Wave-Mechanical Electron-Optical Modeling of Field-Emission Electron Sources

Faruk Krečinić ⬤* and Ralph Ernstorfer ⬤

Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, Berlin 14195, Germany

Electron source coherence has a very important influence on the imaging capabilities of modern electron microscopes. However, conventional electron source models that are based on geometrical electron optics implicitly assume that the emission from the source surface is fully incoherent, which can complicate the treatment of highly coherent field-emission sources. In an attempt to treat the wave-optical properties of electron sources, models inspired by light optics treatments of (partially) coherent sources, which assume a planar source and free wave propagation, have been developed. In this case the underlying assumptions are problematic, because the source surface of a field emitter can have a radius of curvature on the nanometer scale, and the emitted electrons are accelerated by a strong, inhomogeneous electrostatic field following emission. We introduce a model based on wave-mechanical electron optics that draws on a quantum mechanical description of electron emission and propagation to obtain a physically consistent treatment of the wave-mechanical properties of electron sources. We apply the model to investigate spatial resolution limits in low-energy electron holography and microscopy, where it is shown that aberrations and coherence properties of the electron source are crucial and interrelated. The wave-mechanical electron-optical model can, furthermore, be readily generalized to assess and improve electron source performance in other scenarios and techniques where spatial and temporal coherence, and electron-optical aberrations, are relevant.

I. INTRODUCTION

In modern high-resolution electron microscopy the spatial and temporal coherence properties of the electron source play a crucial role in determining the instrument resolution [1–3]. A key concept in the conventional theory of electron sources is the brightness, which is typically assessed by simulating the classical trajectories of electrons emitted by the cathode [4–6]. For extremely small sources that produce highly coherent electron beams, it was shown that the concept of brightness, which is based on geometrical optics principles, can become meaningless and a wave-optical treatment of the source becomes necessary [7]. Even for sources that are not extremely small, the spatial coherence of the electron wave plays an important role that is not captured by conventional electron source theory.

Low-energy electron holography (LEEH) [8–11] is a particularly illustrative example where the consistent treatment of the wave-optical properties of the electron source can contribute to an improved understanding of the spatial resolution limits affecting the technique. In LEEH a sharp metallic tip is used as a point source of field-emitted electrons to project a magnified image of a sample, without further electron-optical lens elements. At high magnifications, the projected image is a Fresnel diffraction pattern of the sample that can be inverted to form a real-space image of the object, which is equivalent to the in-line holography originally proposed by Gabor [12]. Some initial publications investigated the possibility of atomic resolution imaging with LEEH [13,14], but the experimental observations claiming to have achieved this turned out to be controversial [15,16]. Using an ultraclean graphene membrane as a transparent, grounded sample support, Longchamp *et al.* [17] recently imaged individual proteins at an estimated resolution of 7 to 8 Å, which is still significantly larger than required for atomic resolution imaging. The lack of atomic resolution was attributed to the disturbance caused by external mechanical vibrations.

At the same time, it stands to reason that the spatial and temporal coherence of the electron beam has an important influence on the spatial resolution in LEEH. Wave-optical models based on light optics theory, assuming free-space propagation and a planar source surface, were developed in order to study these effects [18–20]. The free-wave,

*krecinic@fhi-berlin.mpg.de

planar-source models indicated that an increase in the size of a fully coherent emitter leads to a reduction of the angular diameter of the emitted beam and (due to the classical Abbe criterion) to a corresponding decrease of the spatial resolution. Latychevskaia [20] concluded, for example, that the spatial resolution is then approximately equal to the size of the coherent source. However, it is unclear how to reconcile this conclusion with the experimental observations of Cho *et al.* [21], who showed that cryogenically cooling the source, and thus increasing the spatial coherence length of the electron wave packet at the source surface, results in a marked increase in the opening angle at which fringes are visible. For these applications, the free-wave, planar-source model clearly has shortcomings: (a) the electron wave is not free, but is accelerated rather strongly after being emitted, which introduces lensing and aberration effects [22]; (b) the field-emission electron source surface can have a radius of curvature on the nanometer scale. Simulations based on geometrical electron optics that can take such effects into account have therefore been employed in order to study the role of the initial electron momentum distribution [23], electron-optical aberrations, and the spatial emission probability distribution [24] on the resolution. However, in this case

it is challenging, especially when dealing with a partially coherent and spatially extended source, to relate the classical electron trajectories in a straightforward and consistent way to the electron wave function, which is required to simulate holographic image formation.

Starting from the basic quantum mechanical principles of electron emission and propagation we introduce a model based on wave-mechanical electron optics that can treat the geometric and wave-optical properties of electron sources in a consistent manner. We apply this model specifically to LEEH, where it is shown to provide physical insight into the interrelated influences of the tip geometry, spatial coherence, and electron-optical aberrations on the spatial resolution. The model should not only be useful in the specific setting of LEEH, but also for other forms of modern electron microscopy that require a more thorough understanding of the coherence and aberration properties of the source.

II. THEORY

We model the point-projection microscope used for LEEH with the concentric spherical electrode model that is illustrated schematically in Fig. 1. The inner spherical
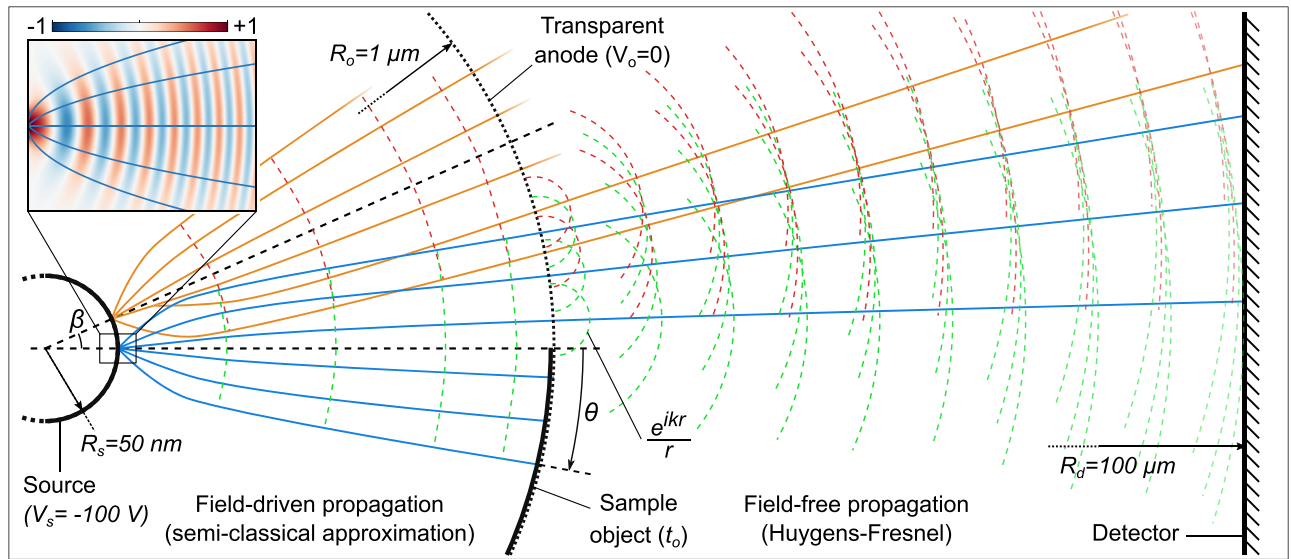

