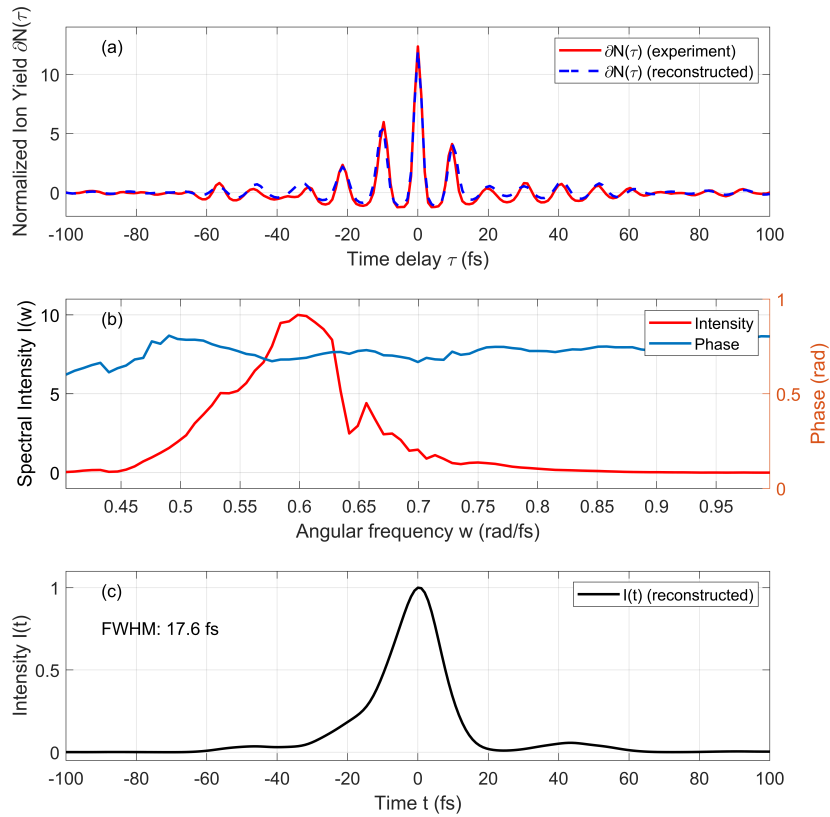


# MIR laser CEP estimation using machine learning concepts in bulk high harmonic generation: supplemental document

## 1. MIR LASER CHARACTERISATION



**Fig. S1.** Characteristics of the MIR laser: (a) Experimental (red solid line) and reconstructed (dashed blue line) ionization yield versus delay provided by TIPTOE measurement; (b) measured intensity spectrum (red) used for reconstruction and the reconstructed spectral phase (blue); (c) reconstructed temporal intensity profile of the MIR pulse.

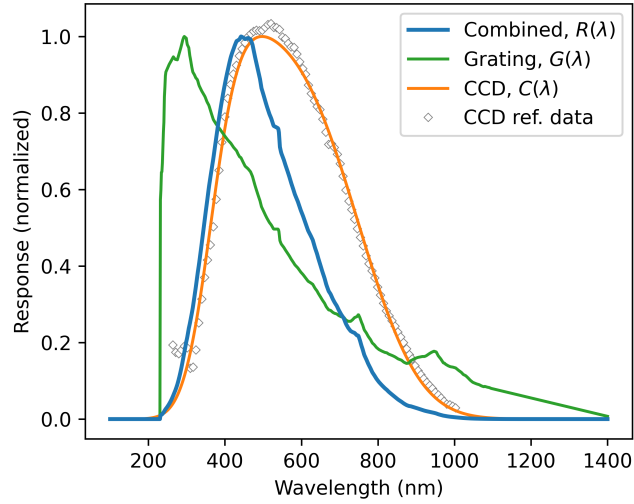
Figure S1 (a) shows the results of TIPTOE measurement on the post-compressed pulse by BaF2 and Si windows [1]. The beam was coupled out for the measurement right before the focusing mirror. The spectrum in Figure S1 (b) is centered at  $3.16 \mu\text{m}$  and spans from  $2.1 \mu\text{m}$  to  $4.5 \mu\text{m}$  (at 1% intensity level). The measured spectrum combined with the reconstructed spectral phase is then used to calculate the temporal profile of the laser beam in Figure S1 (c). Its FWHM pulse duration is estimated to be 17.6 fs.

## 2. RESPONSE CURVES

The spectral response of the spectrometer is modeled by a response function  $R(\lambda) = C(\lambda)G(\lambda)$ , which is composed of the quantum efficiency curve of the CCD ( $C(\lambda)$ ) and the grating response of the spectrometer ( $G(\lambda)$ ). The CCD response curve is modeled as

$$C(\lambda) = \frac{1}{4} \left( 1 + \operatorname{erf} \left( 2 \frac{\lambda - \lambda_l}{\Delta\lambda_l} \right) \right) \left( 1 + \operatorname{erf} \left( 2 \frac{\lambda - \lambda_h}{\Delta\lambda_h} \right) \right), \quad (\text{S1})$$

where  $\operatorname{erf}(x)$  is the error function, and the parameters  $\lambda_l = 747.3 \text{ nm}$ ,  $\Delta\lambda_l = 356.6 \text{ nm}$ ,  $\lambda_h = 364.0 \text{ nm}$ ,  $\Delta\lambda_h = 160.4 \text{ nm}$  were numerically fitted to correspond to the spectral response of a typical commercially available CCD camera (see Fig. S2, gray open circles). The grating response function  $G(\lambda)$ , represented by the orange line, is obtained by linear interpolation on the reference data provided by the manufacturer of the spectrometer. The model response curve  $R(\lambda)$  and its two components are shown on Fig. S2.

