nature research

https://doi.org/10.1038/s41567-020-01042-w

Supplementary information

Resonant phase-matching between a light wave and a free-electron wavefunction

In the format provided by the authors and unedited

Resonant phase-matching between a light wave and a free-electron wavefunction

Raphael Dahan[‡], Saar Nehemia[‡], Michael Shentcis, Ori Reinhardt, Yuval Adiv, Xihang Shi, Orr Be'er, Morgan H. Lynch, Yaniv Kurman, Kangpeng Wang and Ido Kaminer*

[‡] *equal contributors*

* kaminer@technion.ac.il

Supplementary Information

Contents

Note 1: Extended PINEM theory 3

1a. The coherent interaction for the extended PINEM theory 3

1b. Incoherent contributions and the final electron probability density function 7

1c. Reduction to the conventional PINEM theory 8

1d. Comparison of the conventional and extended PINEM theories 9

Note 2: Electric field and coupling constant for prism geometry 11

2a. Derivation of the electric field 11

2b. Derivation of the coupling constant 14

Note 3: Comparison of quantum and classical interaction 17

3a. Classical interaction 17

3b. Quantum interaction 18

3c. Similarities and differences between the two descriptions 20

Note 4: Intuition for the phase-matching effect 22

Note 5: Additional information on the data analysis 26

5a. Correlations to combine multiple energy spectra **26**

5b. Fixing fluctuations in the time scan data **26**

5c. The initial electrons' classical probability density **27**

5d. Side lobes in the time scan map **31**

5e. Fitting theory to the experimental results **31**

Note 6: Finite-difference time-domain (FDTD) simulation **33**

References **34**

Note 1: Extended PINEM theory

This section presents an extension to the conventional photon-induced nearfield electron microscopy (PINEM) theory[1-9]. We reconstruct the first steps of the PINEM theory and then generalize it to precisely describe the interaction between electron and light *pulses*. By considering the **spatiotemporal dynamics** of the pulses, we provide a more general description to common PINEM experiments. In the end of the following derivation, we provide the **master equation** that describes typical acquired data in PINEM experiment: e.g., the electron spectral density and time scan measurement. This theoretical advancement provides us better tools to fit our experimental data and also explain other intriguing ideas where some we discuss below. All is built on our generalization of the coupling constant to be dynamical variable, introducing a time-dependent coupling constant that under the more aggressive assumptions reduces to its time-independent variant of the conventional theory. At the end of this section, we discuss the approximations required to retrieve the conventional case from our extended theory and provide a comparison between the two theories.

1a. The coherent interaction for the extended PINEM theory

The conventional PINEM theory[1-9] describes a free-electron wavefunction interacting with a classical EM field and provides an analytical derivation for the resulting electron wavefunction under several approximations. One of these approximations is neglecting the combined spatiotemporal dependence of both the electron and the laser, and considering only their temporal dependence. This is a reasonable approximation for short-distance interactions, but in our long interaction setup, we cannot longer use that approximation. Thereby, we had to extend the conventional PINEM theory to include the spatiotemporal dynamics of the electron and the electric field in order to explain our observations. We begin by introducing a general derivation for the interaction in the framework of PINEM up to the point this approximation is made. Then, we split the discussion and consider each case separately.

The general time-dependent Schrödinger equation electron–light interaction (relativistic corrections are added below) is

$$\left[\frac{\left(\widehat{\boldsymbol{P}}+q_{\mathrm{e}}\boldsymbol{A}\right)^2}{2m_{\mathrm{e}}}-q_{\mathrm{e}}V\right]\psi=i\hbar\frac{\partial\psi}{\partial t}, \tag{S1}$$

where $\widehat{\boldsymbol{P}}=-i\hbar\boldsymbol{\nabla}$ is the momentum operator, q_{e} and m_{e} are the electron charge and mass, respectively, $\boldsymbol{A}$ is the EM field vector potential, V is the EM field scalar potential, and $\psi=\psi(\boldsymbol{r},t)$ is the electron wavefunction.

Two main assumptions are used in the analytical treatment of the PINEM interaction: (1) electron paraxiality (i.e., the electron trajectory is constrained to a *linear axis* which w.l.o.g we denote by z) and (2) small recoil of the electron due to photon absorption/emission. Under these assumptions, the Schrödinger equation for an electron travelling along the $\hat{z}$ axis with a primary momentum of $\hbar k_0\hat{z}$ reduces to

$$\left[U_0-\hbar v_{\mathrm{e}}\left(i\frac{\partial}{\partial z}+k_0\right)-\frac{iq_{\mathrm{e}}v_{\mathrm{e}}}{\omega}\left(\tilde{E}_z e^{-i\omega t}-\tilde{E}_z^{*}e^{i\omega t}\right)\right]\psi=i\hbar\frac{\partial\psi}{\partial t}, \tag{S2}$$

where $\boldsymbol{v}_{\mathrm{e}}=v_{\mathrm{e}}\hat{z}$ is the electron velocity, $\boldsymbol{k_0}=k_0\hat{z}$ is the electron initial wave-vector, and U_0 is the electron initial energy. The z component of the electric field is the dominant one and can be written as $E_z(\boldsymbol{r},t)=\tilde{E}_z(\boldsymbol{r},t)e^{-i\omega t}$, with ω being the photon laser frequency and $\tilde{E}_z$ the complex electric field phasor. The physical field is equal to $2Re\{E_z(\boldsymbol{r},t)\}$. The fastest component of the field is dominated by $e^{-i\omega t}$, and any additional time dependency in $\tilde{E}_z$ is much slower than ω. In the transition from Eq. S1 to Eq. S2, relativistic corrections are accounted by considering the relativistic Lorentz factor for the velocity v_{e} in the expression for the electron mass γm_{e}. Note that the same derivation can be made through the relativistic Klein–Gordon or Dirac wave equation, yielding the same result[10]. It is convenient to separate the wavefunction to $\psi(r,t)=e^{i\left(k_0z-\frac{U_0}{\hbar}t\right)}\phi(r,t)$, where the first term is the plain-wave-

electron with its initial energy-momentum components and $\phi(\boldsymbol{r},t)$ describes the slower dynamics arising from the interaction. Now, Eq. S2 reduces to

$$v_{\mathrm{e}}\frac{\partial\phi}{\partial z}+\frac{\partial\phi}{\partial t}=-\frac{q_{\mathrm{e}}v_{\mathrm{e}}}{\hbar\omega}\left(e^{-i\omega t}\tilde{E}_z-e^{i\omega t}\tilde{E}_z^{*}\right)\phi, \tag{S3}$$

and the solution is found to be

$$\phi(\boldsymbol{r},t)=\phi_0(x,y,T)e^{-2i\,\mathrm{Im}\{g(\boldsymbol{r},t)e^{-i\omega T}\}}, \tag{S4}$$

Where we further define $T\equiv t-z/v_{\mathrm{e}}$ as the time variable associated with the electron, i.e. the time in its frame of reference. $\phi_0(x,y,T)$ is a general function that represents the *coherent* electron wavefunction before the interaction and $g(\boldsymbol{r},t)$ is the generalized-time-dependent coupling constant, defined by:

$$g(\boldsymbol{r},t)=\frac{q_{\mathrm{e}}}{\hbar\omega}\int_{-\infty}^{z}\tilde{E}_z\left(x,y,z';T+\frac{z'}{v_{\mathrm{e}}}\right)e^{-i\omega\frac{z'}{v_{\mathrm{e}}}}dz'= \tag{S5}$$

$$=e^{i\omega T}\frac{q_{\mathrm{e}}}{\hbar\omega}\int_{-\infty}^{z}E_z\left(x,y,z';T+\frac{z'}{v_{\mathrm{e}}}\right)dz'.$$

The last transition exploits the relation $E_z(\boldsymbol{r},t)=\tilde{E}_z(\boldsymbol{r},t)\mathrm{e}^{-i\omega t}$. Like the conventional PINEM coupling constant, the generalized-time-dependent one is a dimensionless complex parameter that describes the overall coupling efficiency between the electron and light. One can think of this dimensionless parameter as a quantitate value to compare between different electron–laser interactions strengths.

Note that the integration is computed only along the electron trajectory (z axis) **up to** the point of interest. Note also that the transverse dependence of g (in x and y directions) is set *solely* from the electric field E_z **at the** point of interest. In practice, the electrons are measured far away from the interaction region; thus, we can safely look at $z\to\infty$ as the

upper limit of the integral in Eq. S5. Notice that now, none of the above quantities depend *explicitly* on z.