FIG. 1. Schematic illustration of the concentric sphere model of the point-projection microscope, with some typical values for the parameters used in our simulations as indicated in the figure. Electrons are emitted from the spherical cathode with radius R_s, biased at voltage V_s, and are accelerated towards the grounded, transparent anode with radius R_o. The electron wave function at point θ on the anode surface produced by an atomic emitter on the source surface, with angular position β, is obtained using the semiclassical approximation. The classical trajectories (blue and orange solid lines) are normal to the equiphase surfaces (green and red dashed lines) of the corresponding semiclassical wave function. The inset (upper-left corner) shows the classical electron rays (solid blue lines) originating from an atomic emitter and the corresponding real part of the wave function (red-blue heat map) very close to the emitter surface. This illustrates the fact that the wave function of the atomic emitter has a Heisenberg-limited size at the source surface that is given by the transverse momentum distribution of the set of classical trajectories used to construct the semiclassical wave. The sample object is an infinitely sharp edge located on the anode surface, and is defined by the transmission function t_o. Electrons that are transmitted through the transparent anode can propagate field-free to the detector, and the wave function can then be obtained using the Huygens-Fresnel principle [see Eq. (2)].

electrode with radius R_s represents the electron source, i.e., the cathode, and is biased at a potential V_s. The outer spherical electrode, i.e., the anode, with radius R_o is a virtual surface that is transparent to all electrons and is grounded, so that $V_o = 0$. The radial electrostatic field in the region between the two electrodes is given by

$$E_r = \frac{\Delta V}{r^2} \frac{R_s}{1-p}, \qquad (1)$$

where $p = R_s/R_o$ and $\Delta V = V_s - V_o$. The sample object, which is represented by a function modulating the amplitude of the transmitted electron wave function, is located on the anode surface. The detector plane is located 100 μm from the center of the cathode and is also grounded, so that the electrons propagate in a field-free region after passing the anode. The field-free wave propagation between anode and detector is simulated with the well-known classical Huygens-Fresnel principle (see, for example, Ref. [20]). The electron wave function at the detector surface is then given by

$$\Psi_d(\mathbf{r}_d) = \int \Psi_o(\mathbf{r}_o) t_o(\mathbf{r}_o) \frac{\exp(ikr)}{r} d\mathbf{r}_o, \qquad (2)$$

where $\mathbf{r}_d$ is the detector coordinate, $\mathbf{r}_o$ is the object coordinate, $r = |\mathbf{r}_d - \mathbf{r}_o|$, $t_o(\mathbf{r}_o)$ is the object modulation function, and $\Psi_o(\mathbf{r}_o)$ is the electron wave function at the anode surface. We shall simulate an object with an infinitely sharp edge at $\mathbf{r}_o = 0$, so that its transmission function $t_o(\mathbf{r}_o)$ is the Heaviside function. We use the edge as a test sample to demonstrate the effect of changing source conditions, e.g., the coherence length at the source surface, on the simulated projection image, with all other parameters, e.g., the tip-sample distance, remaining fixed. The projection image from an ideal edge can also give us an approximate idea of the resolution based on the Abbe criterion. Ultimately, one would use the wave-mechanical electron-optical model to simulate an experimentally obtained projection image of an object, such as an edge, to evaluate the physical conditions at the source surface.

In the following subsections we discuss the two main aspects of our wave-mechanical electron-optical model. In Sec. II A we consider an atomic-sized emitter on the source surface, which is an approximation for an ideal point source, and show how its wave function at the anode surface can be obtained from a set of classical electron trajectories using the semiclassical approximation. In Sec. II B we show that spatial and temporal coherence are crucial in modeling the electron beam emitted by an extended source surface. With the help of optical coherence theory we first specify the form of the electron wave function at the source surface using its cross-spectral density (CSD). The atomic emitter wave function obtained in Sec. II A is then shown to be an approximation of the

impulse response function that is required to propagate the CSD from the source surface to the anode surface. Finally, we obtain the CSD at the detector and show how an electron wave packet with an energy spread, which gives rise to temporal coherence effects, is to be modeled. We also discuss how to apply our model to a source with varying degrees of spatial coherence, and define the concepts of fully coherent, partially coherent, and fully *in*coherent sources used in the simulations.

Finally, we note that the wave-mechanical electron-optical model we outline here can in principle be readily applied to simulating more realistic electron source and sample geometries, i.e., nanotips and nano-objects, respectively. Nevertheless, the concentric spherical electrode model captures the essential electron-optical characteristics of the problem, e.g., aberrations [25,26], while remaining relatively simple and transparent, which is the goal of the current work.

A. Semiclassical wave function

In the semiclassical approximation one starts from a wave function of the form

$$\Psi(\mathbf{r}) = \mathcal{G}(\mathbf{r}) \exp[i\mathcal{S}(\mathbf{r})], \qquad (3)$$

where $\mathcal{G}(\mathbf{r})$ is the amplitude function and $\mathcal{S}(\mathbf{r})$ is the phase function [27,28]. This wave function is then substituted into the time-independent Schrödinger equation, and the resulting equation is expanded in powers of $\hbar$. The zeroth-order approximation S of the phase function in the classically allowed region is given by the partial differential equation [28]

$$[\nabla S(\mathbf{r})]^2 = 2m[E - V(\mathbf{r})], \qquad (4)$$

where E is the energy and $V(\mathbf{r})$ is the electrostatic scalar potential. Applying the method of characteristics, solutions to Eq. (4) and its corresponding approximate wave function can be obtained from an appropriate set of classical particle trajectories (see the Appendix). The initial conditions, i.e., $\mathbf{r}_i$ and $\mathbf{p}_i$, for the set of classical particles from which the semiclassical wave function is constructed are obtained from boundary conditions specific to the problem. The phase function is then given by

$$S(\mathbf{r}_f) = \int^{t_f(\mathbf{r}_f)} 2K(\tau) d\tau, \qquad (5)$$

where K is the kinetic energy of a classical particle, as a function of time τ, with its trajectory starting at $\mathbf{r}_i$, and ending at $\mathbf{r}_f$, where $\tau = t_f$.

Concretely, to obtain the wave function of a single atomic field emitter in our semiclassical model implementation, we emit a set of classical electrons with starting points on the optical axis, i.e., with $\beta = 0$ (see also

Fig. [1]). Since the electrons are produced by a tunneling process, their longitudinal momentum $p_{i,\parallel}$ is zero as they emerge from the tunneling barrier. However, they can have a range of initial transverse momenta $p_{i,\perp}$. The corresponding radial starting point r_i is obtained from the electrostatic potential and energy conservation such that $E = p_{i,\perp}^2/2m + V(r_i)$. The transverse momentum distribution is assumed to be a Gaussian function

$$G_i(p_{i,\perp}) = \exp\left(-\frac{p_{i,\perp}^2}{2\sigma_{p\perp}^2}\right). \tag{6}$$

