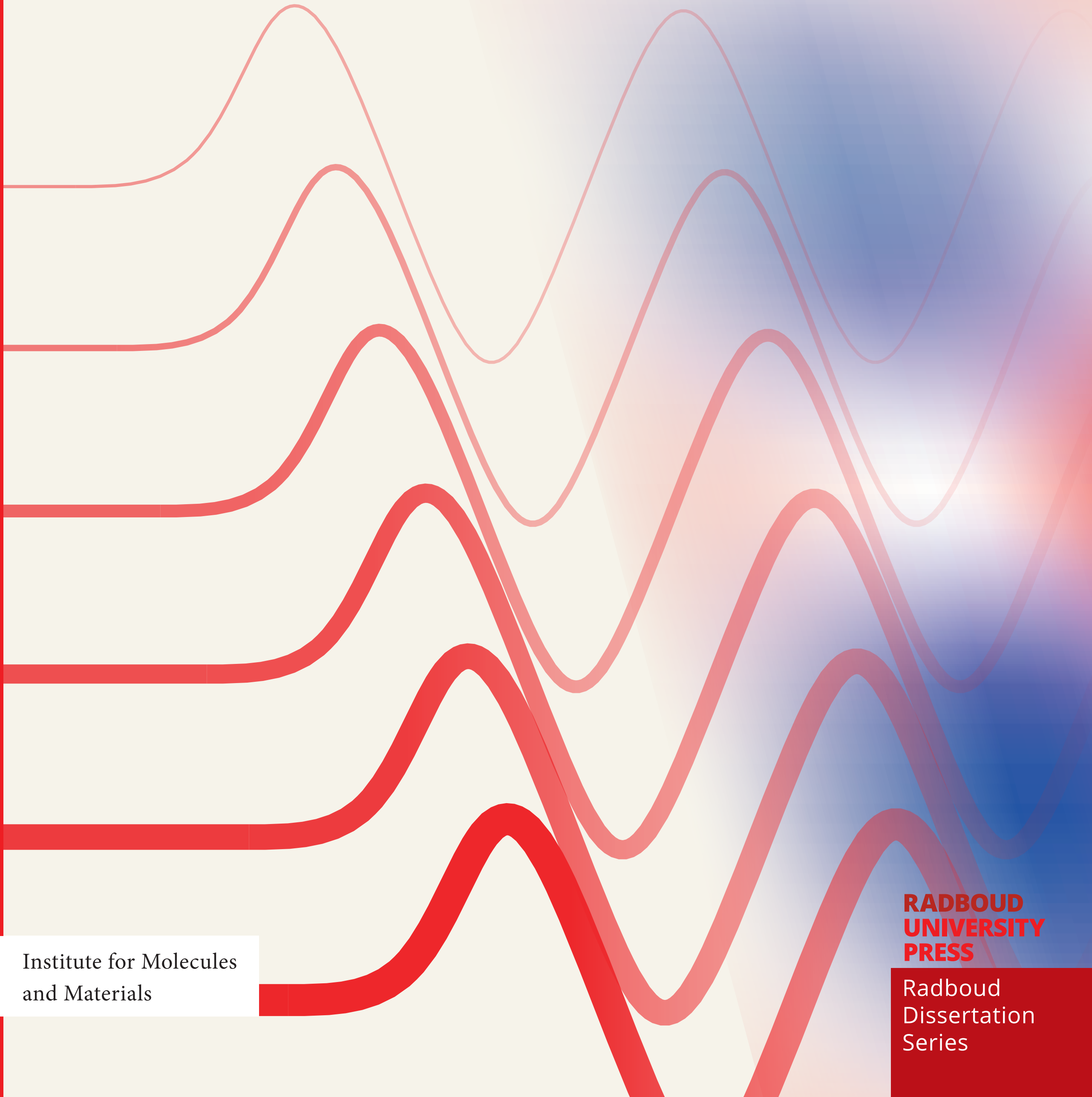
Semiclassical theory for quantum plasmons in two-dimensional inhomogeneous media

Tjacco Mikal Koskamp



Institute for Molecules
and Materials

RADBOUD
UNIVERSITY
PRESS

Radboud
Dissertation
Series

Semiclassical theory for quantum plasmons in two-dimensional inhomogeneous media

T. M. Koskamp

Semiclassical theory for quantum plasmons in two-dimensional inhomogeneous media
T. M. Koskamp

Radboud Dissertation Series
ISSN: 2950-2772 (Online); 2950-2780 (Print)

Published by RADBOUD UNIVERSITY PRESS
Postbus 9100, 6500 HA Nijmegen, The Netherlands
www.radbouduniversitypress.nl

Design: T. M. Koskamp
Cover: T. M. Koskamp
Printing: DPN Rikken/Pumbo

ISBN: 9789465151847
DOI: 10.54195/9789465151847
Free download at: https://doi.org/10.54195/9789465151847

© 2025 T. M. Koskamp

RADBOUD
UNIVERSITY
PRESS

This is an Open Access book published under the terms of Creative Commons Attribution-Noncommercial-NoDerivatives International license (CC BY-NC-ND 4.0). This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator, see http://creativecommons.org/licenses/by-nc-nd/4.0/.

Semiclassical theory for quantum plasmons in two-dimensional inhomogeneous media

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J. M. Sanders,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 10 november 2025
om 16:30 uur precies

door

Tjacco Mikal Koskamp

geboren op 8 augustus 1995
te Boxmeer

PROMOTOR	Prof. dr. M. I. Katsnelson
COPROMOTOR	Dr. K. J. A. Reijnders
MANUSCRIPTCOMMISSIE	Prof. dr. H. M. Cuppen (voorzitter)
	Prof. dr. T. Wehling (Universität Hamburg, Duitsland)
	Dr. J. McIver (Columbia University, Verenigde Staten)

Contents

1 Towards an analytical theory for plasmonic systems 1
- 1.1 Introduction to plasmons 2
  - 1.1.1 From a classical to quantum description 5
  - 1.1.2 Plasmons in different dimensions 7
  - 1.1.3 Bridging the gap: quantum plasmons in complex inhomogeneous environments 9
- 1.2 Theoretical description of plasmons 11
  - 1.2.1 Maxwell's equations and classical plasma oscillations . . 12
  - 1.2.2 Vlasov equation and hydrodynamic model 14
  - 1.2.3 Quantum plasmons from the random phase approximation 18
- 1.3 Semiclassical analysis for inhomogeneous systems 26
  - 1.3.1 The WKB approximation 27
  - 1.3.2 Derivation of the formalism for plasmons 30
  - 1.3.3 Density operator and induced electron density 31

2 Plasmons in arbitrary dielectric environment 37
- 2.1 Effective description of plasmons in 2D systems with arbitrary dielectric environments 38
  - 2.1.1 Separating in-plane and out-of-plane degrees of freedom in the Poisson equation 40
  - 2.1.2 General construction of the effective classical Hamiltonian 45
  - 2.1.3 General construction of the induced potential 49
  - 2.1.4 Interpretation through the energy density 53
  - 2.1.5 Dimensionless parameters and applicability of the semiclassical approximation 55
- 2.2 Application of the theory to two distinct models 57
  - 2.2.1 Plasmons in a simple dielectric environment 58
  - 2.2.2 Plasmons in layered structures with effective height d . 60
- 2.3 Conclusion 62

3 Scattering on radially symmetric inhomogeneities 65
- 3.1 Introduction to scattering 66
- 3.2 Scattering cross sections and phase shifts 68
- 3.3 Derivation of the semiclassical phase shift 70
  - 3.3.1 Alternative expression for the phase shift 76

3.4 Numerical results for the scattering cross section 78
3.5 Conclusion 84

4 (Quasi-)localized states for waveguiding 87
4.1 Introduction to plasmonic waveguiding 88
4.2 Plasmon localization through bound states 89
4.2.1 General analysis of the effective classical Hamiltonian: formation of bound states 89
4.2.2 Numerical implementation semiclassical bound states . 94
4.3 (Quasi-)localization of plasmons through local screening 101
4.3.1 The effect of screening on the amplitude 101
4.3.2 The effect of local screening on the localization of the plasmonic excitation 104
4.3.3 Plasmonic waveguide from amplitude effect 106
4.4 Conclusion 109

5 Plasmonic crystals and band structures 111
5.1 Introduction periodic structures 112
5.2 Periodic modulation for plasmons 113
5.3 The optical theorem for plasmons 117
5.4 The connection problem 119
5.4.1 Stokes diagrams 121
5.4.2 Quantization condition for tunneling 123
5.4.3 Quantization condition for above-barrier scattering . . . 128
5.5 Numerical computation of explicit periodic modulations 130
5.5.1 Gaussian Modulation 132
5.5.2 Cosine Modulation 135
5.6 Conclusion 137

6 Conclusion and Outlook 141
6.1 Outlook 145

Appendix 151
A Alternative derivation of the induced potential $V(\mathbf{x}, z)$ 151
A.1 Leading-order term 152
A.2 Subleading-order term 153
A.3 Transport equation 155
B Energy density in the plane $z = 0$ 156
C Additional derivations for the general description 159
C.1 Derivation of subprincipal symbol Γ_1 at $z = 0$ 159
C.2 Derivation of an expression for the energy density . . . 164
D Simple turning point 167

Bibliography 169

Catching the Wave: a Guide through the Electron Sea (Popular Summary)175

Een Handleiding door de Elektronenzee: Geen Golf te Hoog (Populaire Samenvatting) 179

Acknowledgments 183

Curriculum Vitae 185

1

Towards an analytical theory for plasmonic systems

Plasmonics, the study of controlling plasmons, quasi-particles formed by collective electron oscillations, holds significant promise for manipulating light at the nanoscale. This chapter serves as an introductory overview of plasmons, with a particular focus on their behavior in two-dimensional inhomogeneous media. We begin by establishing the fundamental concepts of plasmons, exploring their characteristics across different regimes, including classical and quantum descriptions. This is followed by a discussion of the influence of dimensionality on plasmonic properties and the added complexities introduced by inhomogeneities. Subsequently, we delve into various theoretical approaches for describing plasmons, highlighting their respective advantages and limitations. To address the challenges posed by inhomogeneous systems, we propose the application of semiclassical techniques within the random phase approximation. We introduce these semiclassical techniques in a general context, providing the essential background for their application. Finally, we lay the theoretical groundwork for the development of our general theory by applying semiclassical techniques to the initial steps of the random phase approximation. This foundation will be further expanded upon and utilized in the subsequent chapters of this thesis.

The content of Section 1.3 in this chapter is based on the research and findings presented in Ref. [1]. Copyright by the American Physical Society.

1.1 Introduction to plasmons

Plasmonic systems, particularly quantum plasmons in two-dimensional inhomogeneous media, offer a rich field of study within condensed matter physics. At the heart of plasmonics lies the concept of a plasma, a state characterized by a significant separation of positive and negative charges, while maintaining overall charge neutrality. This condition necessitates the presence of at least one mobile charge species. Plasmas are ubiquitous in nature, spanning from the ionized gases in the ionosphere and the solar atmosphere to the condensed matter systems found in metals [2]. In the latter, the free electrons, moving amidst a background of positively charged, immobile ions, exemplify a condensed matter plasma. These mobile electrons, under the influence of Coulomb interactions, give rise to collective electron oscillations known as plasma oscillations in classical terms or plasmons in quantum mechanical terms. In this thesis, we consider a quantum description of plasmons and, therefore, use the term plasmons to refer to these excitations.

A simplified explanation of such a collective oscillation is as follows. In a system with free electrons, the total negative charge of the electrons is balanced by the positive charge of the static ions. When free electrons are pushed together to one side by an external electric field, they begin to repel each other due to their negative charges. Simultaneously, they are attracted to the locally uncovered positive ions on the opposite side. When the electric field is removed, this interplay of repulsion and attraction drives the electrons back toward equilibrium, resulting in a coherent motion of charge carriers. The electrons overshoot the equilibrium position and oscillate back and forth, creating a self-sustained oscillation of the electron density. Quantization of these collective oscillations leads to the concept of plasmon quasi-particles, a fundamental excitation in condensed matter systems.

The emergence of plasmons is rooted in the long-range Coulomb interactions between conduction electrons and static ions. These interactions lead to collective oscillations, representing a fundamental excitation of the electron gas. Analogous to longitudinal sound waves in gases and liquids, which are collective oscillations of atoms, plasmons are longitudinal modes, meaning the electron displacement is parallel to the direction of propagation. However, unlike charge-neutral sound waves, plasmons involve the motion of charged electrons, making them strongly coupled to electromagnetic fields. This coupling makes plasmons particularly interesting for optical devices. These characteristics enable the field of plasmonics [3–6], where one seeks to control plasmons for information processing, such as exciting, manipulating, and detecting them.

One of the most compelling features of plasmons is their ability to be excited by visible light, which allows for direct manipulation and excitation using electromagnetic fields. Experimentally, plasmons are excited and detected using techniques such as near-field scanning optical microscopy (SNOM) [7, 8] or excitation via periodic gratings [9]. More importantly, this interaction allows for the confinement of light to extremely small dimensions [4, 10], significantly smaller than those achievable with conventional optics. This sub-wavelength confinement is essential for a wide range of applications, as it enables the miniaturization of optical devices and the enhancement of light-matter interactions.

Plasmonics offers a competitive edge over other methods, particularly in its ability to complement photonic systems. While photonics can significantly enhance electronic systems due to faster speeds and lower losses, it faces limitations in miniaturization due to the large wavelength of visible light. Plasmonics addresses this by enabling the confinement of light to sub-wavelength dimensions, thus allowing for smaller and more compact devices. However, the goal is not to replace photonic systems entirely, as plasmon propagation has its limitations. Plasmons experience damping, which limits their propagation distance. This damping arises from the transfer of collective electron motion into incoherent electron-hole excitations, known as Landau damping [2, 11, 12], or into other forms of excitations such as magnons [13] or phonons [14]. Therefore, in practical applications, plasmons are best suited for short-distance signal processing and manipulation. For long-distance information transfer, other lower-loss carriers like photons in optical fibers are more appropriate. The strategy could, therefore, be to perform complex manipulations in the compact plasmonic regime and then transfer the information to the photonic regime for long-distance transport, acknowledging that plasmons are relatively short-lived particles and not suited for long-term storage of information.