The electron wavefunction (Eq. S4) and the coupling constant (Eq. S5) now take the form

$$\phi(x,y,T) = \phi_0(x,y,T)e^{-2i\,\mathrm{Im}\{g(x,y,T)e^{-i\omega T}\}}, \tag{S6}$$

with

$$g(x,y,T) = \frac{q_e}{\hbar\omega}\int_{-\infty}^{\infty}\tilde{E}_z\left(x,y,z';T+\frac{z'}{v_e}\right)e^{-i\omega\frac{z'}{v_e}}dz'. \tag{S7}$$

The next step is to calculate the coherent electron energy probability density. Since in the general sense, the coupling constant (Eq. S7) can depend on T at a potentially complicated way, the calculation of the electron probability density is not immediate and we need to find the Fourier transform of the electron wavefunction (Eq. S8) with respect to T,

$$\phi(x,y,T) = \int_{-\infty}^{\infty} f(x,y,U)\,e^{\frac{iUT}{\hbar}}dU, \tag{S8}$$

where U is the energy variable and $f(x,y,U)$ is the coherent probability density for the electron to be at each energy state U. The probability for the electron to have an energy u is therefore given by

$$P(U \le u \le U+\Delta U) = |f(x,y,U)|^2\Delta U. \tag{S9}$$

Using the Jacobi-Anger expansion $e^{iu\sin\varphi} = \sum_{\ell=-\infty}^{\infty}J_\ell(u)e^{i\ell\varphi}$ for the electron wavefunction in Eq. S6, with $u = 2|g(x,y,T)|$ and $\varphi = \arg\{-g(x,y,T)\} + \omega T$, we obtain the following series representation for the electron wavefunction[9],

$$\phi(x,y,T) = \phi_0(x,y,T)\sum_{\ell=-\infty}^{\infty}f_\ell(x,y,T)e^{i\ell\omega T}, \tag{S10}$$

where $f_\ell(x,y,T) = J_\ell(2|g(x,y,T)|)e^{i\ell \arg\{-g(x,y,T)\}}$. This is the most general and exact expression for the electron wavefunction after the interaction with the laser in the framework of PINEM.

Generally, the coherent electron probability density function cannot just be inferred from equation S10 because it does **not** represent the Fourier series of the electron wavefunction since the coefficients f_ℓ depend on T. However, in our scenario, the bandwidth of the initial electron wavefunction ϕ_0 is narrow enough in energy as well as the energy width of the $J_\ell^2(2|g|)$ terms, compared to the incoherent contributions discussed in the following subsection (e.g., the width of the zero-loss peak and the temporal pulse durations of the electron and the laser). In this case, the coherent electron probability density is given by

$$\rho^{\text{coherent}}(x,y,U,\Delta t) = \sum_{\ell=-\infty}^{\infty} P_\ell^{\text{coherent}}(x,y,\Delta t)\delta(U-\ell\hbar\omega) \tag{S11}$$

where $P_\ell^{\text{coherent}}(x,y,\Delta t) = J_\ell^2(2|g(x,y,\Delta t)|)$ with Δt is the time delay between the electron and laser. Now, we only need to consider the incoherent contributions to the coherent electron energy probability density function to get the final electron energy probability density. This is discussed in the following section.

1b. Incoherent contributions and the final electron probability density function

In practice, the electrons in our UTEM arrive at random times with different energies, and they are also distributed in the transverse directions according to specific distributions (e.g., Gaussian). We incorporate these incoherent contributions by convolving, in time and energy, and integrating over the transverse direction, the initial electrons *classical* probability density, denoted by $\rho_0(x,y,U,T)$, and the *coherent single* electron probability density, $\rho^{\text{coherent}}(x,y,U,\Delta t)$ and. The final electron probability density function for a given delay Δt between the electron and laser is therefore given by

$$\rho(x,y,U,\Delta t) = \iint d\mathcal{U}dT\, \rho_0(x,y,\mathcal{U},\mathrm{T})\rho^{\mathrm{coherent}}(x,y,U-\mathcal{U},\Delta t - T). \quad \text{(S12)}$$

In practice, we measure the weighted average of $\rho(x,y,U,\Delta t)$ over the transverse spatial directions $\rho(U,\Delta t) = \iint dxdy\, \rho(x,y,U,\Delta t)$. More information about the transverse part of the distribution and the rest of the incoherent broadening in Section 5c.

This expression for the electron probabilities with respect to the delay Δt is the most general formula regarding the PINEM interaction developed so far. It represents the spectrum of the electron for any given delay and can be compared to the acquired time scans discussed in the main text. Note that $\rho^{\mathrm{coherent}}(x,y,U,\Delta t)$ of Eq. S11 contains a set of delta functions centered around $U_\ell = \ell\hbar\omega$, making the final electron probability density function (denoted by $\rho(U,\Delta t)$ above) describe a continuous spectrum with distinct components around the set of energies $U_\ell = \ell\hbar\omega$.

1c. Reduction to the conventional PINEM theory

In the conventional PINEM theory[1-9], $\tilde{E}_z$ is assumed to be time-independent, and the coupling constant according to equation S7 reduces also into its time-independent version, namely, $g(x,y,T) = g(x,y)$ with

$$g(x,y) = \frac{q_\mathrm{e}}{\hbar\omega}\int_{-\infty}^{\infty} \tilde{E}_z(x,y,z')e^{-i\omega\frac{z'}{v_\mathrm{e}}}dz'. \quad \text{(S13)}$$

Substituting this coupling constant to the electron wavefunction of Eq. S10 we get

$$\phi(x,y,T) = \phi_0(x,y,T)\sum_{\ell=-\infty}^{\infty} f_\ell(x,y)e^{i\ell\omega T}, \quad \text{(S14)}$$

where $f_\ell(x,y) = J_\ell(2|g(x,y)|)e^{i\ell\arg\{-g(x,y)\}}$. These f_ℓ are time-independent, and therefore, the above series representation for the electron wavefunction is the Fourier series and we can simply write the probabilities as

$$P_\ell^{\mathrm{coherent}}(x,y) = |f_\ell|^2 = J_\ell^2(2|g(x,y|). \quad \text{(S15)}$$

To completely describe the interaction, we shall now provide the coherent electron probability density of the conventional theory. For this, we use the $P_\ell^{\text{coherent}}(x,y)$ probabilities (Eq. S15) , which resemble the probabilities presented in Eq. S11, without the longitudinal spatiotemporal dependence. Thus, the final coherent electron probability density of the conventional theory takes the form,

$$\rho^{\text{coherent}}(x,y,U) = \sum_{\ell=-\infty}^{\infty} P_\ell^{\text{coherent}}(x,y)\delta(U-\ell\hbar\omega). \tag{S16}$$

Now, to account for the incoherent contributions, we follow this paper[4] and apply the same method of the discussion in the previous subsection with one small adjustment. Since we assume the laser field is time-independent, we must "artificially" add a temporal dependence for the laser. Therefore, the formula for the electron probabilities P_ℓ is given by[9] as:

$$P_\ell^{\text{conventional}}(\Delta t) = \frac{1}{\sqrt{2\pi}\sigma_{\text{e}}}\int e^{-\frac{t^2}{2\sigma_{\text{e}}^2}} J_\ell^2\left(2|g|e^{-\frac{(\Delta t-t)^2}{2\sigma_{\text{I}}^2}}\right)dt, \tag{S17}$$

where σ_{e} is the electron pulse duration and σ_{I} is the laser pulse duration. For the case when the coupling constant is set to be just a number $g(x,y)=g$, the electron beam spatial distribution is averaged out. One can use our derivation in subsection 1b to incorporate all electron incoherent contributions (i.e., in energy, space and time) by substituting the coherent electron probability distribution for the conventional theory (Eq. S16) in our master equation (Eq. S12) which describes the general expression that incorporates both the coherent and incoherent contributions of the electron and laser. More information about the incoherent broadening is presented in Section 5c.

1d. Comparison of the conventional and extended PINEM theories

As shown in the main text (Fig. 3), the conventional PINEM theory isn't sufficient to describe the time scans we observed in this experiment. In Fig. S1 we show a comparison between

theoretical spectra produced using our extended theory and the conventional PINEM theory in respect to our experimental results.

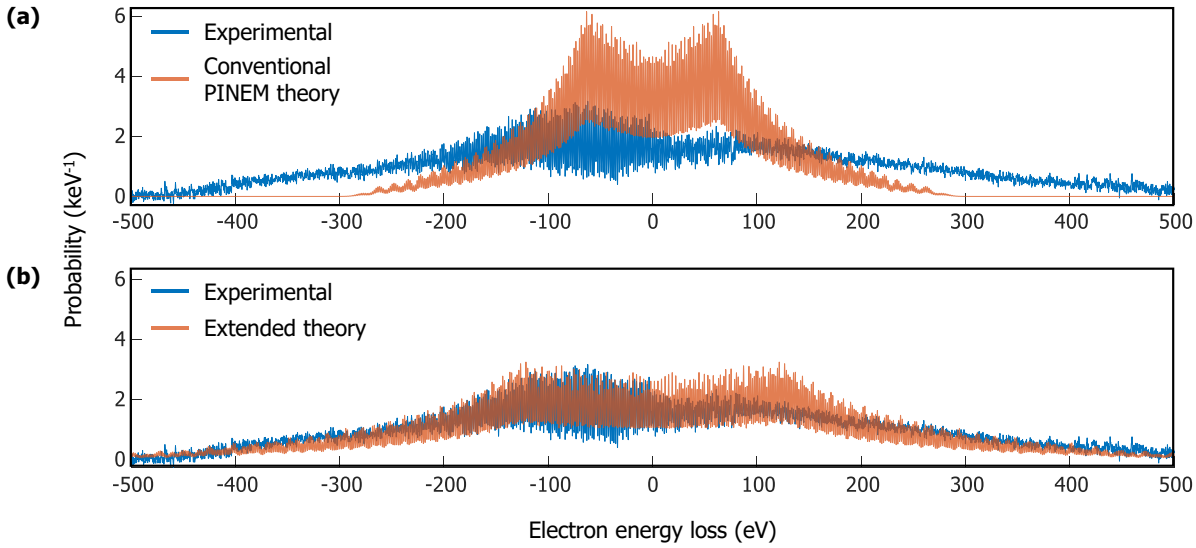

Fig. S1: Comparison of the extended PINEM theory and the conventional PINEM theory. Calculation of the electron energy spectrum by both theories for the same laser field shown in orange in both panels. The extended theory shows a coupling constant that is roughly twice that of the conventional theory. The blue curve presents the experimental electron energy spectrum at zero delay. The conventional PINEM theory does not match the experimental data as well as the extended PINEM theory.