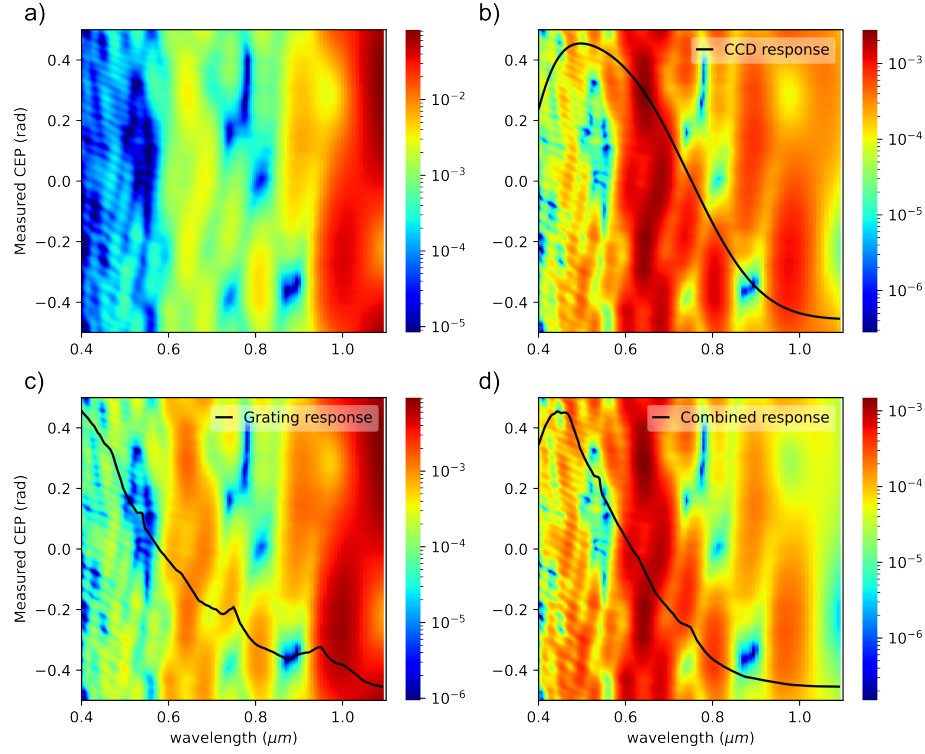


**Fig. S2.** The combined model response curve and its two components used on the simulated CEP-dependent HHG spectra. The reference data used for fitting the parameters of the model CCD response function (Eq. S1) is also shown.

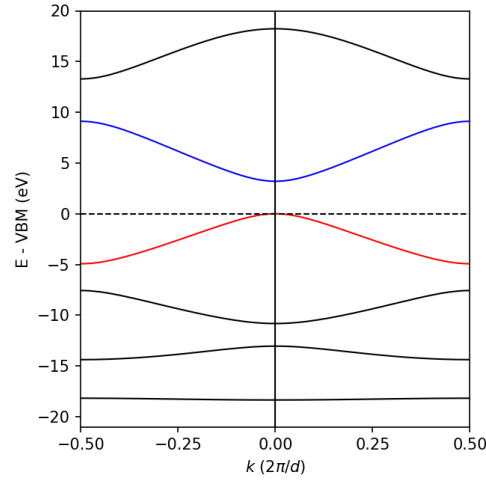
To highlight the effect of the response curves, we plot the simulated CEP dependence (Fig S3 a) and how it is altered by the filters (Fig S3 c-d). In the latter figures, the corresponding response functions are put on the top of the CEP scans. The response curve causes a relative amplification of the low-wavelength signal, around  $0.4 \mu\text{m}$  to  $0.6 \mu\text{m}$ , while the low-wavelength part of the spectrum (above  $0.8 \mu\text{m}$ ) is suppressed. An effect of specific interest is that the interference-like fringes below  $0.6 \mu\text{m}$ , which have low amplitude in the unmodified result, become prominent. On the other hand, the features slowly varying with CEP in the low-wavelength part, which have dominant intensity originally, are significantly reduced. This two effect probably contribute to an improved accuracy of the ML models, as the model may become more sensitive to small variations in the CEP. Together with the exponential-like decrease of the harmonic intensity with increasing photon energy, the response function has an overall effect of smoothing out the spectral intensity in the observed range.

## 3. MODEL BAND STRUCTURE

The band structure of the one-dimensional model lattice, showing the dispersion relation between the electron wavenumber  $k$  and the eigenenergies of the Bloch state within the first Brillouin zone, is shown on Fig. S4. As discussed in Section 3 (Simulations) in the main text, the model parameter  $U_0$  is chosen so that the band gap corresponds to the experimental value of  $3.27 \text{ eV}$ . The band structure reproduces the direct band gap of ZnO at the  $\Gamma$  point.



**Fig. S3.** The effect of the response curves on the simulated CEP-dependent HHG spectra. a) unmodified simulation result; b)-d) Result after multiplication with b) the CCD response  $C(\lambda)$ ; c) the grating response  $G(\lambda)$ ; d) the combined response  $R(\lambda)$ . The response curve, which is normalized to 0-1, is shown as an overlay.



**Fig. S4.** The band structure of the one-dimensional model. The valence and conduction band are marked with red and blue respectively. The electron wavenumber  $k$  is displayed in units of the brilluoin zone size. The energies are offset so that the valence band maximum (VBM) is set to 0.

## REFERENCES

1. R. Flender, M. Kurucz, T. Grosz, *et al.*, "Dispersive mirror characterization and application for mid-infrared post-compression," J. Opt. **23**, 065501 (2021).