The emergence of plasmonics has significantly impacted the field of nanophotonics, driven by the unique capability of plasmons to confine and manipulate light at the nanoscale. Plasmonic devices have found applications in a wide range of areas. Nanoscale target substances can be detected by monitoring shifts in the plasmon resonance frequency, leading to applications in chemical and biological sensing [15]. Additionally, plasmonics plays a crucial role in lithographic fabrication, enabling the creation of nanoscale patterns with higher precision [6, 16]. Furthermore, plasmonic properties are essential for the applicability of metamaterials, artificial materials with engineered electromagnetic properties [17, 18]. These metamaterials can exhibit exotic phenomena like negative refraction. The strong light-matter interaction offered by plasmons is also crucial for quantum optics. By coupling quantum emitters to plasmonic modes, one can enhance light-matter interactions, paving the way for advancements in quantum information processing [10, 19]. Plasmonics can play a crucial role in the further development and competitiveness of integrated photonics and quantum optics.

In most applications, plasmons can be considered classically, since we are usually in the regime where the plasmon momentum is much smaller than the Fermi momentum of electrons. In this regime, the classical theory of plasma waves is sufficient [2, 20]. While classical plasmonics has achieved significant success in areas like sub-wavelength imaging and enhanced spectroscopy, the field of quantum plasmonics is increasingly relevant [5, 21]. Recent experiments have reached regimes where the quantum nature of plasmons becomes evident, particularly when their wavelengths approach the Fermi wavelength of the electrons [22]. In this quantum regime, electrons no longer behave as classical particles, and a quantum mechanical treatment of the electrons becomes necessary. One prominent effect at these smaller wavelengths is intrinsic Landau damping, caused by single electron-hole excitations, which can decohere the plasmon [2, 12]. The impact of the quantum regime varies with the dimensionality of the system. For instance, in two-dimensional systems, plasmons exhibit the classical square-root dispersion relation at long wavelengths. However, at short wavelengths, their behavior deviates from this classical picture, and a quantum mechanical description of the electron-electron interactions becomes increasingly prominent. These effects become more important at higher energies or frequencies, and in smaller systems with higher confinement. Section 1.1.1 discusses the theoretical descriptions of such systems, focusing on the evolution from classical to quantum models.

The behavior of plasmons is governed by the intricate interplay of charged particles via Coulomb interactions and electron kinematics, manifesting distinct characteristics depending on the dimensionality of the electron gas. In three-dimensional (3D) systems, electrons are free to move in all three spatial dimensions, whereas in two-dimensional (2D) systems, electrons are confined to move in a plane, with motion restricted to two dimensions. Section 1.1.2 provides a foundational overview of plasmon physics, emphasizing the contrasting features observed in 3D and 2D electron systems, with a particular focus on the latter due to their growing importance in modern plasmonics. The versatility of plasmons is amplified in 2D materials, which offer enhanced tunability compared to their 3D counterparts. The experimental discovery of graphene paved the way for the exploration of other 2D materials [23, 24]. Advances in fabrication techniques have made these materials widely accessible, enabling the experimental realization of the plasmonic quantum regime [3, 5, 21]. Research groups are actively exploring the properties of plasmons in 2D materials, highlighting their potential for various applications [7, 8].

Despite the significant progress in plasmonics, a comprehensive theoretical framework for quantum plasmons in inhomogeneous media is still lacking. Plasmonic systems, where the manipulation of plasmons is key, are inherently inhomogeneous due to variations in material properties, geometry, or doping profiles. While existing theories adequately describe homogeneous systems, they fall short in capturing the complexities introduced by inhomogeneities. This thesis aims to address this gap by developing a semiclassical theory for plasmons in two-dimensional inhomogeneous media, which will be discussed in more detail in Sec. 1.1.3. This theoretical framework provides a foundation for understanding and predicting the behavior of plasmons in realistic devices, paving the way for the design and optimization of future plasmonic technologies.

To facilitate this comprehensive foundation, the remainder of this introductory section explores the core characteristics of plasmonics, focusing on the distinctions between classical and quantum behavior, the impact of dimensionality, and the complexities of inhomogeneous systems. This overview aims to establish a conceptual framework without overwhelming the reader with detailed formulas. Subsequently, Sec. 1.2 delves into the theoretical underpinnings, tracing the evolution from classical to quantum descriptions and presenting the essential formulas for both regimes. Finally, Sec. 1.3 introduces the semiclassical analysis, detailing the relevant formulas that will be applied to the study of plasmonic systems in this work.

1.1.1 From a classical to quantum description

As discussed in the previous section, recent experimental progress has enabled access to the quantum regime for plasmons [3, 21, 22], opening new avenues for plasmonic waveguides and necessitating a refined theoretical understanding for the practical application of plasmonic systems. In this regime, a quantum description of the electron-electron interactions plays a crucial role in modifying the plasmon dispersion, and a description solely based on classical theories is no longer sufficient. The fundamental distinction between classical and quantum plasmonics can be characterized by the relative magnitudes of the plasmon momentum $|\mathbf{q}|$ and the Fermi momentum p_{F}:

- $|\mathbf{q}| \ll p_{\mathrm{F}}$: Classical limit, classical treatment of electron-electron interactions is sufficient.
- $|\mathbf{q}| \approx p_{\mathrm{F}}$: Quantum limit, quantum mechanical treatment of electron-electron interactions is required.

Consequently, the theoretical description of plasmons has evolved from classical electrodynamics to sophisticated quantum mechanical frameworks, reflecting the increasing complexity of experimental observations and technological advancements.

In the classical limit, plasmons can effectively be described using macroscopic models that treat the electron gas as a continuous fluid. These models, based on classical electrodynamics and fluid dynamics, accurately predict the behavior of plasmons at long wavelengths, where the collective motion of electrons dominates. However, as plasmon wavelengths approach the Fermi wavelength, the quantum nature of electrons becomes significant, and a quantum mechanical treatment of electron-electron interactions becomes necessary. In this quantum regime, the classical description of electron behavior is no longer sufficient.

At the foundation, classical theories for plasma oscillations rest on Maxwell's equations, which describe the propagation of electromagnetic fields in materials [2, 25]. In 3D bulk materials, the uniform electron density leads to well-defined longitudinal bulk plasma oscillations, with a finite energy, commonly known as the plasma frequency. Moreover, at an interface between a conductor and a dielectric, Maxwell's equations predict the existence of a different plasma wave, surface plasma oscillations or surface plasmons. In contrast, these field equations are not suited for describing plasma oscillations in 2D systems, as one needs to incorporate the behavior of the electrons in such a layer.

Going beyond Maxwell's equations, and treating the electrons as kinetic particles, the oscillatory motion of charged particles in a plasma is described classically by the Vlasov equations [20]. This becomes necessary because Maxwell's equations alone do not account for the diffusion of electrons [2]. The Vlasov equations, which combine Maxwell's equations with the Boltzmann equation, provide a kinetic description of electron dynamics. This classical theory accurately describes plasmons in the regime where the Fermi wavelength of electrons is much smaller than the plasmon wavelength [12]. Hydrodynamic theories simplify the Vlasov equation by taking velocity moments, resulting in equations for macroscopic quantities like density, velocity, and pressure [3, 20, 21]. These theories offer a simplified description of long-wavelength plasmons. For inhomogeneous systems within this classical regime, numerical particle-in-a-box simulations are often employed [3, 21].

Vlasov's initial attempt to derive a dispersion relation for plasma waves was refined by Landau [26, 27], who recognized the importance of treating the problem as an initial value problem, leading to the discovery of Landau damping. This phenomenon, which is essentially classical, highlights a limitation of the Vlasov approach.

At shorter wavelengths, when the plasmon wavelength is on the order of the Fermi wavelength of electrons, we enter the regime of quantum plasma oscillations. The associated quasi-particle, termed plasmon, was first described by Pines and Bohm [2, 28, 29]. The plasmon dispersion is determined by the roots of the dielectric function, which describes the electron response to external fields. Within the random phase approximation (RPA), this dielectric function is described by the Lindhard dielectric function [2, 28]. The RPA predicts distinct plasmon dispersions, applicable even in the short-wavelength limit, and offers a microscopic understanding of Landau damping through electron-hole excitations.

From an analytical perspective, the Lindhard dielectric function can be derived analytically in homogeneous systems with parabolic electronic Hamiltonians [2, 28]. However, these solutions rely heavily on Fourier transforms, which are not applicable to inhomogeneous systems. Notably, for graphene, which is described by the Dirac Hamiltonian, a distinct theoretical framework has been developed [30, 31]. Furthermore, for 2D plasmons, an analytical derivation of the full plasmon dispersion is attainable [28, 32, 33], valid for all wavelengths. To accurately capture nonlocal effects arising from the Coulomb tail, it is crucial to account for the effective finite height of materials [34–36]. This introduces additional complexities that significantly impact electron screening and plasmon properties, as demonstrated by Stern and Howard [37]. While numerical methods, such as real-space RPA and tight-binding propagation, have been successfully employed to study plasmons in inhomogeneous systems, including fractals and 2D waveguides [38–41], a comprehensive analytical theory for inhomogeneous systems remains elusive.

In summary, the transition from classical to quantum descriptions of plasmons is essential to accurately capture their behavior, particularly in 2D inhomogeneous systems where higher frequencies and, consequently, shorter wavelengths are excited. These quantum effects are also significantly influenced by smaller spatial variations. The analytical description of quantum plasmons across the full spectrum in inhomogeneous systems remains a challenge. Typical theoretical techniques applicable to translationally invariant systems, such as Fourier analysis, are not applicable, necessitating the development of new theoretical frameworks. In Sec. 1.2, we provide a more in-depth review of the analytical theories commonly used to describe plasmons, and elucidate the need for a novel theoretical framework.

1.1.2 Plasmons in different dimensions

The dimensionality of an electron system profoundly influences its plasmonic properties, leading to distinct behaviors. Understanding these differences between three-dimensional and two-dimensional systems is fundamental to advancing plasmonic technologies. This distinction arises primarily from the confinement of electrons and the resulting modifications to Coulomb interactions, which dictate the characteristics of plasmon excitations [28]. Within 3D systems, we can further distinguish between bulk plasmons, where electrons move freely throughout the material, and surface plasmons, which are confined to the surface of the conductor, while still subject to screening from the bulk electrons. This crucial difference highlights that surface plasmons, despite their surface localization, are fundamentally different from 2D plasmons, where electrons are strictly confined to a plane, devoid of bulk electrons. In the following discussion, we explore the unique properties of plasmons in both 3D and 2D environments, examining how confinement and dimensionality shape their behavior.

Plasmons exhibit diverse characteristics depending on the dimensionality of the system, some of which are apparent in daily life. To understand these plasmon characteristics, it is instructive to consider the dispersion relation, which relates the energy of the excited plasmon to the wavelength or momentum. For a three-dimensional electron gas, the most fundamental plasmon mode is the bulk plasmon, a collective oscillation of the electron density throughout the material in all three directions, possessing a minimal finite frequency. This finite frequency, known as the plasma frequency ω_p, is the reason why metals are reflective in the visible range [2, 25]. Photons with energies below $\hbar\omega_\mathrm{p}$ are strongly reflected because they lack the energy to excite bulk plasmons.

Beyond bulk plasmons, surface plasmons (SPs) are collective oscillations of the electron density confined to the interface between a bulk metal and a dielectric material. These surface-bound modes, known as surface plasmon polaritons (SPPs), propagate along the interface and decay exponentially into both media. The dispersion relation for SPPs at a planar metal-dielectric interface, within a classical model for the metal's dielectric function, exhibits a characteristic long-wavelength limit. In this limit, the energy approaches zero linearly [2]. This implies that these excitations can be excited at visible energies and below, making them crucial for nanophotonics. Due to their ability to confine light to nanoscale dimensions, they enable the development of highly compact optical devices [42, 43]. An important class of surface plasmons are nonpropagating localized surface plasmons (LSPs), confined to metallic nanoparticles or nanostructures. LSPs arise from collective oscillations of electrons within a confined nanostructure with defined boundaries, differing from SPPs in that they do not propagate but are quantized standing waves, due to confinement in all directions. The resonance frequency of LSPs is highly sensitive to the geometry of the nanoparticle and the surrounding dielectric environment, making them ideal for sensing applications. The vibrant colors observed in stained glass windows can be attributed, in part, to LSPs, where metallic nanoparticles embedded in the glass selectively interact with light depending on their size and shape, thus contributing to the observed color.

Reducing the dimensionality of the electron system itself leads to two-dimensional plasmons, which occur in monolayer systems like graphene, TMDCs, and other 2D electron gases (2DEGs). In these systems, the plasmon dispersion relation takes a characteristic square-root form $\hbar\omega_{\mathrm{pl}}(\mathbf{q}) \propto \sqrt{|\mathbf{q}|}$, where $\mathbf{q}$ is the plasmon momentum. This distinct dispersion relation, different from both three-dimensional bulk plasmons and surface plasmons, arises from the reduced dimensionality and the altered Coulomb interaction between electrons. Two-dimensional plasmons are particularly interesting due to the possibility of long-wavelength excitations, similar to SPPs, and their enhanced tunability, facilitated by their strong interaction with the substrate environment [39, 40]. This tunability stems from the strong influence of the dielectric environment surrounding the two-dimensional electron gas on the 2D plasmon frequency, offering a means to control the plasmon properties without altering the active material itself. Furthermore, the low energy of the plasmon excitation makes them accessible with a wider range of experimental techniques.