Note 2: Electric field and coupling constant for prism geometry

In this section, we introduce the full analytical expression for the electric field interacting with the electron in our experiment. Then, we incorporate this field into the coupling constant developed in the previous section and present its closed form formula.

2a. Derivation of the electric field

In our experiment, the electron and laser pulses are two moving Gaussian pulses – a Gaussian beam in the transverse directions and a Gaussian pulse shape in both propagation direction and time. As our structure is 0.5 mm long and the Rayleigh length of the laser is $z_R = \frac{\pi w_0^2}{\lambda} \approx 1$ cm (for spot size (radius) of 50 μm and wavelength 730 nm), we neglect the divergence of the Gaussian beam (i.e., curvature $R(z) \to \infty$). Furthermore, we can approximate the Gaussian beam width to be the waist of the beam, i.e., $w(z) = w_0$. Since the interaction is not y-dependent in our case, we simplify our discussion to the $x - z$ plane (see definition of axes in Fig. S2).

The incident electric field is a Gaussian pulse moving downwards at a small angle β relative to the z axis. The pulse enters the prism after it is refracted from the hypotenuse. In order to treat the laser inside the prism as a Gaussian pulse, we first have to prove that $t_1 = t_2$ (see definition in the zoom-in of Fig. S2). The resulted laser pulse shape inside the prism is Gaussian with a projected spot size w_0' (definition below). The distances related to t_1 and t_2 are:

$$ct_1 = w_0 \tan\alpha_2 \quad , \quad \frac{c}{n} t_2 = w_0' \tan\varphi \tag{S18}$$

where $\alpha_2 = \alpha - \beta$ is the incident angle to the prism, $\varphi = \mathrm{asin}(\frac{1}{n}\sin\alpha_2)$ is the refraction angle (Snell's law), w_0 is the incident laser spot size (radius) and $w_0' = \frac{\cos\varphi}{\cos\alpha_2} w_0$ is the emerging laser spot size inside the prism.

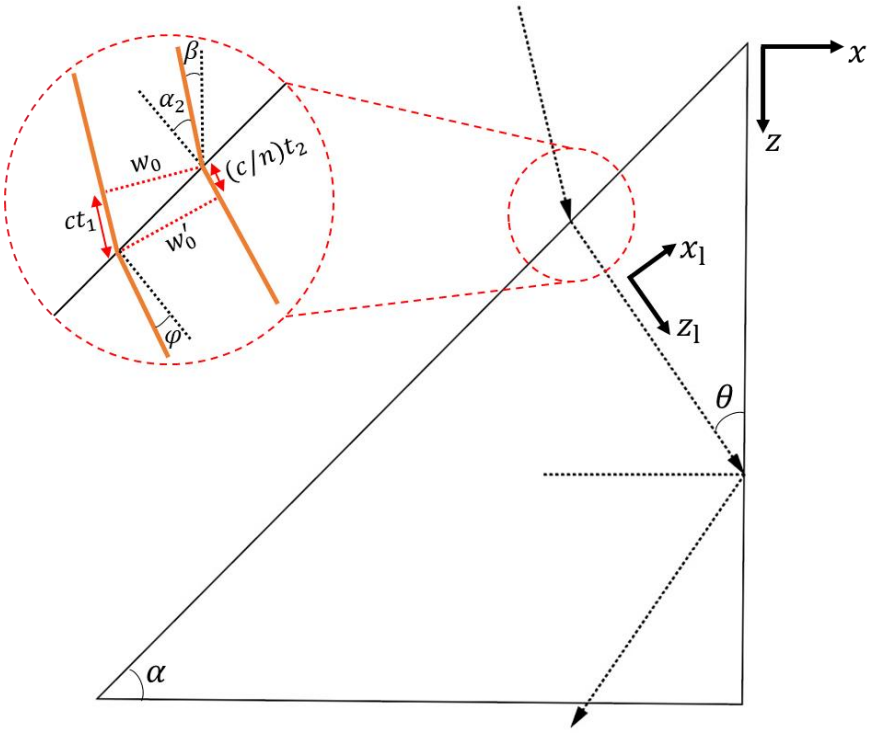

Figure S2. The incident electric field: a Gaussian pulse. An illustration of the prism setup and propagating laser pulse used in our experiments.

Following these relations:

$$t_1 = \frac{w_0}{c}\tan\alpha_2 = \frac{w_0}{c}\frac{\sin\alpha_2}{\cos\alpha_2} = \frac{nw_0}{c}\frac{\sin\varphi}{\cos\alpha_2} = \frac{nw_0'}{c}\tan\varphi = t_2. \tag{S19}$$

Therefore, As stated above, after a short time, the laser pulse shape is a Gaussian with a projected spot size w_0'. Now, the electric field inside the prism can be expressed as:

$$\mathrm{E}_{\mathrm{prism}}(\boldsymbol{r},t) = E_{\mathrm{envelope}}^{\mathrm{prism}}(\boldsymbol{r},t)e^{ink_0z_1-i\omega t}, \tag{S20}$$

where $E_{\mathrm{prism}}^{\mathrm{env}}(\boldsymbol{r},t)$ is the Gaussian envelope of the field and (x_1, z_1) are the axes of the laser propagating inside the prism (see Fig. S2) given by

$$x_1 = x\cos\theta - z\sin\theta \quad , \quad z_1 = x\sin\theta + z\cos\theta \tag{S21}$$

The Gaussian envelope is therefore given by

$$E_{\mathrm{envelope}}^{\mathrm{prism}}(\boldsymbol{r},t) = t_{p1}E_0e^{-{x_1}^2/{w_0'}^2}\, e^{-\left(\frac{z_1}{v_1}-t\right)^2/2\sigma_1^2}, \tag{S22}$$

where E_0 is the incident field amplitude, t_{p1} is the Fresnel transmission coefficient of the first prism's surface, w_0' is the laser's spot size (radius) inside the prism, $\mathrm{v_l} = \frac{c}{n}$ is the velocity of light inside the prism ($n = 1.512$ at $\lambda = 730\mathrm{nm}$) and σ_l is the laser's pulse duration. The transmission coefficient t_{p1} is given by:

$$t_{p1} = \frac{2n\cos\alpha_2}{n^2\cos\alpha_2 + \sqrt{n^2 - \sin^2\alpha_2}} \sim 0.74. \quad \text{(S23)}$$

Following this and assuming that θ is small enough such that the laser is going through total internal reflection, the field becomes evanescent outside the prism ($x > 0$). The field is then defined by its profile on the prism surface (at $x = 0$) with a new complex wavevector. The parallel component (k_z) remains continuous and the perpendicular (k_x) is changed:

$$k_z = nk_0\cos\theta = \frac{\omega}{\mathrm{v_l}}\cos\theta \quad , \quad k_x = k_0\sqrt{1 - n^2\cos^2\theta} \equiv i\kappa_x \quad \text{(S24)}$$

where $\kappa_x \equiv k_0\sqrt{n^2\cos^2\theta - 1} \in \mathbb{R}$ and $k_0 = \frac{\omega}{c}$ is the wave vector in vacuum. Finally, the z-component of the electric field in vacuum, after propagating through the prism, is given by

$$E_z(\boldsymbol{r}, t) = t_{p2}^z E_{\text{envelope}}^{\text{prism}}(x = 0, y, z, t)e^{i(k_x x + k_z z - \omega t)} = \quad \text{(S25)}$$

$$= E_0' \mathrm{e}^{-\left(\frac{z\cos\theta}{v_\mathrm{l}} - t\right)^2/2\sigma_\mathrm{l}^2} e^{-z^2/w_0''^2} e^{-\kappa_x x} e^{i(k_z z - \omega t)} = E_{\text{envelope}}(z, t)e^{-\kappa_x x}e^{i(k_z z - \omega t)}$$

where we define $w_0'' = w_0'/\sin\theta$ as the projected spot size on the prism surface The envelope amplitude is defined as $E_{\text{envelope}}(z, t) = E_0' e^{-\left(\frac{z\cos\theta}{v_\mathrm{l}} - t\right)^2/2\sigma_\mathrm{l}^2} e^{-z^2/w_0''^2}$ with $E_0' = t_{p2}^z t_{p1} E_0$. t_{p2}^z is the Fresnel transmission coefficient of the of the evanescent electric field z-component from the prism to vacuum, and is given by[11]:

$$t_{p2}^z = \frac{2n\sin\theta\sqrt{1 - n^2\cos^2\theta}}{\sin\theta + n\sqrt{1 - n^2\cos^2\theta}} \sim 0.41. \quad \text{(S26)}$$

In PINEM interactions we are only interested in the z-component of the field, as shown in the previous section. In the following section we incorporate the electric field computed above into the time-dependent coupling constant developed in the previous section.

