decreases the bias and variance in the model compared to a single decision tree. The trees are built by randomly selecting a subset of features in the data, and using random threshold values (as opposed to a certain criteria, such as information gain) to split the nodes of the tree. [39]. This high degree of randomization significantly reduces the bias of the model, and makes it less likely to overfit compared to other ensemble models (such as Random Forest) [40].

Opposed to the ExtraTree model, where the trees are created by random selections, the Gradient Boosting regression builds the trees iteratively. In each iteration, the model accuracy is estimated

by a loss function (typically by the mean squared error, MSE), and new trees are created to correct the error of the previous trees. The final prediction is then given as a weighted average of the prediction of the trees. The weights are updated using a gradient descent method in each iteration of the training. This error-correcting strategy enables higher accuracy, but also can make the model more prone to overfitting [40–42].

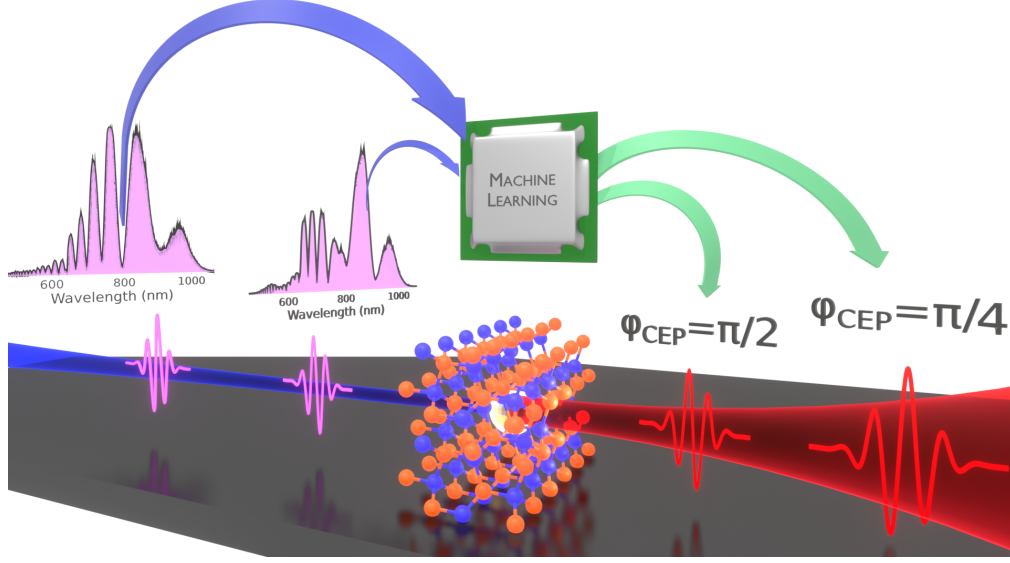


Fig. 1. Concept of the experiment. The driving MIR pulses are focused on the ZnO crystal and high order harmonics are generated and detected with a spectrometer. By changing the CEP of the input laser the corresponding harmonic spectrum changes too. A machine learning model is trained to link the laser CEP to the corresponding harmonic spectrum. Applying this model the CEP of the the fundamental laser can be estimated from the observed harmonic spectra.

3. Simulations

In order to simulate the CEP-dependent HHG process, we employ an one-dimensional semi-classical model, which provides a computationally efficient way to compute the response of electrons in a periodic potential to a strong laser field [43]. Herein we adopt the single-electron approximation to derive the Bloch states and their associated energies, enabling us to discern the distinct contributions of various initial states to high harmonic radiation. In more detail, for an electron with charge e and mass m , we consider the following Hamiltonian:

$$H(t) = \frac{1}{2m}(\mathbf{p} - e\mathbf{A}(t))^2 + U(x), \quad (1)$$

where velocity gauge is used, $\mathbf{A}(t)$ denotes the time-dependent vector potential component of the external field along the x direction, while $U(x)$ represents the periodic (model) potential of the solid. This simplified model of the potential adequately captures essential features of the crystal lattice's atomic structure, including the band gap width; however, it ignores the three-dimensional symmetries including the broken inversion symmetry of C6v. Consequently, in our investigation of ZnO targets, we appropriately consider factors such as the lattice constant of 5.2 Å along c -axis, and a band gap measuring 3.27 eV. On the other hand, scattering events and finite temperature can be taken into account by considering density matrix ρ instead of pure quantum mechanical

states (which are practically Bloch waves in this case). That is, we solve the von Neumann equation

$$\frac{\partial}{\partial t}\rho(t) = -\frac{i}{\hbar} [H(t), \rho(t)] + \frac{\partial}{\partial t}\rho(t)\Big|_{scatt}, \quad (2)$$