The perpendicular momentum spread parameter $\sigma_{p\perp}$ is related to the size of the electron wave function at the cathode by the uncertainty principle. In our semiclassical model, the set of classical electrons originating from the atomic emitter provide a geometric approximation of a Gaussian wave emitter, which is illustrated in the inset of Fig. [1]. Tsujino [29] has recently derived an expression for the wave function of a field-emitted electron using Fowler-Nordheim tunneling theory, which assumes a planar tunneling barrier, and pointed out its connection to an optical Gaussian beam. We note that, although Fowler-Nordheim theory is widely applied to the analysis of field-emission electron sources (see, for example, Refs. [6,24]), the transverse momentum distribution obtained with it implies a Heisenberg-limited spatial extent that is too large to represent our semiclassical atomic emitter. For typical field strengths in the range of 3 to 7 V/nm, the transverse momentum spread $\sigma_{p\perp}$ according to Fowler-Nordheim theory is 2.5×10^{-2} to 3.8×10^{-2} Å^{-1} [30]. By applying the Heisenberg principle, i.e., $\sigma_x\sigma_k \geq \hbar/2$, we find that the *minimum* spatial extent of a wave packet corresponding to the assumed field strengths is 13 to 20 Å. If we consider the limiting case such that the emission originates from a single tungsten atom with a size of about 2 Å, we obtain a $\sigma_{p\perp}$ of 2.5×10^{-1} Å^{-1}, which is approximately an order of magnitude larger than that obtained with the Fowler-Nordheim theory assuming a flat surface. We therefore use $\sigma_{p\perp} = 2.5 \times 10^{-1}$ Å^{-1} for the transverse momentum distribution of the semiclassical atomic emitter. As will be shown in the next section, the atomic emitter wave function gives an approximation of the impulse response function, with which the wave function emitted by an extended surface with an arbitrary geometry can be constructed. It should be noted that our atomic emitter does not necessarily accurately represent a real single-atom emitter, which may have a nontrivial phase and energy distribution that depends on the detailed electronic structure and tunneling mechanics of the emitting atom [31]. We consider the semiclassical atomic emitters in this approximate sense.

We numerically integrate the classical equations of motion and obtain the corresponding semiclassical phase
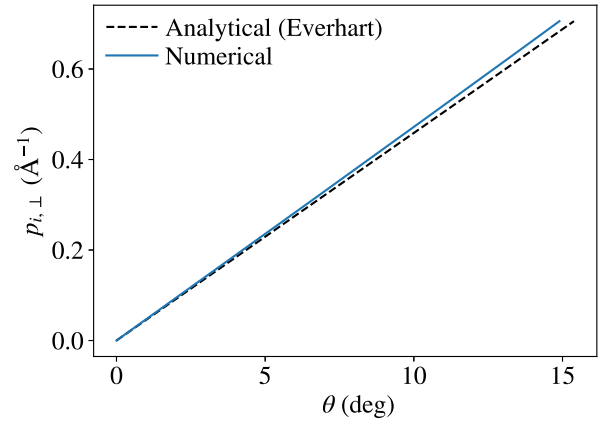

FIG. 2. Plot of $g(\theta)$ from Eq. [7] for the on-axis, i.e., $\beta = 0$, atomic emitter (solid blue line), which maps the anode angular position θ to the initial transverse momentum $p_{i,\perp}$. The dashed black line plots the analytical approximation for the concentric sphere model given by Everhart [32].

function [5] for all electrons originating from the on-axis atomic emitter. The phase values at the intersection points between this set of classical paths and the anode surface are interpolated to obtain the phase function S_o at the anode as a function of the angular coordinate θ (see also Fig. [1]). The semiclassical wave function amplitude is obtained from the transverse momentum distribution [6]. The on-axis atomic emitter wave function at the anode surface can therefore be written as

$$\phi_o(\theta, E) = G_i[g(\theta)] \exp[iS_o(\theta, E)], \tag{7}$$

where g is a function mapping the anode angular position θ to the corresponding initial transverse momentum $p_{i,\perp}$. Figure [2] shows a plot of this function for the on-axis emitter, i.e., with $\beta = 0$. For the concentric spherical electrode model, Everhart [32] has given an analytical approximation of this function, which we obtain numerically, that is plotted in the same figure for comparison.

B. Spatial and temporal coherence

In this section we clarify how the electron wave function produced by an extended source surface is to be composed from the atomic wave functions introduced in the previous section. A crucial consideration here is the spatial and temporal coherence of the emitted electron beam. The emitter is stochastic in nature: the electrons in the metal experience random collisions that are caused by electron-phonon, electron-impurity, and electron-electron interactions. According to the quantum theory of decoherence, these random interactions lead to a localization of the electron within the metal, i.e., to the suppression of spatial coherence [33]. At thermal equilibrium the coherence

length of the electron within the material is given by [34]

$$\ell = \frac{\hbar}{\sqrt{4m_e^* k_B T}}, \qquad (8)$$

where m_e^* is the electron effective mass, k_B is the Boltzmann constant, and T is the temperature. By cryogenically cooling a field emitter, Cho *et al.* [21] showed experimentally that there is a clear connection between the spatial coherence of the electrons within the material, which depends on temperature as indicated in Eq. (8), and the coherence properties of the emitted electron beam. We use optical coherence theory to model the beam after it emerges from the source with a given spatial and temporal coherence, which depends on the material conditions and properties, and propagates to the object and detector.

The semiclassical wave function obtained in the previous section is a solution of the *time-independent* Schrödinger equation. It is therefore natural to start our treatment in the space-frequency (energy) domain. In the space-frequency domain, the coherence, i.e., the statistical degree of correlation between the spectral amplitudes of a particular frequency component at any two points on the source surface, β and β', is given by the CSD $W(\beta, \beta')$ [35]. By making use of a coherent mode decomposition [36], we can write

$$W(\beta, \beta') = \int \rho(\eta) \chi^*(\beta' - \eta) \chi(\beta - \eta) d\eta, \qquad (9)$$

where the variable of integration η is an angular coordinate on the source and the integral is performed over its entire surface. The coherent mode decomposition expresses the CSD as a sum of contributions from spatially fully coherent, elementary sources [36], described by the coherent mode function χ, that are weighed by the probability density function ρ. Physically, it represents here the fact that electrons are emitted as localized wave packets, described by the function χ, at random positions on the tip surface with probability density $\rho(\eta)$. In the current model we assume that the coherent mode function χ is a Gaussian function with a standard deviation σ_χ that is a function of the coherence length of electrons within the material, as well as the field-emission conditions. The spectral degree of coherence for a single mode can be obtained by substituting $\rho(\eta) = \delta(\eta - \eta_0)$ in Eq. (9), which gives

$$\mu^{(\eta_0)}(\beta, \beta') = \frac{W^{(\eta_0)}(\beta, \beta')}{[W^{(\eta_0)}(\beta, \beta)]^{1/2}[W^{(\eta_0)}(\beta', \beta')]^{1/2}}$$
$$= \frac{\chi^*(\beta' - \eta_0)\chi(\beta - \eta_0)}{|\chi(\beta' - \eta_0)||\chi(\beta - \eta_0)|}. \qquad (10)$$

Since $|\mu^{(\eta_0)}(\beta, \beta')| = 1$ for any values of β and β', this demonstrates that an individual mode is indeed spatially

fully coherent, which is consistent with the interpretation that it represents a localized electron wave packet emitted at point η_0.