2b. Derivation of the coupling constant

In this part, we derive a closed formula for the coupling constant which describes results of our experiment. We begin from Eq. S7 and first we simplify the electric field expression:

$$\mathrm{E_z}\left(x,y,z',\frac{z'}{v_\mathrm{e}}+t\right)=E_0' e^{-\kappa_x x} e^{-\left(\frac{z'\cos\theta}{v_\mathrm{l}}-\frac{z'}{v_\mathrm{e}}-t\right)^2/2\sigma_\mathrm{l}^2} e^{-z'^2/w_0''^2} e^{i\left(k_z-\frac{\omega}{v_\mathrm{e}}\right)z'-i\omega t} =$$

$$= E_0' e^{-\kappa_x x} e^{-\left(\left(\cos\theta-\frac{v_\mathrm{l}}{v_\mathrm{e}}\right)z'-v_\mathrm{l}t\right)^2/(2v_\mathrm{l}^2\sigma_\mathrm{l}^2)} e^{-z'^2/w_0''^2} e^{\frac{i\omega z'}{v_\mathrm{l}}\left(\cos\theta-\frac{v_\mathrm{l}}{v_\mathrm{e}}\right)} e^{-i\omega t}. \quad \text{(S27)}$$

where we used $k_z = nk_0\cos\theta = \frac{\omega}{v_\mathrm{l}}\cos\theta$. Expressing the field in terms that resemble the well-known Cherenkov emission angle formula[12] (with $\beta_\mathrm{e} = \frac{v_\mathrm{e}}{c}$):

$$\cos\theta - \frac{v_\mathrm{l}}{v_\mathrm{e}} = \cos\theta - \frac{c}{nv_\mathrm{e}} = \cos\theta - \frac{1}{n\beta_\mathrm{e}}. \quad \text{(S28)}$$

Therefore, we get:

$$E_z\left(x,y,z',\frac{z'}{v_\mathrm{e}}+t\right)=E_0' e^{-\left(\left(\cos\theta-\frac{1}{n\beta_\mathrm{e}}\right)z'-v_\mathrm{l}t\right)^2/(2v_\mathrm{l}^2\sigma_\mathrm{l}^2)} e^{-z'^2/w_0''^2} e^{-\kappa_x x} e^{\frac{i\omega z'}{v_\mathrm{l}}\left(\cos\theta-\frac{1}{n\beta_\mathrm{e}}\right)-i\omega t} \quad \text{(S29)}$$

Arranging the above expression in orders of z' gives:

$$\mathrm{E_z}\left(x,y,z',\frac{z'}{v_\mathrm{e}}+t\right)=E_0' e^{-\kappa_x x} e^{-\mathrm{c}_1^2 z'^2+c_2 z'+c_3}, \quad \text{(S30)}$$

where we define:

$$\mathrm{c}_1^2 = \frac{\left(\cos\theta-\frac{1}{n\beta_\mathrm{e}}\right)^2}{2v_\mathrm{l}^2\sigma_\mathrm{l}^2} + \frac{1}{w_0''^2} \quad \text{(S31)}$$

$$c_2 = \frac{i\omega}{v_\mathrm{l}}\left(\cos\theta-\frac{1}{n\beta_\mathrm{e}}\right) + \frac{t}{v_\mathrm{l}\sigma_\mathrm{l}^2}\left(\cos\theta-\frac{1}{n\beta_\mathrm{e}}\right) = \left(\frac{i\omega}{v_\mathrm{l}}+\frac{t}{v_\mathrm{l}\sigma_\mathrm{l}^2}\right)\left(\cos\theta-\frac{v_\mathrm{l}}{v_\mathrm{e}}\right)$$

$$c_3 = -i\omega t - \frac{t^2}{2\sigma_{\mathrm{l}}^2}.$$

Now, we can calculate the integral in $g(x, y, \mathrm{t})$:

$$g(x,y,t) = \frac{\mathrm{q_e}}{\hbar\omega}\int_{-\infty}^{\infty} \mathrm{E_z}\left(x, y, z', \frac{z'}{v_\mathrm{e}} + t\right) \mathrm{d}z' = \frac{\mathrm{q_e}E_0'}{\hbar\omega} e^{-\kappa_x x} e^{c_3} \int_{-\infty}^{\infty} e^{-\mathrm{c}_1^2 z'^2 + c_2 z'}\, dz'. \quad \text{(S32)}$$

Note that

$$\int_{-\infty}^{\infty} e^{-\mathrm{c}_1^2 z'^2 + c_2 z'}\, dz' = \int_{-\infty}^{\infty} e^{-\left(c_1 z' - \frac{c_2}{2c_1}\right)^2 + \left(\frac{c_2}{2c_1}\right)^2}\, dz' = e^{\left(\frac{c_2}{2c_1}\right)^2} \int_{-\infty}^{\infty} e^{-c_1^2\left(z' - \frac{c_2}{2c_1^2}\right)^2}\, dz', \quad \text{(S33)}$$

and using the Gaussian integral $\int_{-\infty}^{\infty} e^{-a(x-b)^2}\, dz' = \sqrt{\pi/a}$ with $a = c_1^2$ and $b = \frac{c_2}{2c_1^2}$, the integral above reduces to

$$\int_{-\infty}^{\infty} e^{-\mathrm{c}_1^2 z'^2 + c_2 z'}\, dz' = \frac{\sqrt{\pi}}{c_1} e^{\left(\frac{c_2}{2c_1}\right)^2}. \quad \text{(S34)}$$

Finally, we arrive at the final expression for g:

$$g(x,y,t) = \frac{\sqrt{\pi}}{c_1}\frac{\mathrm{q_e}E_0'}{\hbar\omega} e^{\left(\frac{c_2}{2c_1}\right)^2 + c_3} e^{-\kappa_x x}. \quad \text{(S35)}$$

For convenience, here are the important quantities that were defined during the derivation and appear in Eq. S35:

$$E_0' = t_{p2}^z t_{p1} E_0 \quad \text{(S36)}$$

$$t_{p1} = \frac{2n\cos\alpha_2}{n^2\cos\alpha_2 + \sqrt{n^2 - \sin^2\alpha_2}} \sim 0.74$$

$$t_{p2}^z = \frac{2n\sin\theta\sqrt{1 - n^2\cos^2\theta}}{\sin\theta + n\sqrt{1 - n\cos^2\theta}} \sim 0.41$$

$$\kappa_x \equiv k_0\sqrt{n^2\cos^2\theta - 1} \in \mathbb{R}.$$

$$\mathrm{c}_1^2 = \frac{\left(\cos\theta - \frac{1}{n\beta_\mathrm{e}}\right)^2}{2v_\mathrm{l}^2\sigma_\mathrm{l}^2} + \frac{1}{w_0''^2}$$

$$c_2 = \frac{i\omega}{v_\mathrm{l}}\left(\cos\theta - \frac{1}{n\beta_\mathrm{e}}\right) + \frac{t}{v_\mathrm{l}\sigma_\mathrm{l}^2}\left(\cos\theta - \frac{1}{n\beta_\mathrm{e}}\right) = \left(\frac{i\omega}{v_\mathrm{l}} + \frac{t}{v_\mathrm{l}\sigma_\mathrm{l}^2}\right)\left(\cos\theta - \frac{1}{n\beta_\mathrm{e}}\right)$$

$$c_3 = -i\omega t - \frac{\mathrm{t}^2}{2\sigma_\mathrm{l}^2}$$

Note: in perfect phase-matching condition, i.e. $\cos\theta = \frac{1}{n\beta_\mathrm{e}}$, we have $c_1 = \frac{1}{w_0''}$ and $c_2 = 0$.

Therefore, the final expression for $g(x, y, t)$ reduces to:

$$g(x, y, t) = \sqrt{\pi}\frac{\mathrm{q_e}E_0'}{\hbar\omega} e^{c_3} e^{-\kappa_x x} w_0''. \tag{S37}$$

Note 3: Comparison of quantum and classical interaction

In this section, we show the similarities and differences between two different theories for the electron–laser interaction, one of which is a derivation based on quantum mechanics and the second relies on classical electrodynamics. First, we consider a classical interaction using the Lorentz force. Then, we describe the quantum interaction (conventional PINEM theory) and compute its envelope analytically. Finally, we compare the two theories and discuss the similarities and differences between them.

3a. Classical interaction

We describe the electron as a classical point particle travelling in an electric field and aim to find the change in its kinetic energy during the interaction. The equation for the electron velocity under an electric force parallel to its direction of motion is given by the Lorentz force:

$$m_{\mathrm{e}}\frac{dv}{dt} = q_{\mathrm{e}}\, Re\{2\tilde{E}_z(z)e^{-i\omega t}\}, \tag{S38}$$

where $\tilde{E}_z(z)$ is the electric field phasor, as denoted in the PINEM derivation in Section 1. We calculate the change in electron velocity by integration over time:

$$m_{\mathrm{e}}\Delta v = q_{\mathrm{e}}\int dt\, Re\{2\tilde{E}_z\big(z(t)\big)e^{-i\omega t}\}. \tag{S39}$$

Assuming the change in electron velocity is negligible during the interaction, i.e., $v = v_{\mathrm{e}} + \Delta v,\ \Delta v \ll v_{\mathrm{e}}$, we can write $t \approx \frac{z-z_0}{v_{\mathrm{e}}}$, where z_0 is the initial position of the electron (at $t = 0$). Under this approximation, the change in the electron kinetic energy can be written as $\Delta E_k = m_{\mathrm{e}}v_{\mathrm{e}}\Delta v$, and changing variables from t to z in Eq. S39 yields

$$\Delta E_k = q_e\int dz\, Re\left\{2\tilde{E}_z(z)e^{-i\frac{\omega}{v_{\mathrm{e}}}(z-z_0)}\right\}. \tag{S40}$$