This subsection has explored the diverse characteristics of plasmons across different dimensions, highlighting the distinctions between 3D bulk and surface plasmons, and 2D systems. Among these, 2D plasmons stand out due to their enhanced tunability and potential for sub-wavelength confinement, making them particularly relevant for advancing integrated photonics. This sub-wavelength confinement, achieved because plasmons have much smaller wavelengths than photons at the same frequency, allows for the miniaturization of optical circuits beyond the diffraction limit. The ability to control plasmon properties through the dielectric environment, without altering the active material, coupled with the low energy excitation, offers significant advantages for practical applications. Therefore, in this thesis, we focus specifically on the description of 2D plasmons and their propagating behavior.

1.1.3 Bridging the gap: quantum plasmons in complex inhomogeneous environments

Plasmons in homogeneous systems, regardless of dimensionality or quantum regime, are analytically straightforward to understand from a theoretical perspective, making them a common starting point. However, experimental realizations are inherently inhomogeneous. Imagine a water wave propagating across a perfectly uniform ocean; while the wave's behavior is predictable, it becomes truly interesting when it encounters a coastline or an obstacle. Similarly, plasmons in homogeneous systems, though valuable for fundamental study, require inhomogeneities to be controlled, excited, and detected. Most notably, experimental systems always have finite sizes, and variations in material properties, geometry, or doping profiles introduce spatial dependencies. These spatial dependencies significantly alter plasmon behavior. Consequently, a theoretical framework capable of describing plasmons in inhomogeneous environments is essential for both fundamental understanding and practical applications.

Obtaining analytical solutions for inhomogeneous systems is significantly more challenging than for homogeneous ones due to the broken translational symmetry. Techniques that are analytically applicable when there is translational invariance, such as Fourier transforms, become ineffective. For instance, within the random phase approximation, analytical solutions for homogeneous systems are often derived by transforming to momentum space. However, in inhomogeneous systems, the spatial dependence of system parameters precludes such transformations. Similarly, in classical descriptions, spatial variations invalidate the analytical plane-wave formalism. Consequently, analytical treatments of inhomogeneous plasmonic systems require more sophisticated approaches that can handle spatial variations without relying on translational symmetry or a single wavevector description.

The current state-of-the-art approach for studying inhomogeneous plasmonic systems is numerical simulation. While powerful, these methods are inherently limited by computational constraints, restricting the size of systems that can be investigated. Real-space diagonalization within the random phase approximation is a common numerical technique [38–40], but it is computationally intensive, requiring the diagonalization of large matrices and a substantial basis set. Consequently, numerical RPA approaches are often limited to small system sizes. While tight-binding methods can handle considerably larger systems [41, 44–46], they introduce additional approximations that are not always controllable. From a theoretical perspective, classical approaches, such as particle-in-a-box models, can describe localized surface plasmons as quantized modes [3, 21], but these models do not capture propagating quantum plasmons. Furthermore, inhomogeneous 2D systems have been studied analytically using the Lippmann-Schwinger equation to describe plasmon scattering [47], but a comprehensive analytical theory remains elusive. Therefore, to gain a better understanding of the underlying physics, it is essential to develop (semi-)analytical approaches that complement numerical methods.

In pursuit of developing (semi-)analytical approaches that complement numerical methods, the semiclassical approximation emerges as a powerful tool for describing inhomogeneous systems. Semiclassical methods are applicable when the wavelength of the underlying particle (in the case of plasmons, the electron wavelength) is much smaller than the characteristic length scale of the inhomogeneity [48, 49]. These methods have been successfully applied to 3D quantum plasmons in inhomogeneous systems [50]. In this thesis, this method is extended to plasmons in two-dimensional inhomogeneous media, addressing its applicability and limitations, and comparing results to numerical approaches. The semiclassical approximation, which constructs asymptotic solutions of the form $\exp(iS(x)/\hbar)$, where $S(x)$ is the classical action, offers an analytical approach to tackling inhomogeneous problems. This method has a rich history, with early applications of the Wentzel-Kramers-Brillouin (WKB) approximation to inhomogeneous electron systems [51, 52] to study the bound states near the edge of a metal [53] and in atoms [54]. Recent advancements [50] have led to a more rigorous formulation using pseudo-differential operators and the Maslov canonical operator [48, 55]. We demonstrate how, for 2D systems, within the semiclassical approximation, the out-of-plane degrees of freedom can be decoupled from the in-plane degrees of freedom using operator separation of variables [56], potentially leading to an alternative derivation of the background screening model.

It is important to clarify that the term "semiclassical approximation" does not imply a hybrid classical-quantum description built upon classical modes. Rather, it refers to a method for approximating quantum mechanical solutions, in our case derived from the random phase approximation, using techniques from classical mechanics. Specifically, it exploits a small dimensionless parameter that relates the length scale of spatial variations (ℓ) to a characteristic wavelength, such as the electron wavelength (λ_{el}). When $\ell \gg \lambda_{\mathrm{pl}} \approx \lambda_{\mathrm{el}}$ (the second approximation ensures the quantum limit), the system locally resembles a homogeneous one, allowing us to approximate the quantum plasmon behavior using classical concepts like trajectories derived from Hamilton's equations and the classical action governed by the Hamilton-Jacobi equation. Crucially, this approach maintains the full quantum mechanical nature of the plasmon, including Landau damping and the quantum mechanical treatment of electron-electron interactions as reflected in the dispersion relation. In the limit where the dimensionless parameter approaches zero (i.e., ℓ becomes infinitely larger than λ_{el}), the full quantum homogeneous RPA solution is recovered. Therefore, semiclassics provides a powerful way to describe quantum plasmons in inhomogeneous systems using techniques from classical mechanics, without sacrificing the essential quantum features.

Building upon the semiclassical theory for inhomogeneous plasmon systems in three dimensions [50], we extend this theory to two dimensions by utilizing the method of operator separation of variables [56] to adiabatically separate the in-plane from out-of-plane degrees of freedom. Methodologically, we utilize the contemporary formulation of the semiclassical approximation, leveraging the Maslov canonical operator [48] to tackle the complexities of true pseudo-differential operators arising from the dielectric function. This approach provides a valuable opportunity to apply this theory to a concrete physical problem. Physically, we describe plasmons in dielectric environments with arbitrary z-dependence. We also describe bound plasmon states in waveguides, band structures in plasmonic crystals (periodic modulations), and plasmon scattering, constructing the total and differential scattering cross-section. Therefore, this thesis aims to bridge the gap by developing a semiclassical analytical framework for describing 2D quantum plasmons in inhomogeneous environments, providing a deeper understanding and facilitating the design of advanced plasmonic devices.

1.2 Theoretical description of plasmons

This thesis aims to develop a comprehensive, almost fully analytical theory for plasmons in two-dimensional inhomogeneous systems. To achieve this, we must first understand the foundational theoretical descriptions of plasmons, recognizing both their strengths and limitations. Plasmons, as collective electron excitations, are intrinsically linked to electromagnetism and the electronic behavior of materials. This section provides a concise historical overview of the theoretical approaches used to describe plasmons, starting from classical descriptions and progressing to quantum theories, ultimately focusing on the challenges posed by inhomogeneous quantum systems. Examining the shortcomings of existing theories, particularly their challenges in handling shorter plasmon wavelengths and inhomogeneities, motivates the development of a robust semiclassical framework capable of describing 2D quantum plasmons in realistic devices.

This section systematically explores the theoretical landscape, starting with the classical regime. We begin with Maxwell's equations, which provide a fundamental understanding of plasma oscillations in bulk and at interfaces, serving as the classical counterpart to plasmon descriptions. We then present a summary of the Vlasov equations and hydrodynamic models, extending the classical picture to describe the unique dispersion of 2D plasmons in the long-wavelength limit. A short discussion of the particle-in-a-box approach highlights its relevance to localized plasmons in inhomogeneous systems such as nanoparticles. Finally, we delve into the quantum mechanical random phase approximation, the cornerstone of modern plasmon theory, and demonstrate how the analytical calculation of the plasmon dispersion relies on the Fourier transform. The analytical description of plasmons is significantly challenged by inhomogeneous systems, which greatly complicates the analytical application of Fourier techniques. This progression from classical to quantum descriptions sets the stage for the development of a semiclassical theory that can address the complexities of 2D inhomogeneous plasmonic environments.

1.2.1 Maxwell's equations and classical plasma oscillations

Maxwell's equations, the foundation of classical electromagnetism, describe the interplay between electric and magnetic fields and their interaction with matter. When the conduction electrons are viewed as a plasma, Maxwell's equations yield solutions describing plasma oscillations, which are self-sustained collective excitations of electrons coupled to electromagnetic waves. Solving Maxwell's equations for 3D systems in simple, isotropic media reveals the existence of two distinct modes: transverse and longitudinal. In the specific case of an interface, applying appropriate boundary conditions reveals the existence of an additional wave type, the surface plasma wave. In this section, we briefly discuss these 3D findings and then examine why this approach is insufficient for 2D systems.

Maxwell's equations in differential form in Gaussian units are given by

$$\nabla \cdot \mathbf{D} = 4\pi\rho \tag{1.1}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1.2}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \tag{1.3}$$

$$\nabla \times \mathbf{H} = \frac{4\pi}{c}\mathbf{J} + \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} \tag{1.4}$$

where $\mathbf{D}$ is the electric displacement field, $\mathbf{B}$ is the magnetic field, $\mathbf{E}$ is the electric field, $\mathbf{H}$ is the magnetic field intensity, ρ is the external charge density, c is the speed of light, and $\mathbf{J}$ is the external current density. In a material, the displacement field $\mathbf{D}$ and magnetic field intensity $\mathbf{H}$ are related to the electric and magnetic fields through the constitutive relations:

$$\mathbf{D} = \varepsilon_{\mathrm{b}}\mathbf{E} + \mathbf{P} \tag{1.5}$$

$$\mathbf{H} = \mathbf{B} - 4\pi\mathbf{M} \tag{1.6}$$

where ε_{b} is the dielectric constant of the material, $\mathbf{P}$ is the polarization and $\mathbf{M}$ is the magnetization of the material. For nonmagnetic materials, the magnetization $\mathbf{M}$ is zero. In the context of self-sustained plasma oscillations, we assume that the external current density and external charge density are both zero.

Following the argumentation in [2], for a 3D system, both transverse and longitudinal solutions emerge. The transverse solutions correspond to propagating electromagnetic waves within the bulk of the material. The longitudinal solutions, on the other hand, represent self-sustained collective oscillations of the electron density, termed bulk plasma oscillations. Both modes exhibit a characteristic frequency in the long-wavelength limit, the plasma frequency, given by

$$\omega_{\mathrm{p}} = \sqrt{4\pi \frac{n^{(0)} e^2}{m_{\mathrm{eff}}}}, \tag{1.7}$$

where $n^{(0)}$ is the equilibrium electron density, e is the elementary charge, and m_{eff} is the effective electron mass. Remarkably, this classical picture accurately predicts the plasma frequency. For longitudinal (plasma) waves, the plasma frequency is the frequency of (electron) oscillations in the long-wavelength limit, whereas for transverse waves (photons), this is the limiting frequency below which photons cannot propagate in metals.

While Maxwell's equations alone are sufficient to describe the fundamental bulk plasma oscillations in three-dimensional systems, their application to two-dimensional plasmas is more subtle. In 3D, the longitudinal plasma oscillations arise naturally from Gauss's law, which relates the electron density fluctuations to an induced restoring electric field. This results in the well-known plasma frequency, which remains non zero even in the long-wavelength limit. The reason for this lies in the Coulomb interaction in three dimensions, where a charge perturbation generates a restoring electric field that scales as $1/r$. Consequently, the interaction between charge fluctuations and the self-consistent electric field directly leads to bulk plasma oscillations without requiring explicit consideration of the electron equation of motion. This derivation relies solely on the electromagnetic response of the material rather than the microscopic motion of individual charges, making it a broadly applicable result within classical electrodynamics.

In contrast, when considering a two-dimensional electron gas embedded in three-dimensional space, the Fourier component of the Coulomb potential scales as $1/|\mathbf{q}|$ rather than $1/|\mathbf{q}|^2$, leading to fundamentally different screening behavior. As a result, the field response to charge perturbations in 2D differs significantly from the 3D case, and Maxwell's equations alone do not determine the full dispersion relation. Additionally, a truly two-dimensional picture presents a challenge within the framework of classical Maxwell's equations due to the instantaneous quantization of electron movement in the out-of-plane direction. This concept, while a useful idealization, inherently requires the explicit inclusion of the electrons or charge particles themselves within the equations. Without this inclusion, the infinitesimally thin layer becomes a passive boundary, unable to interact with the electromagnetic fields, hence rendering the 2D description incomplete. Thus, modified approaches are necessary to accurately describe 2D plasma oscillations.

While truly two-dimensional plasmons require approaches beyond Maxwell's equations, solutions to these equations do describe a related phenomenon at 2D interfaces: surface plasma waves. These surface-bound modes, known as surface plasmons, are collective oscillations of the electron density confined to the interface between a conductor and a dielectric. They propagate along the interface, decay exponentially in the out-of-plane direction, and are screened on one side by the bulk electrons. The dispersion relation for surface plasma waves at a planar metal-dielectric interface, within a classical model for the metal's dielectric function, has two characteristic limits in the long-wavelength and short-wavelength regimes [2], namely

$$\omega_{\mathrm{SP}} = \begin{cases} c|\mathbf{q}|/\sqrt{\varepsilon_{\mathrm{b}}} & |\mathbf{q}| \ll \omega_{\mathrm{p}}/c \\ \omega_{\mathrm{p}}/\sqrt{1+\varepsilon_{\mathrm{b}}} & |\mathbf{q}| \gg \omega_{\mathrm{p}}/c \end{cases}, \tag{1.8}$$

where ω_{SP} is the SP frequency, $\mathbf{q}$ is the SP momentum, and c is the speed of light. The dielectric constant, ε_{b}, is the background static dielectric constant of the dielectric, which equals $\varepsilon_{\mathrm{b}} = 1$ for a vacuum. In the long-wavelength limit, the energy approaches zero, which implies that these excitations can be excited at visible frequencies and below, and are therefore crucial for nanophotonics. Due to their ability to confine light to nanoscale dimensions, they enable the development of highly compact optical devices [4, 6, 10].