We require an expression of the CSD at the object and the detector surfaces. A general expression for the CSD upon propagation (see also Sec. 4.4.3 of Ref. [35]) is given by

$$W(\theta, \theta') = \iint W(\beta, \beta') K^*(\theta', \beta') K(\theta, \beta) d\beta d\beta', \qquad (11)$$

where $K(\theta, \beta)$ is the impulse response function, i.e., the function specifying the wave function at point θ on the target surface produced by a point source located on the source surface at point β. In our model we approximate the source-object impulse response by the semiclassical atomic emitter wave function obtained in Sec. II A [see also Eq. (7)], from which we can obtain the wave function at the object surface produced by an atomic emitter located at an arbitrary point β on the source surface (see also Fig. 1). We note here that approximating the impulse response function in this way implies that the effective size of the electron wave function at the source surface cannot be smaller than the Heisenberg-limited size of the semiclassical atomic emitter, which is given by its transverse momentum spread $\sigma_{p\perp}$. Substituting Eq. (7) for K, where we have made use of the rotational symmetry of the wave function in the concentric spherical model, and Eq. (9) for W we obtain for the CSD at the object surface

$$W(\theta, \theta') = \int \rho(\eta) \int \chi^*(\beta' - \eta) \phi_o^*(\theta' + \beta', E) d\beta'$$
$$\times \int \chi(\beta - \eta) \phi_o(\theta + \beta, E) d\beta d\eta. \qquad (12)$$

We observe that the integrals over β and β' are a convolution of the atomic emitter wave function ϕ_o with the coherent mode function χ. We therefore rewrite Eq. (12) as

$$W(\theta, \theta') = \int \rho(\eta) \Psi_o^{\chi*}(\theta') \Psi_o^\chi(\theta) d\eta, \qquad (13)$$

where

$$\Psi_o^\chi(\theta, \eta, E) = \int \chi(\beta - \eta) \phi_o(\theta + \beta, E) d\beta, \qquad (14)$$

which is the wave function at the object plane due to a coherent superposition of the atomic emitter wave functions within the coherent mode χ centered at η. To propagate the CSD from the object to the detector plane, we again make use of Eq. (11). The object-detector impulse response is characterized by the object transmission function, which modulates the wave impinging on the object

surface, followed by free wave propagation, and is therefore equal to the factor $t_o(\mathbf{r}_o)\exp(ikr)/r$ in the Huygens-Fresnel integral in Eq. (2). Using Eqs. (13) and (11), for the CSD at the detector, we obtain

$$W(\mathbf{r}_d, \mathbf{r}'_d) = \int \rho(\eta) \int \Psi_o^{\chi*}(\theta', \eta, E) t_o^*(\mathbf{r}'_o) \frac{\exp(-ikr')}{r'} d\mathbf{r}'_o$$
$$\times \int \Psi_o^{\chi}(\theta, \eta, E) t_o(\mathbf{r}_o) \frac{\exp(ikr)}{r} d\mathbf{r}_o d\eta, \tag{15}$$

where $r^{(\prime)} = |\mathbf{r}_d^{(\prime)} - \mathbf{r}_o^{(\prime)}|$ and $\mathbf{r}_o^{(\prime)} = R_o[\cos\theta^{(\prime)}, \sin\theta^{(\prime)}]$. An electron wave packet may consist of multiple energy components, which leads to the presence of temporal coherence effects. In order to take the effect of temporal coherence at the detector into account, we obtain the mutual coherence function (MCF) Γ from the CSD given by Eq. (15). The MCF gives the statistical degree of correlation between the wave function at the two points $\mathbf{r}_d$ and $\mathbf{r}'_d$, and at times t and t', respectively. For a stationary (and ergodic) random process, the MCF can be obtained from the CSD by taking the inverse Fourier transform

$$\Gamma(\mathbf{r}_d, \mathbf{r}'_d, \tau) = \int p(E) W(\mathbf{r}_d, \mathbf{r}'_d) \exp(iE\tau) dE, \tag{16}$$

where p is the spectral density function of the wave packet, i.e., energy spectrum, and $\tau = t' - t$ (see Sec. 4.3 of Ref. [35]). We can obtain the time-averaged intensity at the detector from $I_d \equiv \Gamma(\mathbf{r}_d, \mathbf{r}_d, 0)$, so that after substituting Eq. (15) into Eq. (16) we get

$$I_d = \int p(E) \int \rho(\eta) \left| \Psi_d^{\chi}(\mathbf{r}_d, \eta, E) \right|^2 d\eta dE, \tag{17}$$

where

$$\Psi_d^{\chi}(\mathbf{r}_d, \eta, E) = \int \Psi_o^{\chi}(\theta, \eta, E) t_o(\mathbf{r}_o) \frac{\exp(ikr)}{r} d\mathbf{r}_o. \tag{18}$$

Finally, for a monoenergetic electron beam of energy E_0, we have $p(E) = \delta(E - E_0)$, so that Eq. (17) becomes

$$I_d^{E_0} = \int \rho(\eta) |\Psi_d^{\chi}(\mathbf{r}_d, \eta, E_0)|^2 d\eta. \tag{19}$$

How does one apply the theory above to model a source with varying degrees of spatial coherence? We consider a spatially fully *incoherent* source to be such that the coherent mode function reduces to a delta function, i.e.,

$\chi(\beta - \eta) = \delta(\beta - \eta)$, so that from Eq. (14) we obtain

$$\Psi_o^{\chi}(\theta, \eta, E) = \phi_o(\theta + \eta, E). \tag{20}$$

For a monoenergetic electron beam we get, using Eqs. (18) and (19) for the intensity at the detector,

$$I_d^{E_0, i} = \int \rho(\eta) \left| \int \phi_o(\theta + \eta, E_0) t_o(\mathbf{r}_o) \frac{\exp(ikr)}{r} d\mathbf{r}_o \right|^2 d\eta, \tag{21}$$

which shows that the projection image consists of an incoherent summation of the images produced by the individual atomic emitters that make up the surface. A spatially fully coherent source can be considered such that it emits only a single coherent mode, so that $\rho(\eta) = \delta(\eta - \eta_0)$ [see also the discussion following Eqs. (9) and (10)]. When both ρ and χ are given by delta functions, the electron beam originates from a single atomic emitter on the source surface and is, by definition, fully coherent. A partially coherent source is therefore such that both χ and ρ have a finite width, i.e., neither χ nor ρ contain delta functions.

Before proceeding to the results we discuss how the theory outlined above relates to the van Cittert–Zernike theorem, which is frequently used to estimate the electron source size in experiments. The van Cittert–Zernike theorem describes the coherence properties of fields generated by planar, (quasi)monochromatic, spatially fully *in*coherent sources. Moreover, it is assumed that the wave propagates freely at all times. Clearly, the field emitter, with its curved source surface and the strong acceleration of the electrons after they are emitted, is poorly represented by the van Cittert–Zernike theorem. The effective source size that is obtained from the application of the van Cittert–Zernike theorem therefore has no direct bearing on the physical conditions at the source surface, e.g., coherence length and geometry, which one would be interested in when modeling these sources. By contrast, the model outlined above does provides a way to relate, through the use of simulations, the measured coherence properties of the electron beam to the physical conditions and parameters of the source itself.