We define $W_g \equiv q_e\int dz\tilde{E}_z(z)e^{-i\frac{\omega}{v_{\mathrm{e}}}z}$, and therefore, $\Delta E_k = Re\left\{2W_g e^{i\frac{\omega}{v_{\mathrm{e}}}z_0}\right\}$. Note that W_g is equal to the coupling constant g (see Eq. S13 in Section 1a) up to a factor of the photon

energy; namely, $W_g = g\hbar\omega$. To complete the derivation, we assume that the initial electron position z_0 is uniformly distributed over one optical wavelength/cycle of the laser field and thus, $z_0 \sim U\left(0, \frac{2\pi v_e}{\omega}\right) \equiv f_{\text{dist}}(z_0)$. The probability density of the energy change can then be calculated:

$$P_{\text{classical}}(\Delta E_k) = \left.\frac{f_{\text{dist}}(z_0)}{\left|\frac{d\Delta E_k}{dz_0}\right|}\right|_{z_0 = z_0(\Delta E_k)} = \frac{1}{2\pi|W_g|\sqrt{1-\left(\frac{\Delta E_k}{2|W_g|}\right)^2}}. \quad \text{(S41)}$$

Equation S41 describes the electron energy distribution in the classical case. To enable a comparison with the quantum theory presented later, we convert the energy variable ΔE_k to a continuous dimensionless variable ℓ by setting $\Delta E_k = \ell\hbar\omega$, which, in the quantum case, denotes the net number of photons emitted/absorbed during the interaction. Finally, the classical probability for the electron to change its kinetic energy by $\Delta E_k = \ell\hbar\omega$ is given by

$$P_{\text{classical}}(\ell) = \frac{1}{2\pi g\sqrt{1-\left(\frac{\ell}{2g}\right)^2}}. \quad \text{(S42)}$$

One important difference to mention is that the variable ℓ is continuous in the classical sense, whereas in the quantum description it represents the number of photons and is thus discrete.

3b. Quantum interaction

We describe the quantum electron–laser interaction using the framework of conventional PINEM theory. Under a conventional PINEM interaction with a CW laser[1-4], the probability for an electron to gain/loss an energy quanta of $\Delta E_k = \ell\hbar\omega$ is given by $P_\ell = |f_\ell|^2$, where $f_\ell = e^{i\ell arg(-g)}J_\ell(2|g|)$ are the probability amplitudes depending on the coupling constant g defined in Eq. S13 in Section 1a. The peaks in the measured electron energy spectrum are determined by these probabilities; namely, $P_\ell = J_\ell^2(2|g|)$.

To compare the quantum and classical descriptions, we present these probabilities in the form of a carrier-envelope function. The carrier is responsible for the peaks, while the

envelope can be compared to the classical point of view. The electron spectrum (assuming a monochromatic interaction) can be described by section 1 of this supplementary by $f(U)$:

$$\rho(U) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dt e^{iUt+2ig\,sin(\omega t)}, \tag{S43}$$

where the electron amplitude $\rho(U)$ is a function of a continuous energy variable U. Here, ω is the laser photon frequency, and we assume $arg(g) = 0$ for simplicity. We are interested in calculating the spectral distribution of an electron after a strong interaction, i.e., an interaction described by a large coupling constant $g \gg 1$. By using the method of saddle point approximation (SPA), the asymptotic approximated solution is given by

$$\rho(U) \sim \sqrt{\frac{-2\pi}{\Omega q''(t_s)}}u(t_s)e^{q(t_s)} \ , \tag{S44}$$

where $u(t) = \frac{1}{2\pi} = \text{const}$ and $q(t) = Uit/\hbar + 2ig\sin(\omega t)$. We find that there are infinitely many saddle points t_s satisfying the relation $\cos(\omega t_s) = -\frac{U}{2g\hbar\omega}$. This is true for large g values, such that $U/\hbar\omega < 2g$, as in the regime we are currently analyzing. Indeed, for large g, most of the spectrum is in this range, and we notice that it is also possible to use the SPA to analyze the edges of the spectrum. The approximated solution is given by the following infinite sum:

(S45)

$$\rho(U) \sim \sqrt{\frac{1}{\pi g}}\sum_{\ell=-\infty}^{\infty}\delta(U-\ell\hbar\omega)\left(1-\left(\frac{\ell}{2g}\right)^2\right)^{-1/4}\cos\left(-\ell\cos^{-1}\left(\frac{\ell}{2g}\right)+2g\sqrt{1-\left(\frac{\ell}{2g}\right)^2}-\frac{\pi}{4}\right).$$

This is a form of the **Debye approximation** for a Bessel function. Squaring the above expression yields the probability density of gaining/losing an energy of $\ell\hbar\omega$. Finally, we obtain the following function for the electron probability amplitudes under the SPA:

$$P_{\text{quantum}}^{\text{SPA}}(\ell) \sim \frac{1}{\pi g}\left(1-\left(\frac{\ell}{2g}\right)^2\right)^{-\frac{1}{2}}\cos^2\left(-\ell\cos^{-1}\left(\frac{\ell}{2g}\right)+2g\sqrt{1-\left(\frac{\ell}{2g}\right)^2}-\frac{\pi}{4}\right). \tag{S46}$$

The spectrum is therefore governed by an envelope function $\left(\pi g\sqrt{1-\left(\frac{\ell}{2g}\right)^2}\right)^{-1}$.

3c. Similarities and differences between the two descriptions

Finally, we compare the quantum and classical descriptions. The classical result given in Eq. S42 is found to be one half of the quantum envelope function derived via the SPA (Eq. S46). This factor of two can be explained by performing an average over all discrete peaks of the squared cosine terms. In Fig. S3, we show a comparison of the classical and quantum approaches for this electron–laser interaction.

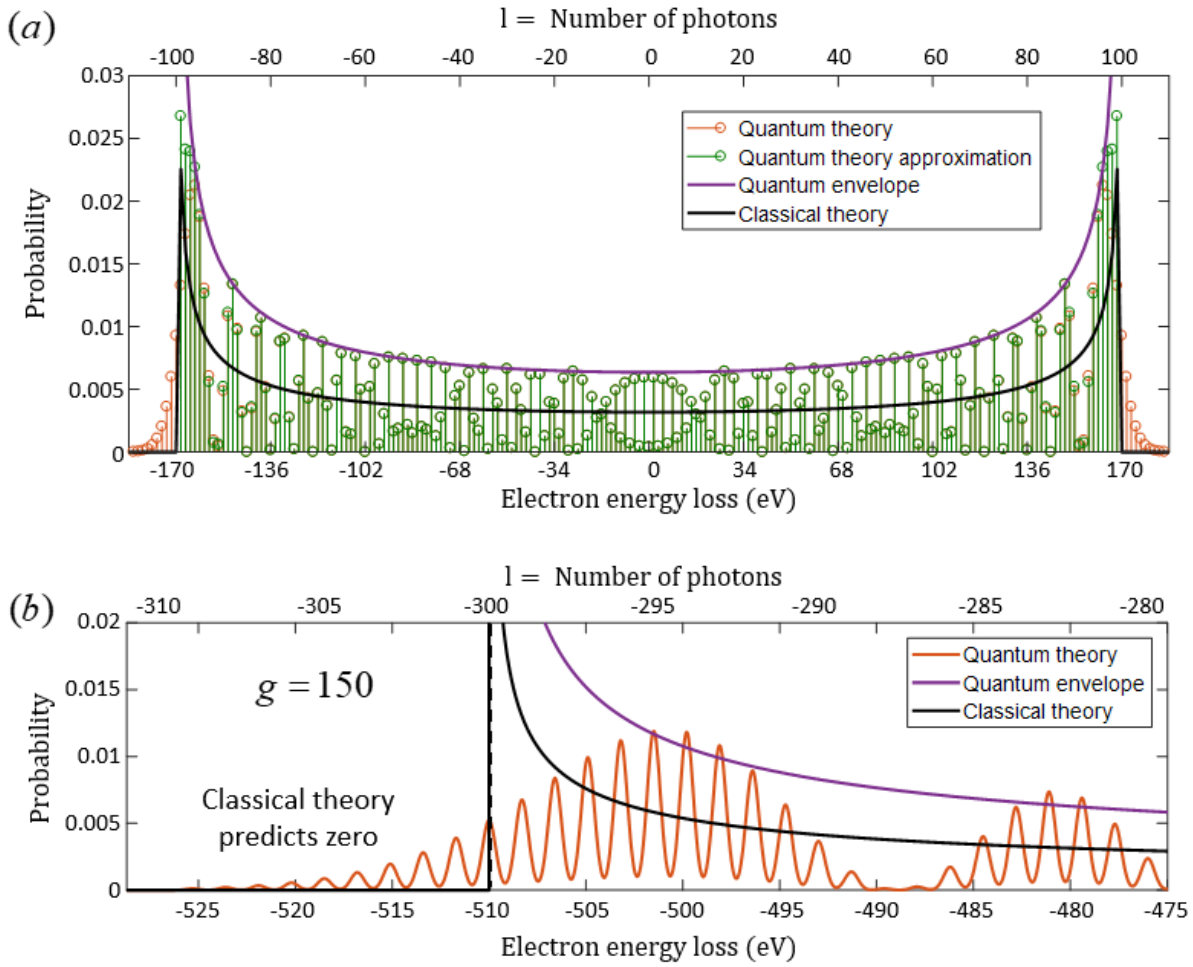

Figure S3. Comparison of the classical and quantum descriptions for electron–laser interactions. (a) The quantum theory (conventional PINEM) alongside its saddle point approximation (SPA), plotted for $g = 50$. The SPA graph provides a good estimation for the PINEM solution in the range $|\ell| < 2|g|$. The SPA's quantum envelope is also presented. The classical result is one half the quantum envelope, appearing as an average over the quantum peaks. (b) Electron energy loss

spectrum after the interaction with $g = 150$, $\hbar\omega = 1.7eV$ around the point U $= 2\hbar\omega|g|$. Whereas the classical theory predicts a spectrum maximum at this exact point (-510 eV in our graph), the real maximum is obtained earlier.