The purely classical picture of plasma oscillations derived from Maxwell's equations for electromagnetic waves, and specifically the linear dispersion of surface plasmons in the long-wavelength limit, contrasts with the expected square-root dispersion for truly two-dimensional plasmons. A key omission in this approach is the treatment of electrons as charged particles. In reality, the inhomogeneous induced electron density not only contributes to the current but also enables electron diffusion. This particle-like behavior is absent in the purely wave-like description provided by Maxwell's equations. Consequently, the dispersion relation for truly 2D plasmons necessitates the explicit incorporation of electron density dynamics, achieved through methods like linearized hydrodynamic equations or the dielectric function approach within the random phase approximation. This results in the characteristic square-root dispersion, $\omega \propto \sqrt{|\mathbf{q}|}$, which differs significantly from the frequency-independent behavior of bulk plasma oscillations in 3D and the light-like dispersion of surface plasmons in the long-wavelength limit. Therefore, to accurately capture the physics of truly 2D plasma oscillations, Maxwell's equations must be supplemented with models that explicitly account for electron density fluctuations.

1.2.2 Vlasov equation and hydrodynamic model

Building upon the foundation of Maxwell's equations, a more refined classical description of plasmon dynamics can be achieved by treating electrons as kinetic particles, using the Vlasov equations [20]. These equations, which combine Maxwell's equations with the Boltzmann equation in the collisionless limit, provide a kinetic description of electron dynamics, incorporating the long-range Coulomb interaction. However, this theory does not include electron-hole pair excitations. This classical theory provides an accurate description of plasmons in the regime where the Fermi wavelength of electrons is much smaller than the plasma wavelength, a condition that ensures the collective behavior of the electrons dominates over quantum effects.

To simplify the complex kinetic description of the Vlasov equation, hydrodynamic theories can be employed [3, 21]. These theories simplify the Vlasov equation by taking velocity moments of the distribution function $f(\mathbf{x}, \mathbf{v}, t)$, resulting in equations for macroscopic quantities such as electron density $n(\mathbf{x}, t)$, average electron velocity $\mathbf{u}(\mathbf{x}, t)$, and pressure. This approach provides a macroscopic description of long-wavelength plasmons, effectively representing the classical limit. Specifically, taking the zeroth moment of the Vlasov equation yields the continuity equation for the electron density, while the first moment results in the Euler equation for the average velocity. Applying plane wave solutions and linearizing these equations, followed by using Poisson's equation in 2D, allows us to derive the dispersion relation.

To derive the dispersion relation for 2D plasmons within the hydrodynamic model, we begin with the Vlasov equation, which describes the time evolution of the distribution function $f(\mathbf{x}, \mathbf{v}, t)$ [57], namely

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f - \frac{e}{m_{\text{eff}}} \mathbf{E} \cdot \nabla_{\mathbf{v}} f = 0, \tag{1.9}$$

where $\mathbf{E}$ is the electric field, e is the electron charge, m_{eff} is the effective electron mass, and $\mathbf{v}$ is the velocity of individual electrons. We define the electron density $n(\mathbf{x}, t)$ as the zeroth moment and the macroscopic average electron velocity $\mathbf{u}(\mathbf{x}, t)$ as the first moment of the distribution function, given by

$$n(\mathbf{x}, t) = \int f(\mathbf{x}, \mathbf{v}, t) d\mathbf{v} \tag{1.10}$$

$$\mathbf{u}(\mathbf{x}, t) = \frac{1}{n} \int \mathbf{v} f(\mathbf{x}, \mathbf{v}, t) d\mathbf{v}, \tag{1.11}$$

where the integrals are taken over the full 2D velocity space.

Integrating the Vlasov equation over the single electron velocity, effectively averaging over the velocity distribution, yields the continuity equation for the electron density $n(\mathbf{x}, t)$, namely

$$\frac{\partial n}{\partial t} + \nabla \cdot (n\mathbf{u}) = 0. \tag{1.12}$$

This equation ensures the conservation of charge within the electron fluid, relating the time variation of the electron density to the divergence of the particle flux density $n\mathbf{u}$. This reflects the fundamental principle that electrons are neither created nor destroyed within the hydrodynamic approximation. Notably, the term $\nabla \cdot (n\mathbf{u})$ arises directly from the $\mathbf{v} \cdot \nabla f$ term in the Vlasov equation, representing a diffusion term omitted in the previous Maxwell's equations approach.

Taking the first moment of the Vlasov equation, which is mathematically represented as the integral $\int \mathbf{v} d^2 v$ applied to the Vlasov equation, leads to the momentum balance equation or the Euler equation, namely

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\frac{e}{m_{\mathrm{eff}}}\mathbf{E} - \frac{1}{m_{\mathrm{eff}} n}\nabla P. \tag{1.13}$$

These hydrodynamic equations, coupled with Poisson's equation, form a closed set that can be solved to describe the dynamics of the electron fluid and, in particular, the collective oscillations associated with plasmons [28]. Linearizing these equations around the equilibrium state and assuming a time-harmonic dependence of the form $e^{-i\omega t}$, we can derive the plasmon dispersion relation. Expressing the frequency as energy, via $E = \hbar\omega$, allows us to represent the result as

$$E = \sqrt{\frac{|\mathbf{q}| e^2 g_{\mathrm{s}} p_{\mathrm{F}}^2}{2 m_{\mathrm{eff}} \varepsilon_{\mathrm{b}}}}. \tag{1.14}$$

It is important to note that this representation, despite using this plasmon energy, remains purely classical, as the underlying Vlasov equation does not incorporate Planck's constant. This energy representation facilitates a more direct comparison with quantum mechanical results. Furthermore, we defined the Fermi momentum p_{F}, which is related to the electron density $n^{(0)}$ via the Thomas-Fermi approximation:

$$p_{\mathrm{F}} = \hbar \left(\frac{4\pi}{g_{\mathrm{s}}} n^{(0)} \right)^{1/2}, \tag{1.15}$$

where g_{s} is the spin degeneracy. This approximation is popular because of its simplicity but it neglects correlation and exchange effects of the electrons. For a more detailed discussion of the Thomas-Fermi approximation and its validity, see Ref. [21, 28, 58, 59]. Equation (1.14) shows that the hydrodynamic model provides a dispersion relation for 2D plasmons, which is not attainable using Maxwell's equations alone. The characteristic square-root dependence on the momentum $|\mathbf{q}|$ is a signature of 2D plasmons, reflecting the unique Coulomb interaction in two dimensions. In principle, one could add higher order corrections by adding an additional force through the pressure gradient ∇P. Here, P is the pressure, which can be approximated, for a noninteracting electron gas at zero temperature, by $P = nE_{\mathrm{F}}/2$ [28]. For a 2D electron gas, we then have $P = m_{\mathrm{eff}} n v_{\mathrm{F}}^2/2$, where v_{F} is the Fermi velocity, with the Fermi energy $E_{\mathrm{F}} = m_{\mathrm{eff}} v_{\mathrm{F}}^2/2$. However, this is a relatively limited approximation, as it relies on the assumption of local equilibrium of the Fermi surface, an assumption that breaks down for plasmons at higher momenta [28].

While the hydrodynamic model provides a valuable starting point, particularly for capturing the long-wavelength behavior, it has limitations. It does not fully account for the microscopic details of electron-electron interactions, which become increasingly important at shorter wavelengths and can significantly influence plasmon properties. Additionally, Vlasov's initial attempt to derive a dispersion relation for plasma waves was shown to be incomplete by Landau [26, 27], who recognized the importance of treating the problem as an initial value problem, leading to the discovery of Landau damping, a phenomenon not captured by the hydrodynamic model. Therefore, while the hydrodynamic model correctly predicts the form of the long-wavelength dispersion, more sophisticated approaches are needed for a complete description of 2D plasmons for all momenta, as well as incorporating Landau damping.

Particle-in-a-box models

The hydrodynamic model, while only applicable for plasmons in the long-wavelength limit, can be adapted to describe confined plasmonic excitations in inhomogeneous nanostructures. In such systems, the details of the electron confinement and the specific geometry play a crucial role in determining the plasmon properties. For these scenarios, an approach analogous to the quantum mechanical particle-in-a-box model can offer valuable insights [3, 21]. This approach is particularly useful for specific inhomogeneous systems, such as nanoparticles, where the electron confinement leads to quantization of plasmonic modes.

In highly confined metallic nanostructures, such as nanoparticles, the confinement of conduction electrons leads to quantization of plasmonic modes, analogous to the quantum mechanical particle-in-a-box model. This simplified picture provides an intuitive understanding of how the geometry and size of a nanoparticle influence its plasmon resonances. The electrons are treated as confined within the nanoparticle's boundaries, and their collective oscillations are quantized due to these boundary conditions. The resonance energies of the confined plasmons are then determined by the size and shape of the box, as well as the effective mass of the electrons and the dielectric environment. This approach effectively captures the essential physics of localized surface plasmon resonances (LSPRs) in nanoparticles, where the electron oscillations are nonpropagating and confined to the nanoparticle's volume.

The particle-in-a-box model offers a straightforward way to estimate the resonance frequencies of LSPRs. For a spherical nanoparticle of radius R, for instance, one can approximate the plasmon modes as standing waves within the sphere. The resonance wavelengths λ_m can then be related to the particle size through a relation of the form $R \approx n\lambda_m/2$, where m is an integer representing the mode number. This leads to a discrete set of resonance frequencies ω_m that depend on the geometry and material parameters. While this model provides a qualitative understanding of the LSPRs, it makes several simplifying assumptions. It often neglects the detailed form of the Coulomb interaction between electrons, the effects of surface polarization, and the influence of the surrounding dielectric medium beyond the immediate vicinity of the nanoparticle [3, 21]. Furthermore, the boundary conditions used in the simple particle-in-a-box model are often idealized and do not perfectly reflect the complex surface interactions present in real nanostructures.

Despite its utility in understanding LSPRs in nanoparticles, the particle-in-a-box model falls short in describing the behavior of plasmons in smaller systems and at higher frequencies, due to the shortcomings of the hydrodynamic model. Furthermore, it fails to describe propagating plasmons, such as those found in thin films, nanowires, or other extended structures. The model's fundamental assumption of strong confinement within a closed boundary does not apply to systems where plasmons can propagate freely. The particle-in-a-box model is therefore not suitable for analyzing the type of 2D (quantum) plasmons that are the focus of this thesis, namely plasmons that propagate in extended inhomogeneous systems for analyzing plasmon scattering and plasmon waveguides. For such systems, a more comprehensive theoretical framework that explicitly accounts for the long-range Coulomb interaction and the specific geometry of the structure is required.

1.2.3 Quantum plasmons from the random phase approximation

The hydrodynamic model, while providing a valuable starting point for understanding plasmons, has limitations. It neglects the changes in the shape of the Fermi surface during plasmon oscillations, leading to inaccuracies in the plasmon dispersion, particularly at shorter wavelengths. This discrepancy arises from the model's assumption of local equilibrium, which fails to hold for plasmon waves due to the high-frequency nature of these oscillations [2, 28]. Additionally, Vlasov's original attempt overlooked the necessity of treating this as an initial value problem, as proposed by Landau, leading to a damping term [26, 27]. In the quantum regime, where the plasmon wavelength is comparable to the Fermi wavelength, a more sophisticated approach is necessary. The quantum description of plasmons was further developed by introducing the concept of the longitudinal dielectric function, which describes the response of the electron to variations in the electron density. This approach led to the development of the random phase approximation, a theoretical framework that allows for the calculation of the plasmon dispersion and other properties in the quantum regime, otherwise known as the time-dependent Hartree approximation [2, 28].

The RPA explicitly accounts for the quantum mechanical nature of electron-electron interactions, providing a more accurate description of the plasmon dispersion across all momenta. It correctly captures the changes in the Fermi surface during plasmon oscillations and can also describe Landau damping [2, 28], which arises from the decay of plasmons into single-particle excitations. Therefore, while the hydrodynamic model and the Vlasov equation offer valuable insights into the classical behavior of plasmons, the RPA is essential for a complete and accurate description of plasmon behavior in the quantum regime, especially at shorter wavelengths. To fully understand quantum plasmons, we employ the RPA to derive the Lindhard dielectric function and the 2D plasmon dispersion in homogeneous systems, and briefly discuss the electron-hole continuum.

The Lindhard dielectric function describes the response of an electron gas to an external perturbation and determines how charges screen external fields. This can also describe the screening of external potentials due to the collective response of the electron system, i.e., plasmons. In this section, we derive this function explicitly by starting from the quantum mechanical evolution of the system based on first principles. Assuming a parabolic energy spectrum for the conduction electrons, we use the equations of motion approach [2] to derive the Lindhard dielectric function.