such that the initial density matrix (before the interaction with the laser field) describes thermal equilibrium, and the second term is responsible for the scattering events. We consider the change of the diagonal elements of the density matrix towards thermal equilibrium at a rate of γ_d , and also the decay of the off-diagonal matrix elements (i.e., the loss of quantum mechanical coherences) with a rate of γ_{od} . By considering realistic rates ($\gamma_{od} = 0.25 \text{ fs}^{-1}$ and $\gamma_d = 0.05 \text{ fs}^{-1}$), not only the CEP dependence of the harmonic peaks can be determined, but also the most important features of the HHG spectra, namely the presence of the plateau, and the intensity dependence of the cutoff and the heights of the harmonic peaks can be investigated [43]. A numerical solution is obtained using a Cash-Karp Runge-Kutta algorithm. The sinusoidal model periodic potential, $U(x)$ is parametrized by the lattice constant d and the potential depth U_0 . Corresponding to the eight valence electrons present in one unit cell of ZnO, the value of U_0 is chosen so that the band gap between the 4th and the 5th band corresponds to the experimentally determined band gap of ZnO, 3.27 eV, as reported by [44]. The band structure of this one-dimensional model is presented on Fig. S4 of the Supplementary document.

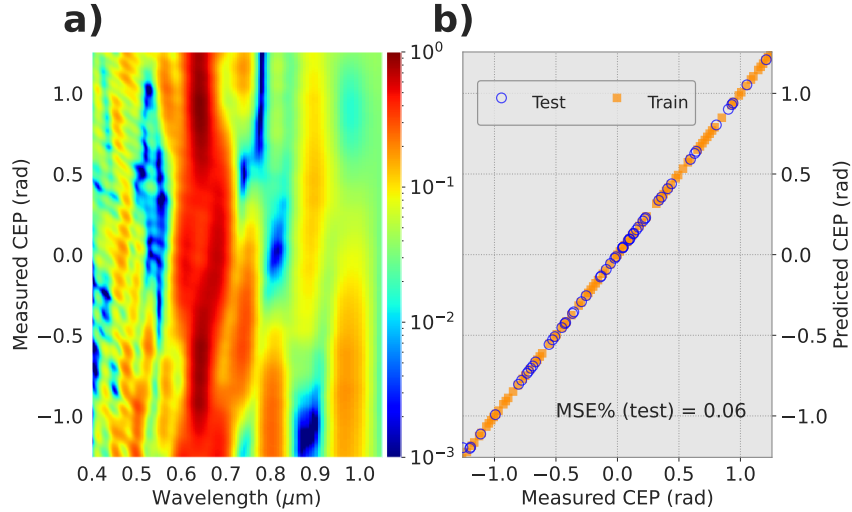


Fig. 2. Theoretical CEP scan performed in ZnO with a 18 fs, $I_0 = 1.36 \times 10^{12} \text{ W cm}^{-2}$, 3.2 μm laser (a). Performance of the machine learning algorithm on estimating the measured relative laser CEP using the train (orange rectangles) and the test (open blue circles) datasets is shown in (b), with respective MSE% values of 0 and 0.06. The slope of both estimations are close to 45°, representing almost perfect estimating capabilities. One fifth of the test data set is plotted to provide better visual transparency between the train and test data.

For simulating the HHG process, a laser with $t_p = 18 \text{ fs}$ pulse duration, $\lambda_0 = 3.2 \mu\text{m}$ and a peak intensity of $I_0 = 1.36 \times 10^{12} \text{ W cm}^{-2}$ (corresponding to a peak field of $E_0 = 3.20 \text{ GV m}^{-1}$) is selected. We employ a \cos^2 temporal pulse profile as an approximation to a Gaussian profile, leveraging its computational efficiency. Our objective is to model an experiment that can be readily executed within a conventional HHG beamline setup, without necessitating specialized equipment. To this end, the most accessible means for harmonic radiation detection is through

the utilization of a spectroscope equipped with a CCD camera. Consequently, we multiply each HHG spectrum with a response function composed of a typical response curve of a grating and a CCD camera, thereby constraining the range of observable harmonics. Details of this response function are reported in the SM.

Using this numerical model, a HHG spectrum is simulated with ~ 800 different CEP values between $\pm 0.5\pi$. The result is shown on Fig. 2a. After generating the simulated CEP-dependent harmonic spectra, they are partitioned into two datasets through random selection. Eighty percent of the data is allocated to the training dataset, while the remaining twenty percent is assigned to the test dataset. The training dataset is employed to train an ExtraTree model, enabling the recognition of patterns linking the spectrum to a specific CEP value. Subsequently, the trained model processes both the spectra from the training and test datasets (the latter representing previously unseen spectra), leveraging the learned patterns to estimate the corresponding laser CEP. As anticipated, the estimated CEP values align perfectly with the training set (particularly considering those data were used to train the model), as shown in Figure 2 b as orange rectangles.