An important difference between the classical and quantum theory can be found at the far edges of the spectrum ($\ell \geq 2|g|$); the classical spectrum is identically zero in this range, whereas the quantum spectrum yields finite values. Consequently, we expect the quantum theory to be important for describing the exact electron acceleration/deceleration and especially for analyzing the regimes of maximal acceleration/deceleration. This comparison emphasizes the importance of the quantum theory for laser-based electron acceleration, as in dielectric laser accelerators (DLA)[7].

To summarize this section, we note that the usage of SPA provides a connection between the quantum and classical descriptions for electron–laser interactions. Both approaches yield envelopes that are identical up to a factor of 2 over the range $\ell < 2|g|$. Most differences emerge when fine features (e.g., the quantized energy peaks) are examined, as expected when one compares a classical and a quantum theory.

Note 4: Intuition for the phase-matching effect

In this section, we provide intuition for the phase-matching condition needed to achieve efficient coupling in electron–light interactions. The key factor of the record strong interaction achieved in this work is maintaining this condition satisfied over long distance.

In general, free electrons are slower than light in free space, leading to an energy–momentum mismatch that prevents a strong interaction. In this scenario, the electron interacts with alternating field directions during its motion, and the net effect cancels out after each optical wavelength/cycle. In our experiment, the electron velocity v_e is matched to the light phase velocity along the electron trajectory ω/k_z, and thus, the electron interacts with a constant field instead of an alternating one, which enables the interaction to accumulate for a long distance. Thus, the interaction becomes stronger by orders of magnitude, with a net effect that increases linearly with the interaction length. To match the phases of the electron and light waves, the light is slowed down by propagating through a medium (e.g., the prism).

For simplicity, it is convenient to discuss plane waves instead of the Gaussian pulses that are used in the experiment (for the full treatment of PINEM with Gaussian pulses see Sections 1 and 2). In this section, we describe the laser by a localized plane wave: $\tilde{E}_z(\boldsymbol{r}) = E_0' e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$ with frequency $\omega = c|\boldsymbol{k}|$ and $E_0' \neq 0$ for $-\frac{L}{2} < z < \frac{L}{2}$. E_0' is the electric field amplitude after going through the prism and we denote the interaction length by L, which relates to the projected spot size w_0'' on the prism's surface (see definitions of $E_0{}'$ and w_0'' in Eq. S25 of Section 2a below). Plugging this field into Eq. 1 of the main text (the same as Eq. S13 in section 1c here), we can calculate the coupling constant explicitly:

$$g(x,y) = \frac{q_\mathrm{e} E_0'}{\hbar\omega} e^{ik_x x + ik_y y} \int_{-\mathrm{L}/2}^{L/2} e^{i\left(k_z - \frac{\omega}{v_\mathrm{e}}\right)z'} dz'. \tag{S47}$$

We define $\Delta k = k_z - \frac{\omega}{v_\mathrm{e}}$ as the phase-mismatch parameter and solve the integral to obtain

$$g(x,y) = \frac{q_e E_0'}{\hbar\omega} e^{ik_x x + ik_y y} \mathrm{sinc}\left(\frac{\Delta k L}{2}\right) \cdot L. \tag{S48}$$

The phase-matching condition can be identified as $\Delta k = 0$ ($k_z = \omega/v_e$). This condition requires k_z to be larger than the wave-vector magnitude in vacuum ($k_0 = \omega/c$) because the electron is slower than light ($v_e < c$).Thus, to satisfy this condition, we must use an evanescent wave so the electric field takes the form $E_z(\boldsymbol{r}) = E'_0 e^{-\kappa_x x + ik_z z}$ (we set $k_y = 0$ and $k_x = i\kappa_x$ for $\kappa_x > 0$). Our coupling constant becomes

$$g(x_0) = \frac{q_e E_0'}{\hbar\omega} e^{-K_x x_0} \mathrm{sinc}\left(\frac{\Delta k L}{2}\right) \cdot L, \tag{S49}$$

where $K_x = \sqrt{k_z^2 - k_0^2}$.

In Fig. S4, we show the coupling constant ($|g|$) as a function of Δk and L. For perfect phase-matching ($\Delta k = 0$), $|g|$ increases linearly with the interaction length L, whereas in the phase-mismatch scenario we see the expected periodic behavior (Fig. S4a). For a fixed interaction length, $|g|$ changes periodically as a function of Δk with decreasing values as we deviate from perfect phase-matching, i.e. $\Delta k = 0$ (Fig. S4b).

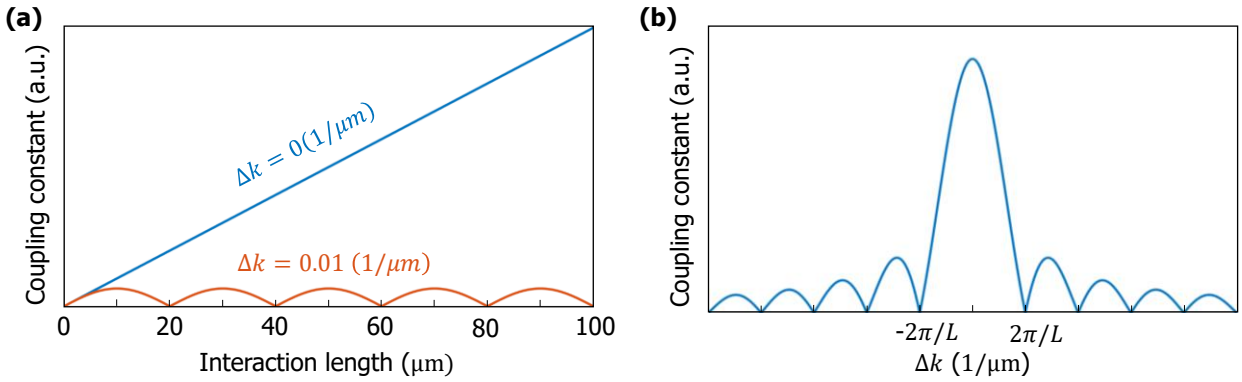

Figure S4. The dependence of the coupling constant on the phase-mismatch and the interaction length. (a) The coupling constant ($|g|$) as a function of the interaction length L for a phase-matched (blue) and phase-mismatched (orange) interaction. **(b)** The coupling constant as a function of the phase-mismatch parameter (Δk). The bandwidth of the phase-matching condition (width of main lobe) is inversely proportional to the interaction length L, meaning smaller interaction lengths "accept" more phase-mismatch than longer interaction lengths.

In comparison with all previous PINEM experiments, we design our experimental scheme such that the entire field has a single k_z value throughout the whole interaction length. From simple geometrical calculations of our setup, we obtain $k_z = nk_0\cos\theta$ (see notations in Fig. 2 of the main text or Fig. S2 here), which reduces the phase-matching condition to a formula that resembles the Cherenkov emission angle formula (using $\beta_\mathrm{e} = v_\mathrm{e}/c$):

$$\cos\theta = \frac{1}{n\beta_\mathrm{e}}. \tag{S50}$$

The phase-matching condition can also be understood as arising from the conservation of energy–momentum between the electron and the light. It is the most efficient energy transfer mechanism between the electron and light, in the sense that the spatial Fourier space representation of the pump laser field inside the prism is described by a single $k_z = \omega/v_\mathrm{e}$ component that carries all of the field energy. Using the precise electron alignment and the field tilt by the prism, we match the electron velocity and this exact k_z value so that the interaction strength is maximized.

To understand the importance of satisfying the phase-matching condition over a long distance in PINEM experiments, it is valuable to compare our experimental scheme with previous PINEM experiments. Previous PINEM interactions involved *localized* nearfields, providing much weaker interactions since the interaction length is cannot exceed the single optical wavelength (typically hundreds of nanometers). Due to their limited interaction length, all previous PINEM experiments can be understood as variants of laser-driven *quantum transition radiation* or *quantum stimulated transition radiation*[4,5,6], in which the electron wavefunction alters the interaction (e.g., quantized energy exchange). In contrast, we report here of the first *extended* interaction measured in UTEM with much longer interaction over hundreds of optical wavelengths/cycles (hundreds of microns). Our longer interaction length and duration is the reason our experiment can be titled as a type of *stimulated*

Cherenkov experiment (also called inverse-Cherenkov effect) near a planar interface. As with the PINEM analogies of transition radiation, we find a dependence on the electron wavefunction.

The comparison between the classical transition radiation and Cherenkov radiation (i.e., spontaneous process) perfectly corresponds to the comparison between the conventional PINEM experiments localized interactions and the extended phase-matched PINEM experiment in this work. Note that in general transition radiation effects are typically weaker than Cherenkov radiation effects because of the limited interaction length. The Cherenkov phase-matched interaction that is maintained over hundreds of microns is what makes classical Cherenkov-type effects stronger than transition radiation effects. In complete analogy, the phase-matching makes our PINEM Cherenkov-type interaction stronger than all previous PINEM experiments that were all transition-radiation-type.