This approach consists of three steps. First, we consider the electrons, which are confined to two spatial dimensions $\mathbf{x} = (x, y)$. We use the Liouville-von Neumann equation to establish a relation between the one-particle density matrix $\hat{\rho}$ and an external potential $V_{\text{ext}}(\mathbf{x}, t)$. Second, we compute the induced electron density $n(\mathbf{x}, t)$ using this density matrix. Third, the induced electron density induces a potential $V_{\text{ind}}(\mathbf{x}, t)$ through the Poisson equation. Finally, we apply the condition that the total induced potential $V_{\text{tot}}(\mathbf{x}, t)$ should equal the induced potential $V_{\text{ind}}(\mathbf{x}, t)$ and the external potential $V_{\text{ext}}(\mathbf{x}, t)$. Afterwards, we argue that plasmons, self-sustained oscillations, exist when the external potential is equal to zero. This system of equations is fully equivalent to the diagrammatic approach to the RPA, that is, the empty loop approximation for the polarization operator [2, 28, 60].

To derive the plasmon dispersion within the RPA, we first investigate the response of the electron gas to an external potential V_{ext}. This external potential has a response in the electron gas, which in turn creates a total screened Hartree potential V_{tot}. The relationship between the Hartree potential and the external potential is characterized by the electronic response function, also known as the dielectric function ε

$$V_{\text{tot}}(\mathbf{x}, t) = \int d\mathbf{x}' \int dt' \varepsilon^{-1}(\mathbf{x}, \mathbf{x}', t, t') V_{\text{ext}}(\mathbf{x}', t'). \tag{1.16}$$

This nonlocal relationship reflects the fact that the electron density at one point in space and time can be influenced by the potential at other points and times. The response function ε encapsulates the dynamics of the electron gas and its collective behavior. In the following, we derive an expression for the response function within the RPA, taking into account the quantum mechanical nature of the electron-electron interactions. This allows us to determine the plasmon dispersion relation, which describes the collective excitations of the electron gas.

The dynamics of a quantum system are governed by the Liouville–von Neumann equation for the density matrix $\hat{\rho}$, namely [2]

$$i\hbar \frac{\partial \hat{\rho}}{\partial t} = [\hat{H}, \hat{\rho}]. \tag{1.17}$$

The Hamiltonian consists of two terms: the equilibrium Hamiltonian $\hat{H}_0$ describing the free electron gas and a time-dependent perturbation $V_{\text{tot}}(\mathbf{x}, t)$ capturing the nonlocal electron-electron interaction, yielding

$$\hat{H} = \hat{H}_0 + V_{\text{tot}}(\mathbf{x}, t). \tag{1.18}$$

Within the RPA, the electron-electron interaction is captured by a scalar potential $V_{\text{tot}}(\mathbf{x}, t)$, called the Hartree potential or induced potential. This is due to the closed system of local equations for the single-particle density matrix [2, 28]. The unperturbed Hamiltonian is given by

$$\hat{H}_0 = \sum_{\nu} \epsilon_{\nu} c_{\nu}^{\dagger} c_{\nu}, \tag{1.19}$$

where $\epsilon_{\nu} = \hbar^2 \mathbf{p}^2 / 2m$ is the free-electron dispersion relation of state ν, and $c_{\nu}^{\dagger}$ and c_{ν} are the fermionic creation and annihilation operators.

We are interested in periodic perturbations of the form

$$V_{\text{tot}}(\mathbf{x}, t) = V(\mathbf{x}) e^{-i\omega t + \eta t}, \tag{1.20}$$

where $\eta \to 0^+$, ensuring causality. Since the perturbation is weak, we assume that the density matrix takes the form

$$\hat{\rho} = \hat{\rho}_0 + \hat{\rho}_1, \tag{1.21}$$

where $\hat{\rho}_0$ is the equilibrium density matrix satisfying $[\hat{H}_0, \hat{\rho}_0] = 0$, and $\hat{\rho}_1$ is the first-order correction due to V_{tot}. This correction term $\hat{\rho}_1$ represents the deviation of the density matrix from its equilibrium state due to the external perturbation. We assume that $\hat{\rho}_1$ has the same time dependence as V_{tot}, which is reasonable for linear response. This allows us to focus on the steady-state response of the system to the perturbation.

To connect the density matrix to the induced potential, we next solve for the induced density. Substituting the density matrix $\hat{\rho}$ into the Liouville–von Neumann equation and keeping only first-order terms in the perturbation, we obtain

$$i\hbar \frac{d\hat{\rho}_1}{dt} = [\hat{H}_0, \hat{\rho}_1] + [\hat{V}_{\text{tot}}, \hat{\rho}_0], \tag{1.22}$$

where we used the fact that the equilibrium density matrix $\hat{\rho}_0$ commutes with the unperturbed Hamiltonian $\hat{H}_0$, and neglected higher-order terms. Evaluating the commutators using the single-particle eigenstates $|\nu\rangle$ (where we use the conventional Dirac notation [2]), we obtain the equation of motion for the matrix elements $(\rho_1)_{\nu,\nu'}$, which depend on two states, given by

$$i\hbar \frac{d(\rho_1)_{\nu,\nu'}}{dt} = (\epsilon_{\nu} - \epsilon_{\nu'})(\rho_1)_{\nu,\nu'} + (V_{\text{tot}})_{\nu,\nu'}(f(\epsilon_{\nu}) - f(\epsilon_{\nu'})). \tag{1.23}$$

Solving for $(\rho_1)_{\nu,\nu'}$ in steady state gives

$$(\rho_1)_{\nu,\nu'} = \frac{f(\epsilon_\nu) - f(\epsilon_{\nu'})}{\hbar\omega + \epsilon_\nu - \epsilon_{\nu'} + i\eta}(V_{\text{tot}})_{\nu,\nu'}, \tag{1.24}$$

where $\eta \to 0^+$ ensures causality. The total induced electron density can be found by taking the trace of the density matrix, or summing over all states. The induced charge density fluctuation becomes

$$n(\mathbf{x}) = g_{\text{s}} \sum_{\nu,\nu'} \frac{f(\epsilon_\nu) - f(\epsilon_{\nu'})}{\hbar\omega + \epsilon_\nu - \epsilon_{\nu'} + i\eta^+}(V_{\text{tot}})_{\nu,\nu'} \tag{1.25}$$

The factor $g_{\text{s}} = 2$ accounts for spin degeneracy in the 2D electron gas.

The total self-consistent potential, denoted as V_{tot}, comprises both the external perturbation and the self-induced potential arising from the charge response. The relationship between the induced charge density and V_{tot} is governed by the Poisson equation [2]:

$$\nabla^2 V_{\text{ind}}(\mathbf{x}) = \frac{-4\pi e^2 n(\mathbf{x})}{\varepsilon_{\text{b}}}. \tag{1.26}$$

While the induced potential on the left-hand side is formally a three-dimensional quantity, reflecting its spatial extent in all directions, we implicitly account for this by first solving the Poisson equation and then evaluating the result at $z = 0$. This approach effectively restricts the potential to the two-dimensional plane of the electron gas. In Sec. 2.1.1, we explicitly perform these steps to clarify the separation of in-plane and out-of-plane coordinates. The total potential V_{tot} experienced by the electrons is the sum of these two potentials: $V_{\text{tot}} = V_{\text{ext}} + V_{\text{ind}}$. For a homogeneous system, we can consider Eq. (1.26) in momentum space. The external perturbation is a time-dependent potential acting on the electrons, which we express in Fourier space

$$V_{\text{tot}}(\mathbf{q}, \omega) = V_{\text{ext}}(\mathbf{q}, \omega) + v_{\mathbf{q}} n(\mathbf{q}, \omega), \tag{1.27}$$

where

$$V_{\text{tot}}(\mathbf{q}, \omega) = \int d\mathbf{x} \int dt V_{\text{tot}}(\mathbf{x}, t) e^{i(-\mathbf{q}\cdot\mathbf{r}+\omega t)}, \tag{1.28}$$

and $v_{\mathbf{q}}$ and $n(\mathbf{q}, \omega)$ represent the Coulomb interaction and the electron density in Fourier space, respectively.

The dielectric function is defined as the response of the system to external perturbations, and therefore, the ratio between the external and total potentials. We, therefore, write

$$\varepsilon(\mathbf{q}, \omega) = \frac{V_{\text{ext}}(\mathbf{q}, \omega)}{V_{\text{tot}}(\mathbf{q}, \omega)}. \tag{1.29}$$

Using Eq. (1.27), we obtain

$$\varepsilon(\mathbf{q},\omega) = 1 - v_{\mathbf{q}}\Pi_0(\mathbf{q},\omega), \tag{1.30}$$

where $\Pi_0(\mathbf{q},\omega)$ is the noninteracting polarization, also known as the Lindhard function or the single-particle response function, and in the continuous limit, where $\sum_\nu \to \int_{\mathbf{p}}$, is given by

$$\Pi_0(\mathbf{q},\omega) = \frac{g_{\mathrm{s}}}{(2\pi\hbar)^2}\int_{\mathbf{p}} \frac{f(\epsilon_{\mathbf{p}}) - f(\epsilon_{\mathbf{p}+\mathbf{q}})}{\hbar\omega + \epsilon_{\mathbf{p}} - \epsilon_{\mathbf{p}+\mathbf{q}} + i\eta^+} d\mathbf{p}, \tag{1.31}$$

in Fourier space. To avoid confusion with the Lindhard dielectric function, we refer to Π_0 as the polarization in this thesis. This function describes the charge response of the free electron gas and plays a central role in determining plasmon dispersion and screening effects. By systematically applying the Liouville–von Neumann equation, we have derived the longitudinal dielectric function for a noninteracting electron gas. This function is characterized by the polarization $\Pi_0(\mathbf{q},\omega)$, which captures the fundamental screening properties of the system. For the homogeneous case, the eigenstates are plane waves.

In essence, the dielectric function provides a comprehensive description of the electron gas's response to external perturbations, encompassing both the collective excitations (plasmons) and the individual particle excitations (electron-hole pairs). It serves as a fundamental tool for understanding the dynamics and properties of the electron gas in the presence of external fields. Its zeros determine the frequencies and momenta of self-sustained oscillations of the electron density, i.e., the plasmon modes. As seen from Eq. (1.29), when $\varepsilon(\mathbf{q},\omega) = 0$, we can have a non-zero potential V_{tot} in the absence of an external potential V_{ext}, indicating the presence of self-sustained collective oscillations. The imaginary part of the dielectric function, $\mathrm{Im}[\varepsilon(\mathbf{q},\omega)]$, is related to the damping of these oscillations. This damping arises from the excitation of electron-hole pairs, a process known as Landau damping [2, 28]. The electron-hole continuum, where such excitations are possible, corresponds to the region where $\mathrm{Im}[\varepsilon(\mathbf{q},\omega)] \neq 0$. We discuss Landau damping in more detail shortly.

Analytic form of the 2D polarization function and plasmon dispersion

Having established the general framework for calculating the polarization function, we now focus on the specific case of a two-dimensional electron gas at zero temperature. In this section, we review the derivation of the polarization function analytically, which is crucial for understanding collective excitations and damping mechanisms in 2D systems, see e.g. Ref. [28]. This 2D polarization function determines the screening and the plasmon dispersion, both playing a central role in the study of electronic properties in two dimensions.

At zero temperature, the noninteracting polarization function for a 2D electron gas, $\Pi_0(\mathbf{q}, \omega)$, is analytically solvable, since the Fermi-Dirac distribution becomes a step function. To simplify the derivation, we introduce the dimensionless variable

$$\nu_\pm \equiv \frac{E m_{\mathrm{eff}}}{|\mathbf{q}|\, p_{\mathrm{F}}} \pm \frac{|\mathbf{q}|}{2 p_{\mathrm{F}}}, \tag{1.32}$$

where the Fermi momentum p_{F} is given by Eq. (1.15) in the Thomas-Fermi approximation.

The total complex polarization at zero temperature, for a parabolic single electron dispersion in a 2D layer, is now given by [28]

$$\Pi(\mathbf{q}, E) = \frac{g_{\mathrm{s}} m_{\mathrm{eff}}}{2\pi\hbar^2} \frac{p_{\mathrm{F}}}{|\mathbf{q}|} \left(-\frac{|\mathbf{q}|}{p_{\mathrm{F}}} - \operatorname{sign}(\operatorname{Re}\nu_-)\sqrt{\nu_-^2 - 1} + \operatorname{sign}(\operatorname{Re}\nu_+)\sqrt{\nu_+^2 - 1} \right). \tag{1.33}$$

For the existence of plasmons, the dielectric function, given in Eq. (1.30), should be equal to zero, $\varepsilon(\mathbf{q}, E) = 0$. Classically, this corresponds to a self-sustained oscillation of charge density, meaning an external perturbation is no longer necessary to maintain the wave. This condition implies that the system possesses an intrinsic oscillation mode, a collective excitation that can exist without any external driving force.

From the condition $\varepsilon(\mathbf{q}, \omega) = 0$, the plasmon dispersion relation, which relates the plasmon energy $E = \hbar\omega$ to the momentum $\mathbf{q}$, can be calculated. For two-dimensional systems encapsulated in vacuum, the Fourier transform of the Coulomb interaction is given by $v_{\mathbf{q}} = 4\pi e^2/(|\mathbf{q}|\varepsilon_{\mathrm{b}})$ [2, 28]. Then, the analytic form of the 2D plasmon dispersion is given by [28, 32]

$$E(\mathbf{q}) = \sqrt{\left(\frac{|\mathbf{q}|\,\hbar\, p_{\mathrm{F}}^2}{4 m_{\mathrm{eff}}^2 \lambda_{\mathrm{TF}}} \right) \frac{\left(1 + \frac{|\mathbf{q}|^3 \lambda_{\mathrm{TF}}}{\hbar p_{\mathrm{F}}^2} + \frac{|\mathbf{q}|^4 \lambda_{\mathrm{TF}}^2}{\hbar^2 p_{\mathrm{F}}^2}\right)\left(1 + \frac{2|\mathbf{q}|\lambda_{\mathrm{TF}}}{\hbar}\right)^2}{1 + \frac{|\mathbf{q}|\lambda_{\mathrm{TF}}}{\hbar}}} \tag{1.34}$$

where $\lambda_{\mathrm{TF}} = (\hbar^2 \varepsilon_{\mathrm{b}})/(2 g_{\mathrm{s}} m_{\mathrm{eff}} e^2)$ is the Thomas-Fermi screening length [28, 33]. It characterizes the screening of the electric field of an electron by other electrons. The full quantum 2D plasmon dispersion is plotted in Fig. 1.1 in black. The dispersion has a finite energy E_{L}, where the dispersion reaches the Landau damped region in gray, which will be discussed shortly.