The predictive accuracy of the model is determined by the relative square root of mean square error between the true (CEP_t) and predicted (CEP_p) carrier-envelope phase values expressed in percentages as

$$\text{MSE}\% = \sqrt{\frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} (\text{CEP}_{t,i} - \text{CEP}_{p,i})^2} \cdot \frac{100}{\pi}. \quad (3)$$

When applying the model on the test set (blue circles), very good agreement is achieved between the true and the estimated CEP values, with an MSE% of 0.06. The outcome proves our assumption, that there is a one-to-one link between the harmonic spectrum and the CEP and this can be captured with a machine learning model. The excellent agreement observed between the true and estimated CEP values can be attributed in part to the extensive sampling of training data. Nevertheless, acquiring a sufficient number of experimental data points may pose challenges in practical scenarios. Generally, increased training data enhances model accuracy. Therefore, it is important to discuss the effect of the number of training data (N_{train}) on the model accuracy (MSE%) using the test data. Furthermore, in a general case, the data used for model training is picked on randomly selected label values. This has the disadvantage of an uneven sampling of the parameter space, which in turn results in a varying model accuracy over the parameter range. However, in an experimental situation where the tag values are known to be within a closed range, it is possible to provide the model with an evenly sampled training data.

To address this question, we train three different models (ExtraTree, Linear Regression and GradientBoosting) using a varying number of train data (N_{train}) and a fixed number of test data (N_{test}). We have simulated a total of 800 solid HHG spectra corresponding to an input of an equidistant CEP grid between -0.4π and 0.4π (in the simulation, because of the 1D model, the CEP has π periodicity, therefore we reduced the examined CEP range, to keep one-to-one mapping). We used random sampling without replacement to select $N_{\text{train}} = 1 - 400$ pairs and for each N_{train} case we selected in a similar manner $N_{\text{test}} = 150$ and calculated the model's performance, shown in Figure 3 a. This random selection is performed 10 times for each N_{train} values. The model is trained using the selected N_{train} HHG spectra, and then tested using the N_{test} data points to estimate the CEP from spectra that was not known for the model before. The MSE% quality parameter is then obtained by performing a linear fit on the predicted versus actual CEP values of the test dataset.

Figure 3 presents a comparative result of this analysis of the effect of N_{train} using randomly (a) and equidistantly (b) picked CEP training data. Regarding the models, we can observe similar trends in both cases. The ExtraTree model presents superior model accuracy in case of low number of training data ($N_{\text{train}} < 100$), which can be attributed to the algorithm's marked resilience against overfitting. However, the linear model, which is the least accurate in this

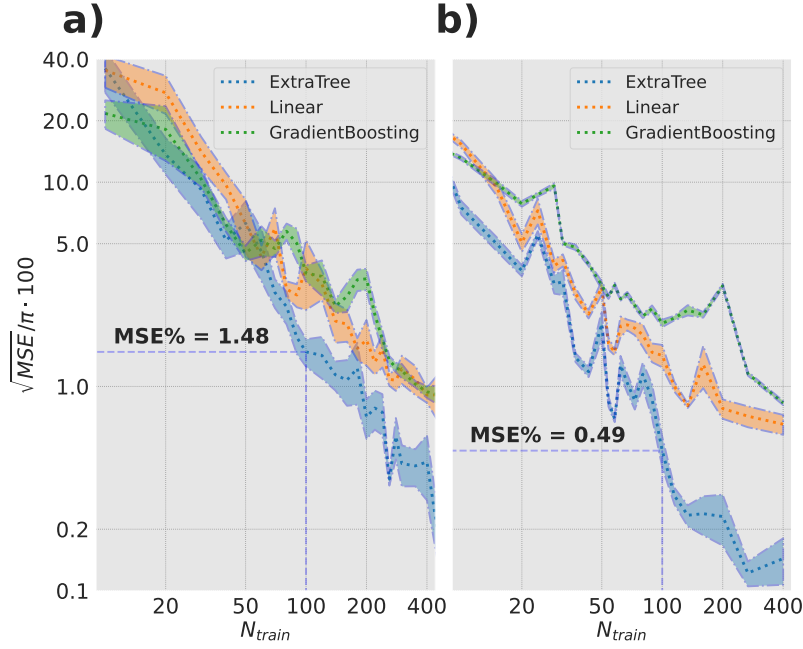


Fig. 3. Comparative analysis of model predictive accuracy between a) random and b) equidistant CEP sampling of the training data. The precision is given by the MSE% of the linear fit on the prediction from 150 test datasets at randomly picked novel CEP values. The random selection is done 10 times at each N_{train} value allowing to give error estimates to the predictions (shaded area on the plots). In both randomly and evenly sampled CEP scenarios, the model performance with the test data is given on the plots for $N_{\text{train}} = 100$ for the ExtraTree model.

sparingly sampled regime, achieves superior accuracy in case of a densely sampled training data ($N_{\text{train}} > 200$). This hints to the existence of a strong linear correlation between the CEP of the laser and the features of the HHG spectra. The GradientBoosting method stays in-between for small train sets, and performs comparatively poorly when a high number of training data is available. As expected, in both sampling cases, at low number of training data the performance improves exponentially (note the log-log plot) with the number of the training data, but around $N_{\text{train}} = 200$ the prediction accuracy starts to saturate, indicating that additional training points will not improve substantially the model's precision.