Interestingly, despite this classical-quantum comparison, the quantum case shows unique features that do not have a classical analogue. For example, the strong energy exchange in our inverse Cherenkov quantum interaction reveals a wide energy *plateau* consisting of quantized energy peaks – in contrast to a flat plateau in the classical case.

Note 5: Additional information on the data analysis

In this section, we describe the necessary data analysis that we had to perform on the raw experimental data. The three key challenges were as follows.

5a. Correlations to combine multiple energy spectra

The strong interaction leads to electron energy spectra that are wider than the physical size of the detector, i.e., the electron energy loss spectrometer (EELS). One could change the detector energy dispersion (focusing the electrons on a smaller part of the spectrometer), but that would cause a reduction in the energy resolution due to the finite number of pixels in the detector. Instead, we used correlation methods to reproduce the full image of interaction with a sufficiently high resolution. We collected the data (e.g., Fig. 3b and 5a-b) in several segments, with a partial overlap. We use these partial overlaps to combine all the acquired spectra into one full spectrum using correlation. This method was proved useful to account for possible inaccuracies of the detector movement between subsequent measurements that influence the measured energy values. In the end of each such process, we have the tailored full spectrum describing the full image of the interaction that occurred.

5b. Fixing fluctuations in the time scan data

A time scan is composed of several electron energy spectra, each describing the interaction with respect to different time delay between the electron and the laser pulses. The result is a map of the electron energy spectrum as a function of the delay between the electron and laser pulses, as presented in Fig. 3 and 5c.

The long acquisition times cause shifts in the electron energy spectra due to inevitable instabilities. These shifts are especially important when tailoring several spectra segments together (as explained in Section 5a above) or when performing time scans. A time scan acquisition takes much longer than a single spectrum acquisition, therefore, fixing system

fluctuations is crucial. The two obstacles we had to overcome were (1) normalization of each spectrum in the time scan, and (2) energy shifts during the measurement due to long acquisition times.

The experimental time scan map presented in Fig. 3a (or larger in Fig. 5c) of the main text underwent several steps of signal processing. Since each row in the time scan is a spectrum for a specific delay, it should represent the electron interaction probability, hence needs to be normalized. For most of the time scan we could easily normalize the integration to unity. However, for delays with strong interaction a significant part of the spectrum exceeded the detector limits, and could not be normalized appropriately. To ensure good image contrast, we normalize these partial spectra with the result of its integration along the measured energy range (-100 eV to 100 eV). To compensate over the energy fluctuations, we determined the zero energy loss position for each delay and shift the energy axes accordingly.

Regarding fluctuations in the electron current, the correction methods used in previous papers on PINEM (e.g., refs. 4-9) are not effective here, because the energy spectrum is wider than the detector size. This poses extra challenges in data analysis, since we cannot always normalize the total probability to account for the fluctuations in the electron current, as stated above.

5c. The initial electrons' classical probability density

In this section, we discuss the dependence of the coupling constant on the transverse spatial coordinates, which in the special case of our prism grazing-angle experiment captures the effect of the evanescent tail. We show why it is crucial to consider this spatial dependence in our setup and observations. This analysis may also be valuable for future grazing-angle experiments.

In our experiment, the interaction between the electron and the laser is strongly affected by the distance between the electron and the interface. This distance (denoted by x) has a huge impact on the coupling constant[3] since the evanescent field exponentially decays in this direction. To fully account for the spatial dependence of the coupling constant, we also need to consider the transverse spatial dependence of the electron beam. To do that, we introduce the normalized initial electrons probability density $\rho_0(x,y,U,T)$, It includes the spatial distribution of the electron in the transverse directions x,y, the propagating Gaussian electron pulse in the longitudinal direction (depends solely on $T=t-z/v_\mathrm{e}$) and the electron incoherent energy broadening, called the zero-loss peak (ZLP).

Now, we shall complete the discussion in Section 1b and provide the master equation for $\rho(U,\Delta t)$ from Eq. S12. We consider a special case (that can be directly generalized) in which the $\rho_0(x,y,U,T)$ can be decomposed to $\rho_0(x,y,U,T)=A(U)B(x,y)C(T)$. $A(U)$ denotes the ZLP, $B(x,y)$ denotes the static the transverse directions probability density in x,y, and $C(T)$ denotes the spatiotemporal incoherent probability density in the longitudinal direction. Finally, we arrive at the **master equation** for extended PINEM interactions by substituting the decomposed expression of $\rho_0(x,y,U,T)$ into Eq. S12 and also integrating over the transverse directions x,y. Therefore, we get:

$$\rho(U,\Delta t)=\iiint A(\mathcal{U})B(x,y)C(T)\rho^{\mathrm{coherent}}(x,y,\mathcal{U}-U,\Delta t-T)d(x,y)d\mathcal{U}dT, \qquad \text{(S51)}$$

where $\rho^{\mathrm{coherent}}(x,y,U,\Delta t)=\sum_{\ell=-\infty}^{\infty}P_\ell^{\mathrm{coherent}}(x,y,\Delta t)\delta(U-\ell\hbar\omega)$ is given in Eq. S11 and $P_\ell^{\mathrm{coherent}}(x,y,\Delta t)=J_\ell^2(2|g(x,y,\Delta t|)$ is the coherent probability to gain/loss energy quanta of $\ell\hbar\omega$.

In the case of our prism experiment, we can assume the spatiotemporal parts, $B(x,y)$ and $C(T)$, to be Gaussian functions. To keep the discussion as much general as we can, we we

shall keep using the energy general function by $A(U)$. Therefore, the initial electrons probability density takes the form

$$\rho_0(x,y,U,T) = \mathcal{A}(U)\, \mathrm{e}^{-\frac{(x-x_\mathrm{e})^2+(y-y_\mathrm{e})^2}{w_\mathrm{e}^2}}\, \mathrm{e}^{-\frac{T^2}{2\sigma_\mathrm{et}^2}}, \tag{S52}$$

Where $\mathcal{A}(U) = \frac{1}{D}A(U)$ with D being the overall normalization factor of the spatiotemporal gaussians, $(x_\mathrm{e}, y_\mathrm{e})$ describe the center of the electron beam in the x-y plane, w_e is the electron spot radius (where the amplitude falls to $1/e$ of the maximum) and σ_et is the electron pulse duration (standard deviation). Note that we use standard deviation notation for temporal quantities and radius/waist notation for spatial quantities, for both the electron and laser.

Since the electric field in our experiment is independent of y, the y dependence of $\rho_0(x,y,U,T)$ averages out in the integration. Now we can substitute the above expression (Eq. S52) in the master equation (Eq. S12) and we arrive to **the master equation describing our interaction**. Note the total probability is preserved, namely, $\int \rho(U,\Delta t)d(U,\Delta t) = 1$. This holds since $\rho_0(x,y,U,T)$ is normalized and $\sum_\ell P_\ell = 1$ (from the Bessel identity[13] $\sum_\ell J_\ell^2(u) = 1$).

The parameter x_e controls the distance of the electron beam from the prism, and thus, large x_e will suppress the interaction. The lower limit of the integration, x_0, comes from the spread angle of the electron beam $\Delta\theta$ (measured to be ~1 mrad in our experiment) that truncates the interaction distance. All electrons that are closer than $x_0 = \frac{L_\mathrm{prism}}{2}\tan\Delta\theta \approx 250\mathrm{nm}$ to the prism are being blocked by the prism and do not reach the detector at all ($L_\mathrm{prism} = 500\mu m$ is the prism side's length). Thus, we integrate from x_0 to infinity and normalize P_ℓ accordingly to preserve the probability. In the experiment, we blocked half of the electron beam to balance a strong enough signal with the closest proximity to the prism; namely, we set $x_\mathrm{e} = x_0$. From fitting our extended PINEM model to the experimental data

(see Section 5e), we suspect the electron beam might have been further away from the prism than assumed, putting numbers we get $x_e = 2x_0 \approx 500\text{nm}$. This is just an estimation since we have a lot of unknown parameters to fit and some of them are entangled (as explained in Section 5e). Because of the evanescent decay of the field, the system is very sensitive to this distance and that's what makes the consideration of the spatial dependence of the coupling constant crucial for this work. Figure S5 demonstrates the effect of this spatial expansion and its significant impact on the electron energy spectrum.