III. RESULTS

We first simulate the point-projection images produced by a monoenergetic, spatially fully coherent source of increasing spatial extent. The spatial extent of the wave function at the source surface σ_s is a summation in quadrature of the Heisenberg-limited size of the semiclassical atomic emitter of 2 Å and the σ_χ of the Gaussian coherent mode distribution. We perform simulations with a σ_s of 2, 13, 26, and 80 Å. A σ_s of 13 and 26 Å corresponds approximately to the coherence length of electrons obtained from Eq. (8) for tungsten at room temperature and at liquid nitrogen temperature (about 77 K), respectively. The $\sigma_s = 2$ and
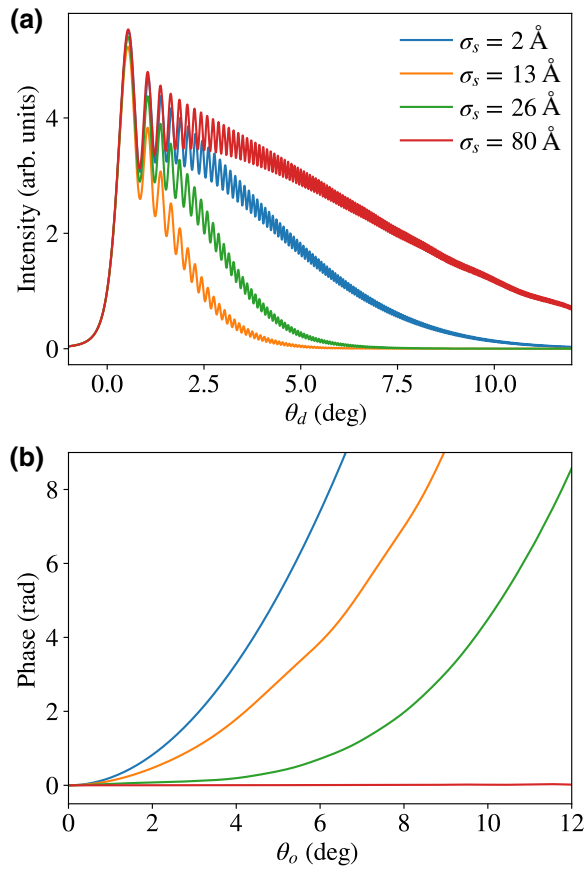
FIG. 3. Simulations of (a) the projection image on the detector and (b) the electron wave function phase at the anode for the concentric spherical electrode model (see Fig. 1: $R_s = 50$ nm, $R_o = 1$ μm, $V_s = -100$ V) with emission from a single monoenergetic, spatially fully coherent Gaussian mode with an extent of $\sigma_s = 2$ to 80 Å at the source surface. The detector plane is located 100 μm away from the center of the cathode.

80 Å simulations represent limiting cases where the emission originates from an atomic-sized area, or from a surface area that is significantly larger than the typical coherence length of electrons in tungsten, respectively. The image at the detector is obtained from Eq. (19) with $\rho(\eta) = \delta(\eta)$. The cathode and anode radii R_s and R_o are 50 nm and 1 μm, respectively, and the cathode voltage V_s is -100 V. The detector plane is located 100 μm away from the center of the cathode. Figure 3(a) shows that the angular extent of the emitted beam reduces significantly as σ_s increases from 2 to 13 Å. This is consistent with the results from free-wave, planar-source models, which show that an increase in the coherent source size leads to a proportional decrease in the beam divergence [18–20]. However, as the source extent σ_s increases from 13 to 26 and 80 Å, the angular extent of the beam increases as well, which is a significant departure from the monotonic decrease in beam divergence that is predicted by the free-wave, planar-source models. This can be explained by the fact that the emitted wave

front is affected by the surface curvature of the cathode that tends to compensate for the emission narrowing caused by interference effects. The effect of the source surface curvature on the wave function can also be seen in Fig. 3(b), which shows the phase of the wave function at the anode surface. The $\sigma_s = 2$ Å emitter shows the strongest deviation from a perfect spherical wave that would have a phase shift of zero across the spherical anode surface. As the coherent emitter size increases to a size comparable to the tip radius, e.g., $\sigma_s = 80$ Å, the emitted wave becomes more spherical, with a more flat phase profile at the anode and also a wider intensity distribution at the detector.

Next, we simulate the projection images produced by a monoenergetic, *partially* coherent source. The emission from such an emitter typically originates from an extended surface area that is significantly larger than the electron coherence length at the source surface. Since in the concentric sphere model the field strength is equal everywhere on the surface of the cathode, we take the emission probability density $\rho(\eta)$ to be constant here. The image at the detector is therefore given by Eq. (19) with $\rho(\eta) = 1$. In realistic sources the emission probability distribution can be obtained from the apex field distribution and the tunneling current density. This is typically a Gaussian-like distribution with an opening angle of greater than 45° [10], so that the constant emission probability assumed in the current simulation is a reasonable approximation as we are mainly concerned with the image at relatively small angles. The simulated projection images are shown in Fig. 4, with the coherence length σ_s corresponding to the fully coherent source simulations shown in Fig. 3. The coherence length σ_s is defined, analogously to the fully coherent source simulations, as the summation in quadrature of σ_χ and the atomic emitter size of 2 Å. A striking effect on the visibility of the projected
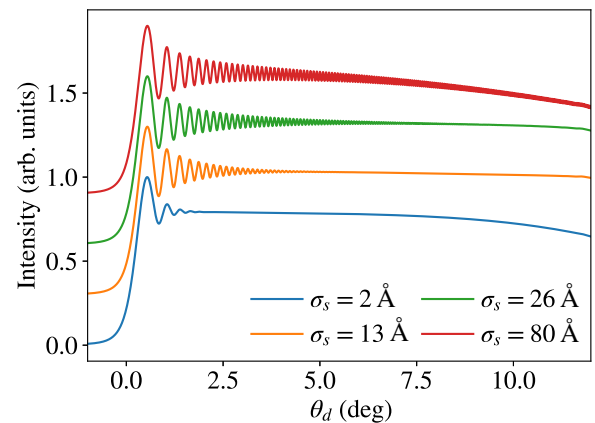

FIG. 4. Simulation of the projection image from a *partially* coherent emitter with a coherence length of the emitted electron at the source surface of $\sigma_s = 2$ to 80 Å (with all other parameters equal to those of Fig. 3).

interference fringes can be observed for a partially coherent emitter with the smallest coherence length of 2 Å. Such a small coherence length implies that the emission from an atomic-sized emitter is fully incoherent with its neighbors. The incoherent superposition of emission from the individual atomic emitters leads to a considerable reduction in the visibility of higher-order fringes to opening angles of $\theta \leq 2°$. According to the Abbe criterion this would correspond to an estimated spatial resolution of ≥ 17 Å. The strong decrease in fringe visibility is actually caused by the electron-optical aberration of the wave function. The aberrations induce phase shifts, as shown in Fig. 3(b), that lead to a shift in the position of the interference fringes on the detector, which become more pronounced as the angle relative to the emission point increases. The net result of this fringe shift, after integration over the incoherent emission from all the atomic emitters on the surface, is a strong suppression of the higher-order fringes. From Fig. 3(b) we deduce that the electron-optical aberration effects should be smaller as the spatial coherence length, i.e., the spatial extent of the coherent mode function χ, increases. Figure 4 indeed shows that the visibility of higher-order interference fringes increases for increasing σ_s, despite the fact that the angular extent of the fully coherent beams for $\sigma_s = 13$ and 26 Å is actually smaller than that for $\sigma_s = 2$ Å [see Fig. 3(a)]. These results are consistent with the experiments by Cho *et al.* [21] and provide a good explanation for the observed increase in the interference fringe visibility as the source temperature is reduced and its spatial coherence length increases (see also the discussion at the beginning of Sec. II B).