It is also observed that the equidistant sampling (Figure 3 b) of the training data results in a significantly improved accuracy in all cases, resulting in an increase of the accuracy by about an order of magnitude, especially in the case of low ($N_{\text{train}} < 200$). This implies that the number of required training data points for a given accuracy can be significantly reduced by performing a methodical CEP sampling in the experimental case. The improvement is especially prominent in case of the ExtraTree model since it gains an order of magnitude prediction performance in favor of the evenly sampled scenario.

4. Experimental setup

The experiments were carried out using the MIR laser at ELI-ALPS [45]. The laser is operating at 100 kHz repetition at 3.2 μm and capable to deliver 140 μJ of pulse energy. The measured spectrum of the driving field can be seen in Figure S1 in the Supplementary document. The 45 fs output pulses are spectrally broadened in BaF2 and Si optical windows and recompressed in bulk

BaF2 windows combined with three reflections on negative TOD dispersive mirrors to reach 18 fs pulse duration. The temporal profile of the pulse was measured by a TIPTOE device by coupling out the beam before the off-axis parabolic mirror. The accurate CEP measurement and control is crucial for this experiment, we used an f-2f setup called Fringeazz by Fastlite which is controlling the acousto-optic programmable dispersive filter (AOPDF, Dazzler, Fastlite) in the OPCPA front-end in a closed loop. The device is able to measure CEP values shot-to-shot at 10 kHz, while the laser is capable to deliver 100 mrad CEP stability. We generate high harmonics in a 90 μm thick ZnO crystal by focusing 1.3 μJ pulses using an off axis parabola with 100 mm focal length. The resulting intensity for the compressed pulses was $1.3 \cdot 10^{12} \text{ W/cm}^2$, by introducing chirp we lowered this intensity to $6.6 \cdot 10^{11} \text{ W/cm}^2$. After solid HHG, a thick 50 mm long BK7 bulk filters out the MIR driving beam and transmits the harmonics, but cuts off radiation below 350 nm. Subsequently, the remaining visible range of the harmonic spectrum is imaged into a commercial spectrometer (AvaSpec-ULS2048 from Avantes). In this configuration we were able to measure from harmonic order 3 to harmonic order 9. During the measurement we recorded 100 spectra for each CEP settings and used the average of these during the analysis, as shown in Figure 4 a. The CEP scan was recorded between $-\pi$ and π having 100 evenly distributed CEP values in between. The spectrum for each CEP value is normalized in this plot.

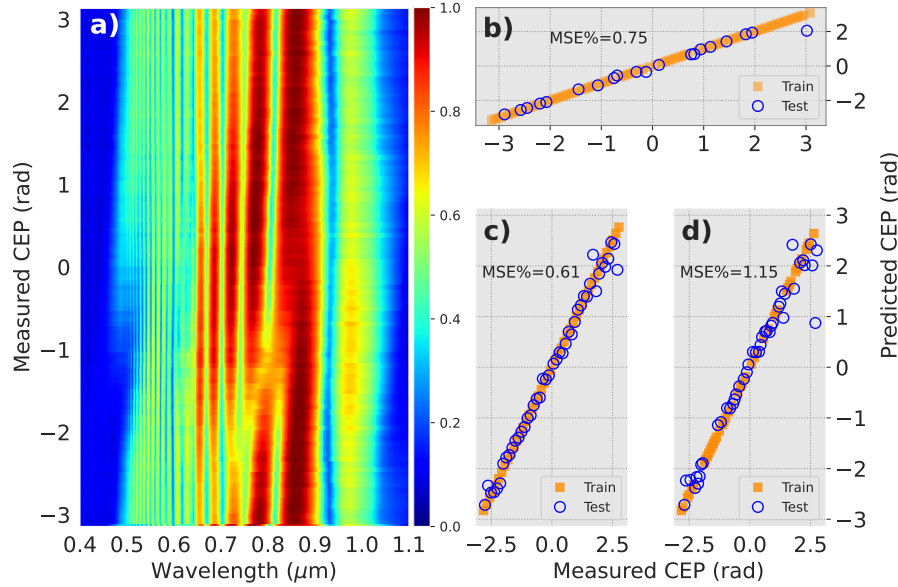


Fig. 4. Experimental CEP scan in ZnO performed with a 18 fs, 0.13 W, 3.2 μm laser (a). The performance of the ExtraTree machine learning algorithm on estimating the measured relative laser CEP using (b) 80% train (orange rectangles), 20% test (open blue circles); (c) evenly while (d) randomly sampled 50% train and 50% test datasets. The slope of each prediction scenario lies very close to 45° , representing almost perfect estimating capabilities, with MSE% values on the test sets: (b) 0.75, (c) 0.61, (d) 1.15.