In the long-wavelength limit ($|\mathbf{q}| \to 0$), we expand the dispersion around $q = 0$, which simplifies the 2D plasmon dispersion relation to

$$E = \sqrt{\frac{|\mathbf{q}| e^2 g_{\mathrm{s}} p_{\mathrm{F}}^2}{2 m_{\mathrm{eff}} \varepsilon_{\mathrm{b}}} + \frac{3|\mathbf{q}|^2 p_{\mathrm{F}}^2}{2 m_{\mathrm{eff}}^2}}, \tag{1.35}$$

The derived dispersion relation exhibits the characteristic square-root dependence on the momentum, a hallmark of 2D plasmons. This result aligns with the hydrodynamic model's prediction Eq. (1.14), which has the same square-root dispersion in this limit, indicating the well-defined nature in the long-wavelength limit. However, the second-order correction differs from the hydrodynamic model, as argued in Ref. [28]. This discrepancy arises from the hydrodynamic model's assumption of local equilibrium, which neglects dynamic variations in the Fermi surface at higher momenta. Both the hydrodynamic dispersion relation Eq. 1.14 (dashed red) and the long-wavelength quantum limit Eq. 1.35 (dashed black) are plotted in Fig. 1.1, illustrating the growing discrepancies at higher momentum $|\mathbf{q}|$.

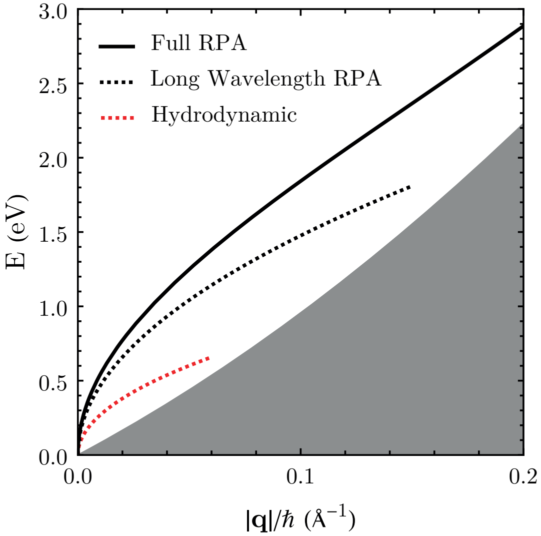


Figure 1.1: Dispersion relations for two-dimensional plasmons, depicting the RPA result (black solid curve), long-wavelength limit (black dashed curve), and hydrodynamic model (red dashed curve). The gray region indicates the Landau damped regime, where single electron-hole pairs can be excited.

Landau damping and the electron-hole continuum

In real systems, plasmons are generally not infinitely long-lived. When the imaginary part of the dielectric function, $\mathrm{Im}[\varepsilon(\mathbf{q}, \omega)]$, becomes non-zero, the collective plasmon mode can lose coherence and decay through single-particle excitations, a process known as Landau damping [2, 28]. Landau damping arises from the interaction of the plasmon with electrons that are moving at velocities close to the phase velocity of the plasmon wave. These electrons can absorb energy from the plasmon, leading to its decay. An electron-hole pair is created when an electron is excited from an occupied state below the

Fermi energy to an unoccupied state above the Fermi energy. The region in energy-momentum space where such excitations are allowed is known as the electron-hole continuum (gray area in Fig. 1.1). Landau damping occurs when the plasmon mode overlaps with this continuum, enabling the transfer of energy from the plasmon to electron-hole pair excitations. Other damping mechanisms, such as magnon [13] or phonon [14] scattering, can also contribute to the decay of plasmons.

The imaginary part of the dielectric function indicates Landau damping. At zero temperature, this occurs when the argument in one of the square roots in Eq. (1.33) becomes negative. The boundary of the electron-hole continuum is therefore given by the condition $\nu_- = 1$. This boundary gives us a minimum value for the energy, namely

$$E_- = \frac{|\mathbf{q}|\, p_\mathrm{F}}{m_\mathrm{eff}} + \frac{|\mathbf{q}|^2}{2m_\mathrm{eff}}, \tag{1.36}$$

and is plotted in Fig. 1.1 as the boundary between the white and gray area. For momenta satisfying the condition $E(\mathbf{q}) = E_-(\mathbf{q}) = E_\mathrm{L}$, or in other words, when the dispersion reaches the electron-hole continuum, a plasmon can decay into electron-hole pairs, leading to a finite imaginary part in the dielectric function. Physically, this means that if the energy of the plasmon is lower than the energy E_L, which is required to excite an electron-hole pair, it remains undamped. However, for higher energies, it enters the electron-hole continuum, and the plasmon can transfer energy to single-particle excitations, leading to damping and energy dissipation.

At finite temperature, the Fermi-Dirac distribution replaces the step function in the zero-temperature polarization function:

$$f(\epsilon) = \frac{1}{e^{(\epsilon-\mu)/k_\mathrm{B}T} + 1}, \tag{1.37}$$

where ϵ is the energy, μ the chemical potential, k_B the Boltzmann constant, and T the temperature. This smoothing of sharp boundaries means Landau damping no longer has a distinct onset at Eq. (1.36), as thermal excitations become probable. Consequently, the plasmon dispersion experiences spectral broadening, and screening properties are altered. While the plasmon peak broadens and becomes less defined with increasing temperature, the essential qualitative features remain. Thus, a practical approach to capturing the essential features is to calculate the plasmon dispersion at zero temperature and then introduce a broadening factor to approximate thermal effects.

In summary, for a 2D electron gas, the polarization function is central to understanding screening and plasmon dispersion. In homogeneous systems, the full quantum plasmon dispersion can be analytically derived, providing a comprehensive description beyond the long-wavelength limit. This dispersion is gapless, meaning that plasmons can exist at arbitrarily low energies. Landau damping occurs when the plasmon loses coherence to single electron-hole excitations, causing energy dissipation [2, 28]. At finite temperature, screening changes, the continuum smears, and modes broaden. In this thesis, we concentrate on undamped plasmonic modes, which allow for clearer semiclassical analysis, with careful attention to the Landau damping onset. While the homogeneous 2DEG is well understood, our focus lies on inhomogeneous systems. Here, spatial variations necessitate numerical methods in real space, as translational invariance is lost. To address these complexities and gain insights into inhomogeneous plasmons, we utilize a semiclassical approach, deriving (semi-)analytical solutions through semiclassical techniques.

1.3 Semiclassical analysis for inhomogeneous systems

Inhomogeneous systems, unlike their homogeneous counterparts, lack translational invariance. This makes them more challenging to analyze but also more interesting, as they offer the necessary complexity for practical applications in plasmonics. To describe plasmons in inhomogeneous systems, we propose using the semiclassical approximation, akin to the WKB approximation for solving differential equations like the Schrödinger equation in smoothly varying inhomogeneous potentials in one direction. This approximation is valid when the typical length scale of the potential variation is much larger than the wavelength of the underlying particle, which allows us to treat the system as locally homogeneous [61].

The WKB approximation was among the initial semiclassical techniques employed to study plasmons in inhomogeneous electron systems. Notably, the semiclassical approximation for quantum plasmons was initially proposed in Ref. [51] and further justified at a heuristic level in Ref. [62]. Progressing beyond these early applications, recent investigations have leveraged pseudodifferential operators and the Maslov canonical operator [48] to establish a more robust theoretical foundation for three-dimensional electron systems [50]. While this advanced methodology offers a significant advancement in the understanding of plasmon dynamics within inhomogeneous environments, further development is required to accurately model the distinctive characteristics observed in two-dimensional systems.

Physically, the semiclassical analysis can be understood in analogy to geometric optics. The Hamilton-Jacobi equation describes the trajectories of plasmons, which can be calculated using Hamilton's equations, allowing us to trace the plasmon motion point by point, similar to how light rays are traced in geometric optics. On top of these classical trajectories, the semiclassical approximation incorporates the wave-like nature of plasmons, capturing interference effects. Semiclassical methods are formally applicable to systems with smoothly varying potentials, where the characteristic length scale of the potential variation is much larger than the relevant wavelength. However, the semiclassical approximate solution often remains qualitatively accurate in regimes beyond its formally defined applicability, extending its usefulness to a wider range of physical scenarios. For the Schödinger equation one can, e.g., consider the harmonic oscillator and the hydrogen atom, for which the semiclassical result coincide with the exact result [61, 63].

For plasmons, the semiclassical approximation is valid when the electron wavelength is much smaller than the characteristic length scale of the system. This ensures that the electrons experience a locally homogeneous potential, which is crucial for the validity of the approximation. In this regime, we can define a dimensionless small parameter h as the ratio of the electron wavelength to the system's characteristic length scale. This characteristic length scale can arise from spatial variations in, for example, the electron density or the substrate dielectric constant. It is important to emphasize that in our context, the semiclassical approximation does not serve as a bridge between classical and quantum plasmons. Rather, in the limit of $h \to 0$, we recover the homogeneous case described by the RPA polarization Eq. (1.31), and not the classical limit from the hydrodynamic model.

In this section, we first provide a theoretical introduction to semiclassics, focusing on the WKB approximation applied to the Schrödinger equation, which introduces essential concepts and techniques that will be applied to plasmons. Then, we apply these semiclassical techniques to electrons in 2D inhomogeneous systems, following the approach outlined by Ref. [50], to derive an analytical expression for the polarization with spatially varying parameters.

1.3.1 The WKB approximation

The semiclassical approximation is a powerful tool in quantum mechanics that bridges classical and quantum techniques for describing quantum systems. A fundamental example of semiclassics is the WKB approximation, which, in the case of the Schrödinger equation, provides an approximate solution in the regime where the characteristic length scale of the system varies slowly compared to the de Broglie wavelength. This method is particularly useful for understanding tunneling phenomena, bound states in slowly varying potentials, and quantization conditions in one-dimensional systems. Here, we give a simplified, physical explanation of the WKB approximation, highlighting the essential concepts for this thesis, in line with Ref. [61]. For a more in-depth discussion of semiclassical analysis, we refer to books like Refs. [48, 49, 63, 64].

Assuming a smoothly varying potential $V(x)$, for which the WKB approximation is applicable, we begin with the time-independent Schrödinger equation for a single particle, namely

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x), \tag{1.38}$$

where $\psi(x)$ is the wavefunction.

Defining the local wavenumber as

$$k(x) = \frac{\sqrt{2m(E - V(x))}}{\hbar}, \tag{1.39}$$

the equation simplifies to

$$\frac{d^2\psi(x)}{dx^2} + k^2(x)\psi(x) = 0. \tag{1.40}$$

We then seek a WKB solution of the form of the Ansatz

$$\psi(x) = (A_0(x) + \hbar A_1(x) + \mathcal{O}(\hbar^2))e^{iS(x)/\hbar}, \tag{1.41}$$

where $S(x)$ is the classical action and $A_i(x)$ the expansion of the amplitude, which is a series in powers of $\hbar$. The WKB approximation is fundamentally based on an expansion in a small dimensionless parameter h, which relates the variations in the potential to the de Broglie wavelength of the particle. While the specific definition of h for plasmon applications will be discussed in detail in Sec. 2.1.5, for this introductory overview, we proceed by expanding in powers of $\hbar$. Substituting this expansion into the Schrödinger equation and solving order by order in $\hbar$ leads to equations for the coefficients $A_i(x)$. The zeroth-order equation is the Hamilton-Jacobi equation, describing the classical motion, resulting in

$$\left(\frac{dS}{dx}\right)^2 - k^2(x) = 0. \tag{1.42}$$

Equivalently, we write the Hamilton-Jacobi equation as $\mathcal{H}_{\text{WKB}}(x, dS/dx, E) = 0$, which defines an effective classical Hamiltonian. Note that when deriving the Hamilton-Jacobi equation, the wavenumber transforms to $|k| = |dS/dx|$; a natural consequence of applying the semiclassical Ansatz. Using Hamilton's equations applied to $\mathcal{H}_{\text{WKB}}(x, k, E)$, we can calculate classical trajectories with Hamilton's equation, analogous to geometric optics.

The first-order equation is the transport equation, governing the amplitude $A_0(x)$, given by

$$\frac{d}{dx}\left(A_0^2(x)\frac{dS}{dx}\right) = 0. \tag{1.43}$$

This ensures the conservation of probability density along classical trajectories. The amplitude derived from the transport equation, through standard derivation, yields a solution proportional to the inverse square root of the Jacobian, $J(x)$ [48, 55]. This amplitude is evaluated along the trajectories in phase space, and the Jacobian ensures that the wavefunction maintains its correct form and normalization under coordinate transformations, reflecting the conservation of probability. However, in this one-dimensional WKB case, this greatly simplifies to $J(x) = k(x)$ with the help of Hamilton's equations. This specific simplification is not possible for plasmons, as is discussed in greater detail in Sec. 2.1.3. The approximate solution, Eq. (1.41), with amplitude A_0 governed by the transport equation, incorporates the wave-like behavior, including quantum interference effects. In classically allowed regions, where $E > V(x)$ and $k(x)$ is real, the solution takes the form

$$\psi(x) \approx \frac{A_0^0}{\sqrt{k(x)}} e^{\pm i \int^x k(x')dx'}, \tag{1.44}$$

where A_0^0, is an initial amplitude factor. The probability of finding the particle at a given point x is given by $|\psi(x)|^2 \propto |k(x)|^{-1}$, which reflects the fact that it is less probable to find the particle in regions where it has a higher velocity. In classically forbidden regions, where $E < V(x)$ and $k(x)$ is imaginary, the wavefunction becomes

$$\psi(x) \approx \frac{A_0^0}{\sqrt{|k(x)|}} e^{\pm \int^x |k(x')|dx'}. \tag{1.45}$$

This exponential behavior in the forbidden region describes quantum mechanical tunneling, a key prediction of the WKB approximation.