The preceding discussion implies that the ratio between the coherence length and the emission surface curvature should have an influence on the properties of the emitted beam. In Fig. 5 we show simulations of a spatially fully coherent source where $R_s = 20$ nm and all other parameters are the same as those in Fig. 3. Though the angular extent of the beam for $\sigma_s = 2$ Å is quite similar to the corresponding simulation in Fig. 3(a), the beams for $\sigma_s = 13$ to 80 Å are significantly wider. The increased surface curvature is essentially transferred to the emitted wave front. The source with $\sigma_s = 26$ Å in fact already has a larger angular extent than the atomic emitter. Furthermore, Fig. 5(b) shows that the phase of the wave function at the object is significantly more spherical for all of the coherent emitter sizes. Figure 6 shows the simulation of the projection images from a partially coherent source with $R_s = 20$ nm. For the partially coherent emitter, the increase in the phase sphericity evident from Fig. 5(b) leads to a reduction in aberration effects that tend to wash out the interference fringes. The angular extent at which fringes can be observed, and the corresponding Abbe-limited resolution, has consequently become higher even for the lowest coherence length of 2 Å. It must be mentioned here that the strong decrease in the aberration for the smaller cathode
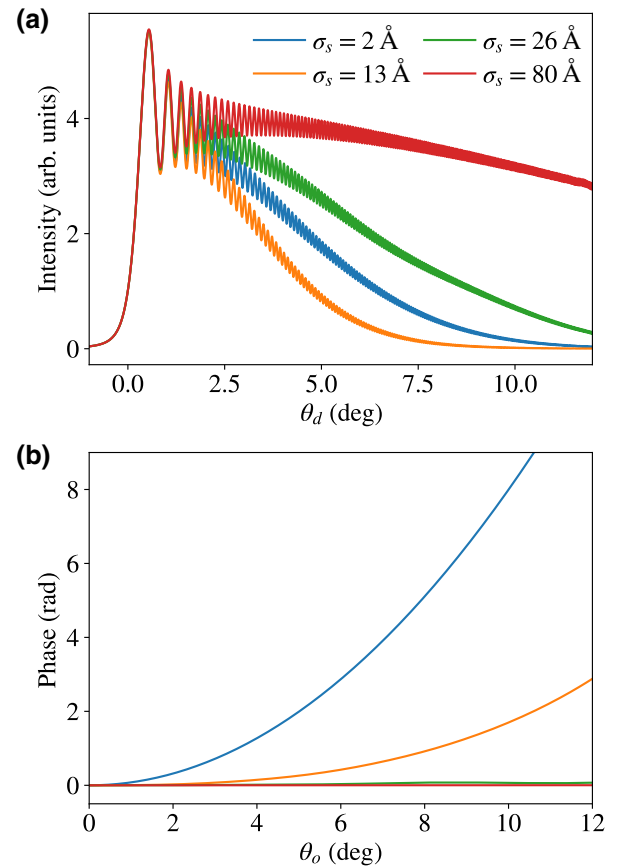

FIG. 5. Simulations of (a) the projection image on the detector and (b) the electron wave function phase at the anode for a monoenergetic, spatially fully coherent emitter with a cathode radius $R_s = 20$ nm, and with all other parameters equal to those of Fig. 3.

radius is partially due to the concentric sphere geometry, which is known to underestimate aberration effects [5]. In a more realistic tip-sample geometry, which includes the
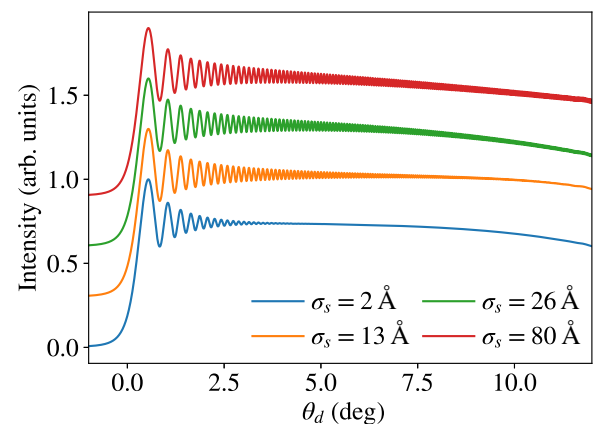

FIG. 6. Simulation of a partially coherent emitter with a cathode radius $R_s = 20$ nm, and with all other parameters equal to those of Fig. 4.
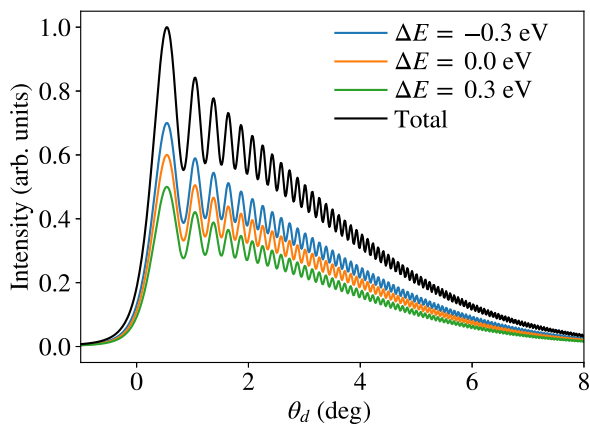
FIG. 7. Simulated projection image (black line) produced by an atomic-sized emitter ($\sigma_s = 2$ Å) with a Gaussian energy spread of $\sigma_E = 400$ meV and otherwise the same parameters as used in the simulations shown in Fig. 3. Monoenergetic simulations (colored lines) for a given energy difference ΔE from the nominal center energy $E_0 = 100$ eV.

influence of the tip shaft, the electrostatic field distribution is much more inhomogeneous. This leads to additional aberrations of the wave front, which can be expected to have an adverse effect on the interference fringe visibility for a partially coherent emitter. The current model also does not include any surface roughness or faceting effects that should have a more pronounced influence as the size of the emission surface is reduced.

Finally, in Fig. 7 we show the effect of an energy spread, i.e., temporal coherence, on the point-projection image. We simulate a source that consists of a single atomic emitter, with a Gaussian energy spectrum with $\sigma_E = 400$ meV, which is a typical value for the energy spread of a field emitter, and with otherwise the same parameters as used in the simulations shown in Fig. 3. The projection image is obtained using Eq. (17), where $\rho(\eta) = \delta(\eta)$ and $\chi(\beta) = \delta(\beta)$. Somewhat surprisingly, the influence of the total energy spread is rather small and does not impose a limit on the LEEH imaging resolution in our simulations, since there is no noticeable effect on the visibility of the interference fringes. This is explained by the fact that fringe positions are affected very little by changes in the electron energy, as illustrated in Fig. 7 by the monoenergetic simulations that are plotted with colored lines.

IV. DISCUSSION

In this section we discuss the most important implications of the results obtained with our wave-mechanical electron-optical model on optimizing the achievable spatial resolution in LEEH. Because of their extremely coherent nature, monoatomic field emitters have been regarded as the ideal realization of a point source for the use

in LEEH. However, as was pointed out by Scheinfein *et al.* [24], and later by Stevens [23], the resolution of a monoatomic emitter is mainly limited by the initial transverse momentum and the corresponding achievable beam opening angle. From the typically observed beam opening angle values of 4°–6° [10], the Abbe criterion implies that the achievable resolution of these sources is 6 to 9 Å for a beam energy of 100 eV. Holenstein *et al.* [37] have shown that, in principle, it should be possible to go beyond this limit by constructing a composite hologram where the sample is moved with respect to the emitter. From our simulations we can furthermore conclude that the composite hologram must be constructed such that phase shifts caused by electron-optical aberrations are accounted for, which is interesting because LEEH is typically considered to be an "aberration-free" technique [9,11]. This can be achieved by characterizing the wave front aberrations with a reference sample and correcting for them during the holographic image reconstruction. At a 100 eV beam energy it should in principle be possible to extend the effective beam opening angle to 8°–12°, enabling a spatial resolution down to about 4 Å.