The fringe pattern extending over the NIR region on the harmonic spectrum on Figure 4 is solely due to interference of harmonics generated from the front and the rear side of the ZnO crystal [23] and is therefore independent of the CEP. However, the harmonic spectrum expresses harmonic minima at different CEP values, which is the result of constructive and destructive interference between the XUV bursts generated in consecutive laser half cycles [32, 34].

Direct comparison between simulation and experimental data is not straightforward, because

on one hand, in simulation the CEP is expressed in absolute terms, while in experiment it is relative. On the other hand, the simulation does not take into account the complexity of the experimental conditions, and the 1D model cannot be expected to reproduce neither 3D symmetry properties of the real crystal, nor propagation-related effects, like phase matching. However, the spectral minima shift as the function of CEP is visible in both simulations and experiments. It is known that the extent of CEP sensitivity has dependence on the instantaneous structural changes in the laser-induced lattice, which is not taken into our simulations, thereby possibly causing some difference in features. Furthermore, it is important to note that the CEP scan confirms a one-to-one mapping between the CEP and the related spectra, as required for the machine learning model to work.

We split the experimentally recorded data in Figure 4 b into train (80%) and test (20%) randomly selecting into these datasets and present the performance of a trained ExtraTree model in estimating the laser's CEP from the input spectra. The orange rectangles represent the CEP predictions using the spectra from the train data set, showing perfect agreement between the predicted and the true laser CEP. The blue circles present the performance of the model on the previously unseen test data, indicating that the model is well trained to recognize spectral patterns in the solid HHG spectrum to estimate the laser CEP correctly with an $\text{MSE}\% = 0.75$. The only major discrepancy between the estimated and the true CEP is where there is a gap in the sampling in the train data set (around π rad in the figure). This issue can be overcome by evenly sampling the training dataset, as described in Figure 3 b. To visualize this effect we resample our dataset into 50% train and 50% test with evenly and randomly sampled input CEP values, as shown in Figure 4 c and d, respectively. As expected, the evenly sampled scenario ($\text{MSE}\% = 0.61$) is outperforming the randomly sampled case ($\text{MSE}\% = 1.15$) by a factor of 2 in estimating the CEP of the laser from the harmonic spectra. These outcomes clearly prove that our concept for CEP estimation works and could be used for monitoring laser CEP during experiments. They also highlight that good model performance can be only achieved if the training CEP values are evenly sampled with small step size.

5. Conclusion and outlook

In conclusion, we have laid out a conceptual scheme for estimating MIR laser CEPs relying on the spectrum of high order harmonics generated from a solid crystal exploiting the ability of complex pattern recognition of a machine learning model. Furthermore, we demonstrated the applicability of this concept both through theoretical simulations and experimental measurements. This proven scheme offers an economic, instrumentally not demanding option to measure the laser CEP. The concept can be generalized to other laser wavelength, assuming that the combination of the laser and the crystal fulfills the requirements of one-to-one mapping of the CEP to the solid HHG spectrum and the existence of 2π periodicity. Furthermore, in principle it is possible to perform single shot recordings of the harmonic spectrum with a sampled beam (only 1% of the total energy was used) while an experiment is carried out with the remaining beam. In case of random CEP laser source, the latter approach will allow CEP tagging.

The current study relies on a reference method as implemented in a fast CEP measurement device, Fringeazz by FastLight, to train the ML model. However with more accurate modelling the simulated and experimental CEP scans could match better, meaning simulations could be a tool to train the ML model, as it was done in [27] for FROG scans. Additional benefit of such simulation would be to retrieve the absolute CEP of the laser, since in simulations it is an input parameter. It is important to notice that this study implies a fixed laser intensity for the training and for subsequent actual measurements and CEP retrieval. The question arises then whether intensity variations would ruin the process. Initial investigations have been carried out that show the ability of the machine learning to recover CEP values from various field configurations and intensities. However, in-depth study of the sensitivity to intensity fluctuations is ongoing.

Overall, our proposed protocol, combined with precise simulations, opens up the route towards multi-parameter estimation and optimization using machine learning concepts, as it has been already proposed by [26, 28, 29]. Solid HHG is a process, where the laser parameters and the material features have a direct manifestation in the measured spectrum, resulting in an economic and adaptable tool for monitoring experimental conditions. Therefore we utilize HHG signal from semiconducting ZnO crystal in MIR regime to infer the CEP, based on ML models trained with combined experimental and simulations results. Our findings demonstrate the effectiveness of the ExtraTree-based models for predicting the one-to-one correlation between CEP and HHG spectra. Our approach can prove instrumental in determination and control of laser parameters, such as intensity, CEP in-situ depending on the target material, prior to an advanced HHG based experiment design. While it is possible to infer CEP from spectral measurement, this usually needs extended data acquisition. Our method achieves the same with significantly less number of measurements, thereby being of prominent use towards reducing cost of running experiment campaign, as relevant for large-scale user facilities like ELI ALPS.