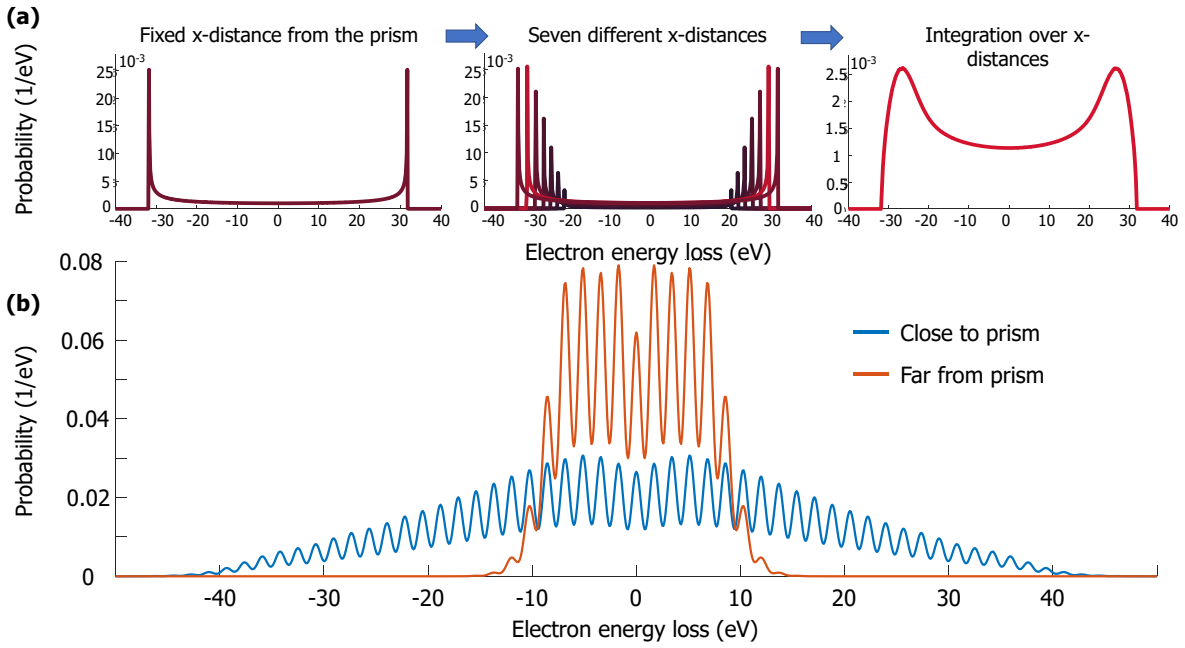

Figure S5. Effects of the electron beam transverse distribution in grazing angle electron–laser experiments. (a) The quantum envelope of the PINEM interaction with and without integration over the electron distribution along the *x*-axis. This integration transforms the two sharp peaks into two smooth peaks and brings them toward the center, as shown in the top three panels (left to right). The rightmost graph resembles the shape of our experimental spectrum envelope, giving validity to our expansion of the theory. Note that the sharp points in the left and center panels of (a) are in fact divergences, and the height differences correspond to the weight they receive in the integration, as can be seen from Eq. S52. **(b)** Theoretical energy spectrum for the interaction between coherent free electrons and an electromagnetic field with an electron beam that is close to the prism (blue) versus electron beam that is far away from the prism (orange). We can see the exponential decrease of the interaction strength with the electron beam distance, as expected from the nature of the evanescent

field. This figure highlights the importance of aligning the system with great precision, so the electron will travel as close as possible to the surface of the prism for the longest possible distance.

5d. Side lobes in the time scan map

In practice, the laser's temporal dependence is not a perfect Gaussian. We performed an autocorrelation measurement and note that it contains two symmetrical small side lobes from each sides of the central Gaussian peak which possibly arises from our optical parametric amplifier (OPA). Since our interaction is very strong, we observed the interaction between the electron and the much weaker side lobes related electric fields, compared to the central peak related electric field. This result is described in Fig. 3 of the main text. We modeled these side lobes as two symmetrical small-amplitude Gaussians, delayed by $\pm t_0^{\mathrm{side}}$ relative to the center Gaussian with an amplitude denoted by E_0^{side} and pulse duration denoted by $\sigma_{\mathrm{l}}^{\mathrm{side}}$. The resulting electron probabilities (defined below Eq. S11 in the end of Section 1a) is the sum of the center peak interaction and these two side lobes contributions. Thus, we get

$$P_\ell^{\mathrm{coherent}}(x,\Delta t) = J_\ell^2\left(2\left|g(x,\Delta t) + g^{\mathrm{side}}\left(x,\Delta t - t_0^{\mathrm{side}}\right) + g^{\mathrm{side}}\left(x,\Delta t + t_0^{\mathrm{side}}\right)\right|\right), \quad \text{(S53)}$$

where g describes the central peak interaction and g^{side} describes the side lobes contributions.

5e. Fitting theory to the experimental results

We fit the experimental time scan presented in Fig. 3a (and Fig. 5c) of the main text with the theoretical time scan computed from Eq. 4 of the main text (same as Eq. S12 in Section 1a here). The estimated parameters are presented in Table S1 below. The fit also gives us several quantities that we were unable to measure directly. We employed an optimization method using nonlinear programming solver with the following parameters: the laser amplitude $E_0 = E_0' e^{-\kappa_x x}$, laser pulse duration σ_{l}, laser spot size w_0, electron pulse duration σ_{e}, and electron beam size w_{e}. We also include three parameters to describe the laser temporal side lobes (discussed in Section 5d): amplitude E_0^{side}, pulse duration $\sigma_{\mathrm{l}}^{\mathrm{side}}$, and time

delay t_0^{side}. The final result is presented in Fig. 3d of the main text and the fit parameters are given here in Table S1.

Name	Variable	Value	FHWM Value
Electron pulse duration	σ_{e}	179 fs	421 fs
Electron beam radius	w_{e}	100 nm	167 nm
Laser pulse amplitude	E_0	1.680×10^6 V/m	-
Laser pulse duration	σ_{l}	273 fs	642 fs
Laser waist	w_0	50.21 μm	83.61 μm
Side lobes amplitude	E_0^{side}	0.0472×10^6 V/m	-
Side lobes duration	$\sigma_{\mathrm{l}}^{\mathrm{side}}$	823 fs	1939 fs
Side lobes delay	t_0^{side}	664 fs	-

Table S1. The optimal fit parameters found from fitting the experimental time scan (in Fig. 3a of the main text) with the extended PINEM theory (in Fig. 3d of the main text). Note that we use standard deviation notation for temporal quantities and radius/waist notation for spatial quantities, for both the electron (defined in Section 5c) and laser (defined in Section 2). To compare with common experimental notation full-width half-max (FWHM) we add a "FWHM value" column which is calculated using $\mathrm{FWHM} = 2\sqrt{\ln 2}w$ and $\mathrm{FWHM} = 2\sqrt{2\ln 2}\sigma$. Also, note that the waist of the laser is much larger than the electron beam radius, and we take it with higher precision since each micron alters the interaction significantly in our extended phase-matched interaction.

These parameters match with the experimental values measured in other ways: estimated laser pulse duration of 600 fs FWHM, side lobes delay of 2 ps FWHM, waist of the laser ~ 100 μm). The field amplitude found from the optimization matches (to within 25%) with the experimental value when considering the exponential decay of the field, calculated at a distance of 500 nm from the prism surface. The field amplitude is extracted from the power measured before the laser couples to the microscope, corresponding to a pulse energy of 250 nJ (laser average power of 250 mW and repetition rate 1 MHz).

Note 6: Finite-difference time-domain (FDTD) simulation

In this section we provide a Finite-difference time-domain (FDTD) simulation of our experimental setup as described in the main text. The electric field propagates inside the prism and incidents the prism's surface above the critical angle. The emerging evanescent wave in vacuum interacts with the free–electron. This simulation visualize why we had to extend the conventional PINEM theory to include the spatiotemporal dependence of the electric field and the electron, thus, fully describing our *extended* interaction.

Figure S6. Finite-difference time-domain simulation of our setup. (the simulation movie is attached as a separate file). The simulation assumes a right-angle glass prism with refractive index 1.513 @ $\lambda = 700$ nm, 15 μm leg size, and base angle of 45°. The laser undergoes total internal reflection inside the prism and generates an evanescent field that interacts with the electron that passes nearby.

References

[1] Barwick, B. et al. Photon-induced near-field electron microscopy. *Nature* **462**, 902-906 (2009).

[2] de Abajo, F. J. G. et al. Multiphoton absorption and emission by interaction of swift electrons with evanescent light fields. *Nano Lett.* **10**, 1859-1863 (2010).

[3] Park, S. T. et al. Photon-induced near-field electron microscopy (PINEM): theoretical and experimental. *New J. Phys.* **12**, 123028 (2010).

[4] Feist, A. et al. Quantum coherent optical phase modulation in an ultrafast transmission electron microscope. *Nature* **521**, 200-203 (2015).

[5] Piazza, L. et al. Simultaneous observation of the quantization and the interference pattern of a plasmonic near-field. *Nat. Comm.* **6**, 6407 (2015).

[6] de Abajo, F. J. G. et al. Electron diffraction by plasmon waves. *Phys. Rev. B* **94**, 041404(R) (2016).

[7] Priebe, K. E. et al. Attosecond electron pulse trains and quantum state reconstruction in ultrafast transmission electron microscopy. *Nat. Photonics* **11**, 793-797 (2017).

[8] Morimoto, Y. & Baum, P. Diffraction and microscopy with attosecond electron pulse trains. *Nat. Phys.* **14**, 252-256 (2018).

[9] Vanacore, G. M. et al. Attosecond coherent control of free-electron wave functions using semi-infinite light fields. *Nat. Comm.* **9**, 2694 (2018).

[10] Park, S. T., & Zewail, A. H. Relativistic effects in photon-induced near field electron microscopy. *J. Phys. Chem. A*, **116**(46), 11128-11133 (2012).

[11] Griffiths, D. *Introduction to Electrodynamics* Ed. **3**, pp. 390 (2005).

[12] Cherenkov, P. A. Visible emission of clean liquids by action of γ radiation. *Dokl. Akad. Nauk SSSR* **2**, 451-454 (1934).

[13] Abramowitz, M. & Stegun, I. A. Handbook of Mathematical Functions (Dover, New York, 1972).