If we consider emission from an extended surface area, our simulations indicate that, aside from the transverse momentum distribution and the spatial coherence length, the emission surface curvature plays a crucial role in determining spatial resolution. In contrast to the free-wave, planar-source models, which imply that the beam divergence must become smaller and thus that the attainable spatial resolution must decrease as the coherent source size increases, our simulations show that the coherent extent of the emitted beam can in fact broaden. By making sure that the coherence length is comparable to the curvature of the emission surface, it may even be possible to extend the coherent opening angle of the emission to values beyond that of an atomic emitter. Experimentally, this would likely require a tip with a very well-defined shape and high-quality apex surface, in addition to cryogenic cooling in order to ensure a large coherence length within the emitting material. An additional intriguing possibility for ensuring an adequate electron coherence length within the source material is through the use of materials with an intrinsically large coherence length. For example, semimetals such as bismuth, antimony, and graphene are characterized by a very small effective electron mass, which, according to Eq. (8), leads to a large spatial coherence length. Recently, Shao *et al.* [38] have produced a graphene coated nanotip that would, for example, be well suited to explore this possibility in that material.

Lastly, we would like to point out the use of photoemission as a possibility for increasing the attainable spatial resolution. Nanotip photocathode sources with coherence and resolution properties comparable to field emitters have been demonstrated in low-energy electron holography [39]. However, using photoemission, it may be possible

to go beyond some of the limitations of monoatomic sources. As we have seen, for field emission, the maximum attainable transverse momentum distribution for atomic-sized emitters is ultimately given by the Heisenberg limit. Increasing the transverse momentum distribution would therefore require a reduction in the size of the coherent emitter. However, for photoemission, the excess energy from a high-energy photon emitting an electron into the vacuum would directly translate to an increase in the transverse momentum spread. In this scenario, two points would be crucial. The first is that the photoemission must be constrained to an as small, preferably atomic-sized, surface as possible. As illustrated in Fig. 4, incoherent photoemission from an extended surface area would inevitably limit the spatial resolution due to aberration effects. The second point is that the photoemission would need to be constrained such that the energy spread does not increase with the excess photon energy. In a metal the presence of an electron sea below the Fermi energy results in a very large increase in energy spread as the photon energy increases beyond the work function. Ideally, the emission would therefore proceed through a two-step process, where first a well-defined intermediate state is populated resonantly, followed by photoemission from the intermediate state by a high-energy photon with a significant amount of excess energy with respect to the vacuum level. This setup could for example be realized with a semiconducting carbon nanotube, where the atomic dimensions of the tube diameter ensure localization of emission. A two-photon photoemission scheme with initial excitation to the bottom of the conduction band, followed by photoemission, would ensure that the excess photon energy does not result in a corresponding increase in the energy spread.

V. CONCLUSION

We introduce a wave-mechanical electron-optical model that can consistently simulate the wave and particle-optical properties of electron sources and provide physical insight into the relation between coherence, aberrations, and the geometry of the emitter. The simulation of field-emission electron sources in LEEH allows us to resolve some inconsistencies between experimental observations and the predictions of the free-wave, planar-source models that have been used so far to perform similar simulations. For example, we show that the predicted narrowing of the emitted beam from a spatially extended coherent source in the free-wave, planar-source models is counteracted in our model by the curvature of the emission surface. For partially coherent emitters, we show how the combination of limited spatial coherence at the source surface and aberrations due to electron propagation effects can lead to a decrease in the effective spatial coherence length of the resulting beam. In the case of LEEH this has lead to the identification of possible strategies for increasing the spatial resolution of the

technique. Increasing the electron coherence at the source, in combination with engineering of the emission surface curvature to values comparable to the coherence length, may lead to a source with a diffraction limit higher than even the monoatomic source. For monoatomic sources, it should be possible to go beyond the limit set by the finite beam opening angle by constructing a composite hologram, provided that image aberrations are addressed during the reconstruction step. The wave-mechanical electron-optical model can be readily developed further to simulate more realistic electron source structures. The results of such models should be useful in evaluating electron sources for more general use in (low-energy) electron microscopy.

APPENDIX: PHASE FUNCTION SOLUTION BY THE METHOD OF CHARACTERISTICS

The zeroth-order approximation S to the phase function $\mathcal{S}$ is given by Eq. (4), which is also known as the eikonal equation. This partial differential equation can be solved using the method of characteristics [40]. We first rewrite Eq. (4) in Cartesian coordinates as

$$\frac{p^2}{2m} + \frac{q^2}{2m} - [E - V(x,y)] = 0, \qquad (A1)$$

where $p = \partial S/\partial x$ and $q = \partial S/\partial y$. The characteristic system of differential equations that corresponds to Eq. (4) is

$$\frac{dx}{d\tau} = \frac{p}{m}, \qquad (A2a)$$

$$\frac{dy}{d\tau} = \frac{q}{m}, \qquad (A2b)$$

$$\frac{dS}{d\tau} = \frac{1}{m}(p^2 + q^2), \qquad (A2c)$$

$$\frac{dp}{d\tau} = \frac{\partial V}{\partial x}, \qquad (A2d)$$

$$\frac{dq}{d\tau} = \frac{\partial V}{\partial y}, \qquad (A2e)$$

where τ is a parametric coordinate describing the characteristic curves. Comparing this system of equations to the system of equations describing classical particle motion, we see from Eqs. (A2a) and (A2b) that p and q are the momentum vector components, and that Eqs. (A2d) and (A2e) can be identified as Newton's second law with

the force on the particle being equal to the electric field vector given by $(\partial V/\partial x, \partial V/\partial y)$. To obtain a particular solution of the eikonal equation, the system of differential equations (A2) must be supplemented by initial values, which, moreover, have to conform to Eq. (A1). For example, the initial conditions describing a semiclassical point source specify that the characteristics start at $\tau = 0$ from a single point (x_0, y_0) with $S(\tau = 0) = 0$. From Eq. (A2) we obtain the additional constraint

$$\frac{p_0^2}{2m} + \frac{q_0^2}{2m} = [E - V(x_0, y_0)], \qquad (A3)$$

where $p_0 = p(\tau = 0)$ and $q_0 = q(\tau = 0)$ are the momentum components of the particle. In other words, the magnitude of the initial momentum is given by the kinetic energy $K_0 = E - V(x_0, y_0)$, while its direction can be specified freely. The full solution of the phase function partial differential equation consists of a union of all characteristic curves obtained from a set of initial values specifying the initial position *and* momentum vectors. The phase is obtained by integrating Eq. (A2c) with respect to τ,

$$S = \int \frac{1}{m}[p(\tau)^2 + q(\tau)^2]d\tau, \qquad (A4)$$

and substituting $K \equiv p^2/2m + q^2/2m$ results in Eq. (5). The path integral form for the above phase function is obtained by a substitution of variables. Using Eqs. (A2a) and (A2b), the differential path length element $d\xi = \sqrt{dx^2 + dy^2}$ becomes

$$d\xi = \frac{1}{m}\sqrt{p^2 + q^2}d\tau. \qquad (A5)$$

Substituting this into Eq. (A2c), and using Eq. (A1), gives

$$S = \int \sqrt{2m[E - V(x, y)]}d\xi. \qquad (A6)$$

[1] L. Reimer and H. Kohl, *Transmission Electron Microscopy Physics of Image Formation*, Springer Series in Optical Sciences (Springer New York, New York, NY, 2008), 5th ed, Vol. 36.