6. Funding

The ELI-ALPS project (GINOP-2.3.6-15-2015-00001) is supported by the European Union and co-financed by the European Regional Development Fund.

7. Acknowledgments

We thank Rajaram Shrestha for his kind help in processing data and plotting figures for the Supplementary Material.

References

1. A. Nayak, M. Dumergue, S. Kühn, S. Mondal, T. Csizmadia, N. Harshitha, M. Füle, M. Upadhyay Kahaly, B. Farkas, B. Major, V. Szaszkó-Bogár, P. Földi, S. Majorosi, N. Tsatrafyllis, E. Skantzakis, L. Neoričić, M. Shirozhan, G. Vampa, K. Varjú, P. Tzallas, G. Sansone, D. Charalambidis, and S. Kahaly, “Saddle point approaches in strong field physics and generation of attosecond pulses,” *Phys. Reports* **833**, 1-52 (2019).
2. P. Tenio, C. Ming-Chang, A. Paul, M. M. M., and K. H. C., “The attosecond nonlinear optics of bright coherent x-ray generation,” *Nat. Photonics* **4**, 822–832 (2010).
3. C. Marceau, T. J. Hammond, A. Y. Naumov, P. B. Corkum, and D. M. Villeneuve, “Wavelength scaling of high harmonic generation for 267 nm, 400 nm and 800 nm driving laser pulses,” *J. Phys. Commun.* **1**, 015009 (2017).
4. A. S. Johnson, D. R. Austin, D. A. Wood, C. Brahms, A. Gregory, K. B. Holzner, S. Jarosch, E. W. Larsen, S. Parker, C. S. Strüber, P. Ye, J. W. G. Tisch, and J. P. Marangos, “High-flux soft x-ray harmonic generation from ionization-shaped few-cycle laser pulses,” *Sci. Adv.* **4** (2018).
5. Y. Fu, K. Nishimura, R. Shao, A. Suda, K. Midorikawa, P. Lan, and E. J. Takahashi, “High efficiency ultrafast water-window harmonic generation for single-shot soft x-ray spectroscopy,” *Commun. Phys.* **3**, 92 (2020).
6. S. Ghimire, A. D. DiChiara, E. Sistrunk, P. Agostini, L. F. DiMauro, and D. A. Reis, “Observation of high-order harmonic generation in a bulk crystal,” *Nat. Phys.* **7**, 138–141 (2010).
7. S. K. Jongkyoon Park, Amutha Subramani and M. F. Ciappina, “Recent trends in high-order harmonic generation in solids,” *Adv. Physics: X* **7**, 2003244 (2022).
8. D. M. Simanovskii, H. A. Schwettman, H. Lee, and A. J. Welch, “Midinfrared optical breakdown in transparent dielectrics,” *Phys. Rev. Lett.* **91** (2003).
9. E. Migal, E. Mareev, E. Smetanina, G. Duchateau, and F. Potemkin, “Role of wavelength in photocarrier absorption and plasma formation threshold under excitation of dielectrics by high-intensity laser field tunable from visible to mid-ir,” *Sci. Reports* **10** (2020).
10. M. Kübel, R. Siemering, C. Burger, N. G. Kling, H. Li, A. Alnaser, B. Bergues, S. Zherebtsov, A. Azzeer, I. Ben-Itzhak, R. Moshhammer, R. de Vivie-Riedle, and M. Kling, “Steering proton migration in hydrocarbons using intense few-cycle laser fields,” *Phys. Rev. Lett.* **116** (2016).
11. V. Hanus, S. Kangaparambil, S. Larimian, M. Dorner-Kirchner, X. Xie, M. S. Schöffler, G. G. Paulus, A. Baltuška, A. Staudte, and M. Kitzler-Zeiler, “Experimental separation of subcycle ionization bursts in strong-field double ionization of H_2 ,” *Phys. Rev. Lett.* **124** (2020).
12. M. Krüger, M. Schenk, and P. Hommelhoff, “Attosecond control of electrons emitted from a nanoscale metal tip,” *Nature* **475**, 78-81 (2011).
13. C. A. Haworth, L. E. Chipperfield, J. S. Robinson, P. L. Knight, J. P. Marangos, and J. W. G. Tisch, “Half-cycle cutoffs in harmonic spectra and robust carrier-envelope phase retrieval,” *Nat. Phys.* **3**, 52-57 (2006).