A crucial application of the WKB method is the Bohr-Sommerfeld quantization condition for periodic trajectories, which determines the allowed energy levels of bound states. For a potential $V(x)$ exhibiting parabolic behavior, the particle's motion resembles that of a particle in a harmonic potential well. When the particle encounters two simple turning points, the wavenumber behaves as $k(x) \propto \sqrt{x}$ near these points and goes to zero at the turning point. The quantization condition for the action then reads

$$\oint k(x)dx = \left(m + \frac{1}{2}\right) 2\pi, \tag{1.46}$$

where the integral is taken over one full periodic trajectory and m is a non-negative integer. This condition establishes a relationship between the energy levels and the integer m. The factor $\pi/2$ in Eq. (1.46) accounts for the phase shift of the solution (1.44) at a simple turning, and can be formalized through the Maslov index [48, 50, 55]. This phase shift arises because the asymptotic solution (1.44) breaks down at a turning point, as the Jacobian vanishes. From a practical perspective, the Maslov index ensures the correct phase evolution as the particle traverses a turning point, compensating for the sign change in the Jacobian within the amplitude. Such a quantization condition, incorporating the Maslov index, will be applied in Ch. 4 of this thesis to determine the bound states in a plasmonic waveguide.

One major limitation of the WKB approximation is its failure at classical turning points, where $E = V(x)$ and $k(x) = 0$. At these points, the amplitude of the WKB solution diverges, indicating a breakdown of the approximation. Physically, this occurs because the classical particle momentarily comes to rest before reversing direction, and the wavelength of the quantum wavefunction changes rapidly in this region. While a more careful treatment using Airy functions, $\mathrm{Ai}(x)$ and $\mathrm{Bi}(x)$, can provide intermediate solutions applicable at the turning point, in most cases, one is not particularly interested in the specific solution at these points, but rather in the general behavior around them. This general behavior can be addressed in two specific scenarios. First, the behavior of a backscattered particle after a turning point is captured by the Maslov index, as discussed previously. Second, one can connect the WKB solutions before and after the turning point, a procedure known as the connection problem. For analyzing behavior in regions beyond the turning point, the complex phase integral method is a suitable alternative, as it circumvents the turning point's direct influence [63, 65]. This method, pioneered by Zwaan [66], involves taking a path through the complex plane to connect the solutions, carefully tracking their behavior along so-called Stokes and anti-Stokes lines. Connection formulas can be used to directly relate the WKB solutions on either side of the turning point. This complex phase integral method will be applied in Ch. 5 of this thesis.

The WKB approximation is significant because it connects quantum mechanics with techniques from classical mechanics in a mathematically transparent way. It provides insight into wave propagation, tunneling, and energy quantization in systems where exact solutions to the Schrödinger equation are difficult to obtain. In essence, this approach allows us to determine classical trajectories with Hamilton's equations, using the effective classical Hamiltonian $\mathcal{H}_{\mathrm{WKB}}\left(x, k, E\right)$. The WKB method incorporates the wave-like solutions in the form of $\psi(x) = A_0(x)\exp(iS(x)/\hbar)$, where $S(x)$ is the classical action satisfying the Hamilton-Jacobi equation, $\mathcal{H}_{\mathrm{WKB}}\left(x, dS/dx, E\right) = 0$, while the amplitude $A_0(x)$ is calculated from the transport equation. Using techniques like the Maslov index and the complex phase integral method, this approximation can be further refined to accurately describe quantum phenomena in experimentally relevant systems.

1.3.2 Derivation of the formalism for plasmons

In this section, we aim to develop a description for electron behavior in inhomogeneous systems, specifically focusing on the derivation of a semiclassical formalism for a two-dimensional inhomogeneous electron gas. Building upon the equations of motion approach, as utilized in Ref. [2] and discussed in Sec. 1.2.3 for a homogeneous system, we derive expressions for the polarization, which in the homogeneous case is given by Eq. (1.31), and the induced electron density in inhomogeneous systems. To account for the spatial variations arising from the inhomogeneity of the system, we employ the formalism of pseudodifferential operators [48, 49, 67]. This approach was used in Ref. [50] to derive the full theory for plasmons in three-dimensional systems. Here, we establish the foundation for our derivation of the full theory for plasmons in 2D inhomogeneous systems by reviewing the procedure based on the results from Ref. [50]. We show that the initial steps of obtaining the induced density from the polarization are generally applicable and independent of dimensionality.

We use the Liouville-von Neumann equation to establish a relation between the one-particle density operator $\hat{\rho}$ and the semiclassical Ansatz for the induced potential $V_{\mathrm{pl}}(\mathbf{x}, t)$ in the plane. Subsequently, we compute the induced electron density $n(\mathbf{x}, t)$ using this density operator. The final steps, which involve relating the time-dependent induced electron density to the three-dimensional potential $V(\mathbf{x}, z, t)$ through the Poisson equation, are deferred to Ch. 2, as these represent novel contributions. For this step, we consider the electric field, a three-dimensional quantity that extends into the out-of-plane dimension z. We apply a self-consistency condition: the potential $V(\mathbf{x}, z, t)$ should equal the in-plane induced potential $V_{\mathrm{pl}}(\mathbf{x}, t)$ at $z = 0$. We remark that this condition is not required for a three-dimensional charge density [50], where a self-consistent solution for the induced potential is obtained directly. In this section, we lay the groundwork for this new derivation by reviewing the initial two steps of the procedure, based on the results from Ref. [50].

1.3.3 Density operator and induced electron density

Moving beyond the homogeneous case, we now specifically apply the equations of motion approach [2, 50] to plasmons in inhomogeneous systems. Consequently, we denote the total potential as the induced potential or Hartree potential $V_{\mathrm{pl}}(\mathbf{x}, t)$, assuming the external potential to be zero, as discussed in Sec. 1.2.3. To ensure self-containment in this derivation, some steps from the homogeneous derivation are repeated. We write the Hamiltonian as

$$\hat{H} = \hat{H}_0 + V_{\mathrm{pl}}(\mathbf{x}, t), \tag{1.47}$$

where the operator $\hat{H}_0$ describes the motion of the individual electrons and $V_{\mathrm{pl}}(\mathbf{x}, t)$ is a scalar potential that expresses the electron-electron interaction within the system.

We are interested in inhomogeneous systems, which necessitate modifications to the Hamiltonian for individual electrons. Throughout this thesis, we consider electrons with a quadratic dispersion moving within a spatially varying scalar potential $U(\mathbf{x})$, as described by

$$\hat{H}_0 = \frac{\hat{\mathbf{p}}^2}{2m} + U(\mathbf{x}), \qquad \hat{p}_x = -i\hbar\frac{\partial}{\partial x}, \tag{1.48}$$

where m is the effective electron mass of the system. It is important to note that while we focus on quadratic dispersion, the semiclassical framework developed here is applicable to other electronic dispersion relations, albeit with additional considerations. The length of the vectors $\mathbf{x}$ and $\hat{\mathbf{p}}$, depend on the dimensionality of the electron gas, i.e. for a two-dimensional electron gas $\mathbf{x} = (x, y)$. The potential $U(\mathbf{x})$ encodes the spatially varying electron density $n^{(0)}(\mathbf{x})$. Within the Thomas-Fermi approximation [2, 28, 58], the position-dependent Fermi momentum $p_\mathrm{F}(\mathbf{x})$ and the spatially varying scalar potential $U(\mathbf{x})$ are related by

$$p_\mathrm{F}(\mathbf{x}) = \hbar \left(\frac{4\pi}{g_\mathrm{s}} n^{(0)}(\mathbf{x}) \right)^{1/2}, \qquad U(\mathbf{x}) = \mu - \frac{p_\mathrm{F}^2(\mathbf{x})}{2m}, \tag{1.49}$$

analogous to the homogeneous case described by Eq. (1.15), this equation uses $p_\mathrm{F}(\mathbf{x})$ to represent the position-dependent Fermi momentum, μ for the chemical potential of the system, and g_s for the spin degeneracy.

When the system is homogeneous, we can decompose this Hartree potential V_pl into Fourier modes. In our case, where the system is “almost” homogeneous on the scale of the electron wavelength, since $\lambda_\mathrm{el}/\ell \ll 1$, we can use the Ansatz that is commonly used in the semiclassical approximation [48, 55], namely

$$V_\mathrm{pl}(\mathbf{x}, t) = V_\mathrm{pl}(\mathbf{x}) e^{-iEt/\hbar}, \tag{1.50}$$

where we assume that the Hamiltonian $\hat{H}_0$ does not depend on time t, which indicates that the potential does not vary in time. Since we compute the retarded response function of the electrons, it would be more correct to write $E + i\eta$ and consider the limit $\eta \to 0^+$. However, we implicitly assume this throughout the thesis.

In the remainder of this subsection, we summarize the main points of the derivation performed in Ref. [50], to which we refer for more detailed arguments. The motion of the system of interacting electrons is described by the Liouville-von Neumann equation for the density operator, i.e.

$$i\hbar \frac{\partial \hat{\rho}}{\partial t} = [\hat{H}, \hat{\rho}]. \tag{1.51}$$

We can decompose the density operator as $\hat{\rho} = \hat{\rho}_0 + \hat{\rho}_1$, where $\hat{\rho}_0$ corresponds to the equilibrium situation originating from the Hamiltonian $\hat{H}_0$. The perturbation $\hat{\rho}_1$ to this equilibrium is caused by the electron-electron interaction, which implies that $\hat{\rho}_1$ should be of the same order of magnitude as V_pl. Moreover, $\hat{\rho}_1$ should have the same time dependence as in Eq. (1.50).

Since V_pl is assumed to be small, we can linearize the Liouville-von Neumann equation (1.51). The operator $\hat{\rho}_0$ is time independent, because it corresponds to the equilibrium. The zeroth-order terms therefore give $[\hat{H}_0, \hat{\rho}_0] = 0$. The terms that are first order in V_pl lead to

$$E\hat{\rho}_1 = [\hat{H}_0, \hat{\rho}_1] + [V_\mathrm{pl}, \hat{\rho}_0], \tag{1.52}$$

where we have taken the time dependence of $\hat{\rho}_1$ into account.

At this point, we apply the semiclassical approximation. We construct an asymptotic solution for $V_{\mathrm{pl}}(\mathbf{x})$ in the form of the semiclassical Ansatz

$$V_{\mathrm{pl}}(\mathbf{x}) = \varphi(\mathbf{x}, \hbar) e^{iS(\mathbf{x})/\hbar}, \tag{1.53}$$

where $S(\mathbf{x})$ is the classical action and $\varphi(\mathbf{x}, \hbar)$ is the amplitude. The latter is a series in powers of $\hbar$, that is,

$$\varphi(\mathbf{x}, \hbar) = \varphi_0(\mathbf{x}) + \hbar \varphi_1(\mathbf{x}) + \mathcal{O}(\hbar^2). \tag{1.54}$$

When this Ansatz is applied to one-dimensional problems in physics, one usually speaks of the WKB approximation [61, 63], as discussed in the previous subsection Sec. 1.3.1.

Unfortunately, we cannot use the same semiclassical approach for the linearized Liouville-von Neumann equation (1.52), since it is an operator equation rather than a standard eigenvalue problem as the Schrödinger equation. We therefore need a more advanced toolbox to derive an effective classical Hamiltonian, or, more fundamentally, to relate quantum operators on Hilbert space to classical observables on phase space. Very naively, one could think of quantum mechanical operators as functions of $\mathbf{x}$ and $\hat{\mathbf{p}}$ and replace all momentum operators $\hat{\mathbf{p}}$ by variables $\mathbf{p}$ to obtain classical observables on phase space. Whilst this approach gives the correct results to the lowest order in $\hbar$, it hardly seems like a well-defined mathematical procedure and gives wrong results for the higher-order corrections.

The most elegant way to express the relation between quantum mechanical operators on Hilbert space and functions on classical phase space $(\mathbf{x}, \mathbf{p})$ is through the formalism of pseudodifferential operators [48, 49, 67]. Within so-called standard quantization, one obtains a function $\sigma(\hat{a}) = a(\mathbf{x}, \mathbf{p}, \hbar)$ on classical phase space from a quantum operator $\hat{a}$ with the formula

$$a(\mathbf{x}, \mathbf{p}, \hbar) = \sigma\left(\hat{a}\right) = e^{-i\langle \mathbf{p}, \mathbf{x}\rangle/\hbar} \left(\hat{a} e^{i\langle \mathbf{p}, \mathbf{x}\rangle/\hbar}\right), \tag{1.55}$$

where $\langle \mathbf{p}, \mathbf{x}\rangle = \sum_j p_j x_j$ is the standard inner product on $\mathbb{R}^2$. The function $a(\mathbf{x}, \mathbf{p}, \hbar)$ is commonly called a symbol. As an example, we may apply this formula to $\hat{H}_0$ in Eq. (1.48). We then find that $\sigma(\hat{H}_0) = H_0(\mathbf{x}, \mathbf{p}) = \mathbf{p}^2/2m + U(\mathbf{x})$, which is exactly what one obtains when replacing $\hat{\mathbf{p}}$ by $\mathbf{p}$ in $\hat{H}_0$.