[2] J. C. H. Spence, *High-Resolution Electron Microscopy* (Oxford University Press, Oxford, 2013).

[3] P. Kruit, in *Transmission Electron Microscopy*, edited by C. B. Carter and D. B. Williams (Springer International Publishing, Cham, 2016), p. 1.

[4] J. Worster, The brightness of electron beams, J. Phys. D: Appl. Phys. **2**, 321 (1969).

[5] J. C. Wiesner and T. E. Everhart, Point-cathode electron sources-electron optics of the initial diode region, J. Appl. Phys. **44**, 2140 (1973).

[6] M. S. Bronsgeest, J. E. Barth, L. W. Swanson, and P. Kruit, Probe current, probe size, and the practical brightness for probe forming systems, J. Vac. Sci. Technol. B: Microelectron. Nanometer Struct. **26**, 949 (2008).

[7] P. Kruit, M. Bezuijen, and J. E. Barth, Source brightness and useful beam current of carbon nanotubes and other very small emitters, J. Appl. Phys. **99**, 024315 (2006).

[8] H.-W. Fink, Mono-atomic tips for scanning tunneling microscopy, IBM J. Res. Dev. **30**, 460 (1986).

[9] H. W. Fink, W. Stocker, and H. Schmid, Holography with Low-Energy Electrons, Phys. Rev. Lett. **65**, 1204 (1990).

[10] V. T. Binh, N. Garcia, and S. Purcell, in *Advances in Imaging and Electron Physics* (Academic Press, San Diego, 1996), Vol. 95, p. 63.

[11] J. Spence, W. Qian, and A. Melmed, Experimental low-voltage point-projection microscopy and its possibilities, Ultramicroscopy **52**, 473 (1993).

[12] D. Gabor, Microscopy by reconstructed wave-fronts, Proc. R. Soc. A: Math. Phys. Eng. Sci. **197**, 454 (1949),

[13] H. J. Kreuzer, K. Nakamura, A. Wierzbicki, H. W. Fink, and H. Schmid, Theory of the point source electron microscope, Ultramicroscopy **45**, 381 (1992).

[14] H.-W. Fink, H. Schmid, H. J. Kreuzer, and A. Wierzbicki, Atomic Resolution in Lensless Low-Energy Electron Holography, Phys. Rev. Lett. **67**, 1543 (1991).

[15] J. C. H. Spence, J. M. Zuo, and W. Qian, Comment on "Atomic Resolution in Lensless Low-Energy Electron Holography", Phys. Rev. Lett. **68**, 3256 (1992).

[16] G. M. Shedd, Electron interference effects in electron projection microscopy, J. Vac. Sci. Technol. A: Vac. Surface Film **12**, 2595 (1994).

[17] J.-N. Longchamp, S. Rauschenbach, S. Abb, C. Escher, T. Latychevskaia, K. Kern, and H.-W. Fink, Imaging proteins at the single-molecule level, Proc. Natl. Acad. Sci. **114**, 1474 (2017), ArXiv:1512.08958

[18] P. A. Serena, L. Escapa, J. J. Sáenz, N. García, and H. Rohrer, Coherent electron emission from point sources, J. Microsc. **152**, 43 (1988).

[19] N. Garcia and H. Rohrer, Coherent electron beams and sources, J. Phys.: Condens. Matter **1**, 3737 (1989).

[20] T. Latychevskaia, Spatial coherence of electron beams from field emitters and its effect on the resolution of imaged objects, Ultramicroscopy **175**, 121 (2017).

[21] B. Cho, T. Ichimura, R. Shimizu, and C. Oshima, Quantitative Evaluation of Spatial Coherence of the Electron Beam from Low Temperature Field Emitters, Phys. Rev. Lett. **92**, 246103 (2004).

[22] H. Liebl, *Applied Charged Particle Optics* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008).

[23] G. Stevens, Resolving power of lens-less low energy electron point source microscopy, J. Microsc. **235**, 9 (2009).

[24] M. R. Scheinfein, W. Qian, and J. C. H. Spence, Aberrations of emission cathodes: Nanometer diameter field-emission electron sources, J. Appl. Phys. **73**, 2057 (1993).

[25] E. Ruska, Zur fokussierbarkeit von kathodenstrahlbündeln großer ausgangsquerschnitte, Z. Phys. **83**, 684 (1933).

[26] H. Liebl, The image aberration caused by the acceleration field between concentric spherical electrodes, Optik **83**, 129 (1989).

[27] P. W. Hawkes and E. Kasper, *Principles of Electron Optics—Wave Optics* (Academic Press, San Diego, 1996), Vol. 3.

[28] Z. H. Huang, T. E. Feuchtwang, P. H. Cutler, and E. Kazes, Wentzel-Kramers-Brillouin method in multidimensional tunneling, Phys. Rev. A **41,** 32 (1990).

[29] S. Tsujino, Transverse structure of the wave function of field emission electron beam determined by intrinsic transverse energy, J. Appl. Phys. **124,** 044304 (2018).

[30] R. D. Young, Theoretical total-energy distribution of field-emitted electrons, Phys. Rev. **113,** 110 (1959).

[31] V. T. Binh, S. T. Purcell, N. Garcia, and J. Doglioni, Field-Emission Electron Spectroscopy Of Single-Atom Tips, Phys. Rev. Lett. **69,** 2527 (1992).

[32] T. E. Everhart, Simplified analysis of point-cathode electron sources, J. Appl. Phys. **38,** 4944 (1967).

[33] E. Joos, in *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer, Berlin, 2003), 2nd ed., p. 41.

[34] D. Querlioz, J. Saint-Martin, A. Bournel, and P. Dollfus, Wigner Monte Carlo simulation of phonon-induced electron decoherence in semiconductor nanodevices, Phys. Rev. B **78,** 165306 (2008).

[35] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, 1995), 1st ed.

[36] E. Wolf, New theory of partial coherence in the space–frequency domain Part I: Spectra and cross spectra of steady-state sources, J. Opt. Soc. Am. **72,** 343 (1982).

[37] R. Holenstein, T. A. Rothwell, and M. R. A. Shegelski, Reconstruction of composite in-line electron holograms using a small emission cone, Ultramicroscopy **94,** 99 (2003).

[38] X. Shao, A. Srinivasan, W. K. Ang, and A. Khursheed, A high-brightness large-diameter graphene coated point cathode field emission electron source, Nat. Commun. **9,** 1288 (2018).

[39] D. Ehberger, J. Hammer, M. Eisele, M. Krüger, J. Noe, A. Högele, and P. Hommelhoff, Highly Coherent Electron Beam from a Laser-Triggered Tungsten Needle Tip, Phys. Rev. Lett. **114,** 227601 (2015).

[40] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley-VCH, Weinheim, 1989), Vol. 2.