14. P. A. Roos, X. Li, R. P. Smith, J. A. Pipis, T. M. Fortier, and S. T. Cundiff, "Solid-state carrier-envelope phase stabilization via quantum interference control of injected photocurrents," *Opt. Lett.* **30**, 735 (2005).
15. T. M. Fortier, P. A. Roos, D. J. Jones, S. T. Cundiff, R. D. R. Bhat, and J. E. Sipe, "Carrier-envelope phase-controlled quantum interference of injected photocurrents in semiconductors," *Phys. Rev. Lett.* **92** (2004).
16. L. Xu, T. W. Hänsch, C. Spielmann, A. Poppe, T. Brabec, and F. Krausz, "Route to phase control of ultrashort light pulses," *Opt. Lett.* **21**, 2008 (1996).
17. C. Feng, J.-F. Hergott, P.-M. Paul, X. Chen, O. Tcherbakoff, M. Comte, O. Gobert, M. Reduzzi, F. Calegari, C. Manzoni, M. Nisoli, and G. Sansone, "Complete analog control of the carrier-envelope-phase of a high-power laser amplifier," *Opt. Express* **21**, 25248 (2013).
18. M. Kurucz, S. Tóth, R. Flender, L. Haizer, B. Kiss, B. Persielle, and E. Cormier, "Single-shot CEP drift measurement at arbitrary repetition rate based on dispersive fourier transform," *Opt. Express* **27**, 13387 (2019).
19. C. Guo, M. Miranda, A.-K. Raab, A.-L. Viotti, P. T. Guerreiro, R. Romero, H. Crespo, A. L'Huillier, and C. L. Arnold, "Single-shot, high-repetition rate carrier-envelope-phase detection of ultrashort laser pulses," *Opt. Lett.* **48**, 5431–5434 (2023).
20. T. Wittmann, B. Horvath, W. Helml, M. G. Schätzel, X. Gu, A. L. Cavalieri, G. G. Paulus, and R. Kienberger, "Single-shot carrier-envelope phase measurement of few-cycle laser pulses," *Nat. Phys.* **5**, 357–362 (2009).
21. D. Hoff, F. J. Furch, T. Witting, K. Rühle, D. Adolph, A. M. Saylor, M. J. J. Vrakking, G. G. Paulus, and C. P. Schulz, "Continuous every-single-shot carrier-envelope phase measurement and control at 100 kHz," *Opt. Lett.* **43**, 3850 (2018).
22. M. Shirozhan, S. Mondal, T. Grósz, B. Nagyillés, B. Farkas, A. Nayak, N. Ahmed, I. Dey, S. Choudhary De Marco, K. Nelissen, M. Kiss, L. Gulyás Oldal, T. Csizmadia, Z. Filus, M. De Marco, S. Madas, M. Upadhyay Kahaly, D. Charalambidis, P. Tzallas, E. Appi, R. Weissenbilder, P. Eng-Johnsson, A. L'Huillier, Z. Diveki, B. Major, K. Varjú, and S. Kahaly, "High-repetition-rate attosecond xuv beamlines at eli alps for studying ultrafast phenomena," *Ultrafast Sci.* (2024).
23. R. Hollinger, D. Hoff, P. Wustelt, S. Skruszewicz, Y. Zhang, H. Kang, D. Würzler, T. Jungnickel, M. Dumergue, A. Nayak, R. Flender, L. Haizer, M. Kurucz, B. Kiss, S. Kühn, E. Cormier, C. Spielmann, G. G. Paulus, P. Tzallas, and M. Kübel, "Carrier-envelope-phase measurement of few-cycle mid-infrared laser pulses using high harmonic generation in zno," *Opt. Express* **28**, 7314–7322 (2020).
24. A. Leblanc, P. Lassonde, G. Dalla-Barba, E. Cormier, H. Ibrahim, and F. Légaré, "Characterizing the carrier-envelope phase stability of mid-infrared laser pulses by high harmonic generation in solids," *Opt. Express* **28**, 17161–17170 (2020).
25. M. Awad, A. Manna, S. Hell, B. Ying, L. Ábrók, Z. Divéki, E. Cormier, B. Kiss, J. Böhmer, C. Ronning, S. H. Han, A. George, A. Turchanin, A. N. Pfeiffer, and M. Kübel, "Few-cycle laser pulse characterization on-target using high-harmonic generation from nano-scale solids," *Opt. Express* **32**, 1325 (2024).
26. A. Sanchez-Gonzalez, P. Micaelli, C. Olivier, T. R. Barillot, M. Ilchen, A. A. Lutman, A. Marinelli, T. Maxwell, A. Achner, M. Agåker, N. Berrah, C. Bostedt, J. D. Bozek, J. Buck, P. H. Bucksbaum, S. C. Montero, B. Cooper, J. P. Cryan, M. Dong, R. Feifel, L. J. Frasinski, H. Fukuzawa, A. Galler, G. Hartmann, N. Hartmann, W. Helml, A. S. Johnson, A. Knie, A. O. Lindahl, J. Liu, K. Motomura, M. Mücke, C. O'Grady, J.-E. Rubensson, E. R. Simpson, R. J. Squibb, C. Sâthe, K. Ueda, M. Vacher, D. J. Walke, V. Zhaunerchyk, R. N. Coffee, and J. P. Marangos, "Accurate prediction of x-ray pulse properties from a free-electron laser using machine learning," *Nat. Commun.* **8**, 15461 (2017).
27. I. Tóth, A. M. M. Gherman, K. Kovács, W. Cho, H. Yun, and V. Toşa, "Reconstruction of femtosecond laser pulses from frog traces by convolutional neural networks," *Photonics* **10** (2023).
28. N. D. Klimkin, Álvaro Jiménez-Galán, R. E. F. Silva, and M. Ivanov, "Symmetry-aware deep neural networks for high harmonic spectroscopy in solids," *Opt. Express* **31**, 20559–20571 (2023).
29. S. Yang, X. Liu, J. Lin, R. Zuo, X. Song, M. Ciappina, and W. Yang, "Reconstructing the semiconductor band structure by deep learning," *Mathematics* **10**, 4268 (2022).
30. N. Tsatrafyllis, S. Kühn, M. Dumergue, P. Foldi, S. Kahaly, E. Cormier, I. Gonoskov, B. Kiss, K. Varju, S. Varro *et al.*, "Quantum optical signatures in a strong laser pulse after interaction with semiconductors," *Phys. Rev. Lett.* **122**, 193602 (2019).
31. S. Ghimire and D. A. Reis, "High-harmonic generation from solids," *Nat. Phys.* **15**, 10–16 (2018).
32. Y. S. You, M. Wu, Y. Yin, A. Chew, X. Ren, S. Gholam-Mirzaei, D. A. Browne, M. Chini, Z. Chang, K. J. Schafer, M. B. Gaarde, and S. Ghimire, "Laser waveform control of extreme ultraviolet high harmonics from solids," *Opt. Lett.* **42**, 1816 (2017).
33. O. Schubert, M. Hohenleutner, F. Langer, B. Urbanek, C. Lange, U. Huttner, D. Golde, T. Meier, M. Kira, S. W. Koch, and R. Huber, "Sub-cycle control of terahertz high-harmonic generation by dynamical bloch oscillations," *Nat. Photonics* **8**, 119–123 (2014).
34. Y. S. You, Y. Yin, Y. Wu, A. Chew, X. Ren, F. Zhuang, S. Gholam-Mirzaei, M. Chini, Z. Chang, and S. Ghimire, "High-harmonic generation in amorphous solids," *Nat. Commun.* **8** (2017).
35. M. Garg, H. Y. Kim, and E. Goulielmakis, "Ultimate waveform reproducibility of extreme-ultraviolet pulses by high-harmonic generation in quartz," *Nat. Photonics* **12**, 291–296 (2018).
36. Z. Wang, H. Park, Y. H. Lai, J. Xu, C. I. Blaga, F. Yang, P. Agostini, and L. F. DiMauro, "The roles of photo-carrier doping and driving wavelength in high harmonic generation from a semiconductor," *Nat. Commun.* **8** (2017).

37. N. Altman and M. Krzywinski, "Simple linear regression," *Nat. Methods* **12**, 999-1000 (2015).
38. M. Krzywinski and N. Altman, "Multiple linear regression," *Nat. Methods* **12**, 1103-1104 (2015).
39. P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.* **63**, 3-42 (2006).
40. M. Ghazwani and M. Y. Begum, "Computational intelligence modeling of hyoscine drug solubility and solvent density in supercritical processing: gradient boosting, extra trees, and random forest models," *Sci. Reports* **13** (2023).
41. J. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *Ann. Stat.* **29**, 1189-1232 (2001).
42. A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurobotics* **7** (2013).
43. V. Szaszko-Bogár, P. Földi, I. Magashegyi, and K. Varjú, "Interference-induced phenomena in high-order harmonic generation from bulk solids," *Appl. Sci.* **9**, 1572 (2019).
44. B. Hussain, A. Aslam, T. Khan, M. Creighton, and B. Zohuri, "Electron affinity and bandgap optimization of zinc oxide for improved performance of zno/si heterojunction solar cell using pc1d simulations," *Electronics* **8**, 238 (2019).
45. N. Thiré, R. Maksimenka, B. Kiss, C. Ferchaud, G. Gitzinger, T. Pinoteau, H. Jousselin, S. Jarosch, P. Bizouard, V. Di Pietro, E. Cormier, K. Osvay, and N. Forget, "Highly stable, 15 w, few-cycle, 65 mrad cep-noise mid-ir opcpa for statistical physics," *Opt. Express* **26**, 26907 (2018).