Generalizing this previous example, we note that most of the symbols that we consider in this thesis are so-called classical symbols. These are symbols $a(\mathbf{x}, \mathbf{p}, \hbar)$ that have an asymptotic expansion in terms of $\hbar$ [49], i.e.,

$$a(\mathbf{x}, \mathbf{p}, \hbar) = a_0(\mathbf{x}, \mathbf{p}) + \hbar a_1(\mathbf{x}, \mathbf{p}) + \mathcal{O}(\hbar^2). \tag{1.56}$$

One can think of the leading-order term $a_0(\mathbf{x}, \mathbf{p})$ as the classical observable (on phase space) corresponding to the quantum operator $\hat{a}$. If one replaces the momentum operators $\hat{\mathbf{p}}$ by coordinates $\mathbf{p}$ in the operator $\hat{a}$, one precisely obtains a_0, which formalizes the very naive procedure that we sketched before.

Within standard quantization, Hermitian operators may have complex symbols. For instance, the operator $\frac{1}{2}(\mathbf{x}\cdot\hat{\mathbf{p}}+\hat{\mathbf{p}}\cdot\mathbf{x})$ has symbol $\mathbf{x}\cdot\mathbf{p}-i\hbar$. One can show that the symbol of a Hermitian operator satisfies the relation

$$a_1(\mathbf{x},\mathbf{p}) = -\frac{i}{2}\sum_j \frac{\partial a_0}{\partial p_j \partial x_j}(\mathbf{x},\mathbf{p}), \tag{1.57}$$

see e.g. Ref. [50]. These complex symbols should therefore be seen as an artifact of the procedure, and by no means imply that we are dealing with non-Hermitian operators. In order to avoid these complex symbols, one may use Weyl quantization [49, 67], in which the relation between operators and their symbols differs from Eq. (1.55) and Hermitian operators correspond to real symbols. Although both quantization schemes are formally equivalent, Weyl quantization typically makes the calculations more complicated. We therefore use standard quantization throughout this text.

Starting from a symbol $a(\mathbf{x},\mathbf{p},\hbar)$, one obtains the corresponding pseudodifferential operator $\hat{a}$, within standard quantization, with the Fourier transform $\mathcal{F}$, namely [48, 49, 67]

$$\left(\hat{a}\, f\right)(\mathbf{x}) = \mathcal{F}^{-1}_{\mathbf{p}\to\mathbf{x}}\, a(\mathbf{x},\mathbf{p},\hbar)\mathcal{F}_{\mathbf{y}\to\mathbf{p}} f(\mathbf{y}). \tag{1.58}$$

The operator constructed in this way corresponds to the situation where the momentum operator always acts first, and the position operator acts second. For instance, quantization of $\mathbf{x}\cdot\mathbf{p}$ with this procedure gives $\mathbf{x}\cdot\hat{\mathbf{p}}$, which is not Hermitian. This is in accordance with our previous statement that a symbol should satisfy Eq. (1.57) to give rise to a Hermitian operator. Equations (1.55) and (1.58) establish a one-to-one relation between operators and symbols [49, 67]. For a general discussion of pseudodifferential operators and their symbols, we refer to Refs. [48, 49, 67]. A short overview in the context of this thesis can be found in Ref. [50].

Let us return to the linearized Liouville-von Neumann equation (1.52). In order to apply the semiclassical approximation to this operator equation, one also needs to employ a semiclassical Ansatz for the induced density operator $\hat{\rho}_1$. This Ansatz is constructed in detail in Ref. [50], and its symbol is expressed in terms of the amplitude $\varphi(\mathbf{x},\hbar)$ and the classical action $S(\mathbf{x})$, cf. Eq. (1.53), by solving the operator equation (1.52). The result for $\hat{\rho}_1$ is subsequently used to compute the induced electron density, defined by

$$n(\mathbf{x}) = g_{\mathrm{s}}\mathrm{Tr}(\delta(\mathbf{x}'-\mathbf{x})\hat{\rho}_1). \tag{1.59}$$

After some lengthy calculations, that are performed explicitly in Ref. [50], one finds that the induced density $n(\mathbf{x})$ can be written as

$$n(\mathbf{x}) = \hat{\Pi} V_{\mathrm{pl}}(\mathbf{x}), \tag{1.60}$$

where $V_{\mathrm{pl}}(\mathbf{x})$ is given by the semiclassical Ansatz (1.53). Where, in contrast to the homogeneous case (cf. Sec.1.2.3), the polarization $\hat{\Pi}$ is a pseudodifferential operator. These pseudodifferential operators can be viewed as generalizations of partial differential operators and are defined through their so-called symbols [49, 67], which are functions on classical phase space. Intuitively speaking, this relation can be understood as the correspondence between quantum mechanical operators and classical observables on phase space [68]. Importantly, the symbols of pseudodifferential operators do not have to be polynomial in 4Its symbol has an expansion in powers of $\hbar$, cf. Eq. (1.56), i.e.,

$$\sigma(\hat{\Pi}) = \Pi(\mathbf{x}, \mathbf{q}, \hbar) = \Pi_0(\mathbf{x}, \mathbf{q}) + \hbar \Pi_1(\mathbf{x}, \mathbf{q}) + \mathcal{O}(\hbar^2), \tag{1.61}$$

where the principal symbol $\Pi_0(\mathbf{x}, \mathbf{q})$ depends on the position $\mathbf{x}$ and the (plasmon) momentum $\mathbf{q}$. It equals

$$\Pi_0(\mathbf{x}, \mathbf{q}) = \frac{g_\mathrm{s}}{(2\pi\hbar)^d} \int \frac{\rho_0\left(H_0(\mathbf{x}, \mathbf{p})\right) - \rho_0\left(H_0(\mathbf{x}, \mathbf{p}+\mathbf{q})\right)}{H_0(\mathbf{x}, \mathbf{p}) - H_0(\mathbf{x}, \mathbf{p}+\mathbf{q}) + E} d\mathbf{p}, \tag{1.62}$$

where the function $\rho_0(z)$ is the Fermi-Dirac distribution. Its argument $H_0(\mathbf{x}, \mathbf{p})$ is the principal symbol of the Hamiltonian $\hat{H}_0$, which we constructed before.

Expression (1.62) bears a strong resemblance to the familiar Lindhard expression for a homogeneous charge density Eq. (1.31) [2, 28], with the key difference being the replacement of the energy eigenvalue $\epsilon_\mathbf{p}$ by the symbol $H_0(\mathbf{x}, \mathbf{p})$ of the electronic Hamiltonian. Consequently, the integral can therefore be evaluated in the same way as in Sec. 1.2.3, see e.g. Ref. [28]. Performing this evaluation, one obtains the Lindhard expression with the replacement $p_\mathrm{F} \to p_\mathrm{F}(\mathbf{x})$, see also the discussion in Ref. [50]. The principal symbol $\Pi_0(\mathbf{x}, \mathbf{q})$ thus represents a local polarization, dependent on the spatial coordinate $\mathbf{x}$ and the plasmon momentum $\mathbf{q}$. Notably, due to the isotropy of the Hamiltonian $\hat{H}_0$ in momentum space, it depends only on the magnitude $|\mathbf{q}|$. The subprincipal symbol $\Pi_1(\mathbf{x}, \mathbf{q})$ satisfies Eq. (1.57), confirming the Hermitian nature of the polarization operator. This holds true in regions devoid of Landau damping, where the principal symbol Π_0 is real-valued. Incorporating the standard time dependence $\exp(-iEt/\hbar)$ into $n(\mathbf{x})$, we arrive at an expression for the induced electron density $n(\mathbf{x}, t)$. While the form of the resulting expression (1.60) is similar across different spatial dimensions, the normalization factor $g_\mathrm{s}/(2\pi\hbar)^d$ in the polarization depends explicitly on the dimensionality d of the electronic system. Therefore, the result is not strictly dimension-independent. Nevertheless, the underlying semiclassical approach and the steps leading to the induced density are applicable to both two and three spatial dimensions. However, throughout this thesis, we exclusively focus on the two-dimensional case ($d = 2$).

As a result, we have obtained the position-dependent polarization, which describes the single-electron response to the collective electron interactions. In contrast to the homogeneous case discussed in Sec. 1.2.3, where a direct Fourier transform of the Poisson equation is feasible, the inhomogeneous case presents a more complex challenge. In Ch. 2, we apply the semiclassical approximation to the Poisson equation, utilizing the electron density given by Eq. (1.60) as the source term. Given that the Poisson equation is inherently three-dimensional, with the electric potential extending into the out-of-plane dimension, we must carefully separate the in-plane and out-of-plane degrees of freedom. Subsequently, by analyzing the first-order and second-order components in the semiclassical parameter h, we derive the Hamilton-Jacobi equation and the transport equation, which collectively govern the plasmonic system.

In Ch. 3, we consider plasmon scattering, a fundamental process in wave physics involving the redirection of waves due to interactions with inhomogeneities. This is particularly interesting in plasmonics, where inhomogeneities such as impurities, local doping, and density variations significantly impact plasmon propagation and are crucial for understanding and controlling realistic plasmonic behavior in nanostructures observed in experiments. By calculating the classical trajectories using our semiclassical approach, we can gain an initial qualitative understanding of the plasmon scattering behavior. We then quantify this by calculating the total and differential scattering cross sections, measures of the effective scattering area, utilizing the wave-like properties of the approximate solution. Specifically, we focus on a simple circular geometry, exploring how the scattering cross section is influenced by variations in electron density and the width of such inhomogeneities.

In Ch. 4, we extend our analysis to investigate plasmonic localization within layered systems, specifically focusing on plasmonic waveguides. Plasmonic localization, the confinement of plasmons to a specific region, is critical for integrated photonics applications, enabling efficient information transport and manipulation through plasmons. This localization can be achieved through various mechanisms, which we explore in detail. We identify two primary types of localization: full localization through the formation of bound states, and quasi-localization via amplitude screening effects. Full localization arises from periodic trajectories within classically allowed and forbidden regions, allowing us to calculate the discrete energy spectrum using the Bohr-Sommerfeld quantization condition, as discussed in Sec. 1.3.1. Additionally, we examine quasi-localization, where plasmon confinement is achieved through amplitude screening. We compare and validate our semiclassical results with numerical findings in the literature.

In Ch. 5, we extend our investigation to plasmonic crystals, which are periodic analogs of waveguides, exhibiting plasmon tunneling. These structures are of significant interest as they enable the creation of plasmonic band structures, akin to photonic and electronic band structures, offering unprecedented control over plasmon propagation. Specifically, we employ connection formulas derived from the Zwaan method [63, 66] to incorporate tunneling effects within these periodic structures. This allows us to accurately calculate the plasmonic band structure, revealing the allowed and forbidden energy bands for plasmon propagation within these complex environments.

2

Plasmons in arbitrary dielectric environment

In this chapter, we continue the semiclassical analysis from Sec. 1.3.2 to investigate plasmons within inhomogeneous systems characterized by arbitrary dielectric environments. Building upon the induced density Eq. (1.60) derived in the preceding chapter, we employ Poisson's equation to establish the relationship with the induced potential. Our primary objective is to develop a comprehensive theoretical framework capable of describing plasmon behavior in such complex scenarios. This involves deriving an effective classical Hamiltonian, analogous to the Lindhard dielectric function but incorporating spatially varying parameters, to extract the plasmon dispersion and the classical action via the Hamilton-Jacobi equation. Furthermore, we determine the amplitude of the semiclassical Ansatz using the transport equation, which can be interpreted physically through the energy density. To demonstrate the versatility of our methodology, we apply it to two distinct models of the out-of-plane dielectric environment, which serves as foundational examples for subsequent chapters.

This chapter draws upon work published in Refs. [1] and [69]. Copyright by the American Physical Society.

2.1 Effective description of plasmons in 2D systems with arbitrary dielectric environments

In Ref. [1], the construction of the effective classical Hamiltonian $\mathcal{H}_0$ was performed using the simplest model for the dielectric environment, in which the background dielectric environment does not depend on z. This leads to an effective dielectric function in the plasmon dispersion, which arises due to screening effects from the surrounding environment. In this model, the function is given by the average of the dielectric properties of the involved media. However, Ref. [69] argues that more complicated models for the dielectric environment are often necessary to more accurately describe the screening effects. In homogeneous systems, such advanced models typically yield more accurate predictions of three-dimensional screening effects, as they account for nonlocal interactions. Consequently, the effective dielectric function, describing this screening, is in these cases no longer a simple averaged quantity, but instead exhibits a $\mathbf{q}$-dependence [34, 40, 70]. We would therefore like to extend the formalism developed in Ref. [1] to arbitrary dielectric environments in order to be applicable to more complicated models.

Throughout this thesis, we describe the dielectric properties of the substrate using the term dielectric constant. While we assume the substrate material is locally characterized by a fixed static dielectric constant, justifying the term “constant”, this value can still exhibit spatial variations as a function of $\mathbf{x}$. This, for example, accounts for material transitions in systems with multiple materials.

The derivation in the previous chapter concluded with an expression for the induced electron density, Eq. (1.60), which we now use to compute the induced potential via the Poisson equation. A fundamental distinction arises when applying this approach to two-dimensional versus three-dimensional systems. In three-dimensional scenarios, both the charge density and electrostatic potential are inherently three-dimensional. However, in two-dimensional systems, while the electrostatic potential remains three-dimensional, the charge density is effectively confined to two dimensions. The induced potential, as dictated by Poisson’s equation, extends into the third dimension. Thus, in Sec. 2.1.1, we consider the electric field, a three-dimensional quantity that permeates the out-of-plane dimension z, and establish a relationship between the time-dependent induced electron density and the three-dimensional potential $V(\mathbf{x}, z, t)$. Our semiclassical derivation employs a separation of variables into slow and fast components. We posit that the in-plane variables are slow, implying that the spatial scale of variations in the charge density and background dielectric constant greatly exceeds the electron wavelength. Conversely, the out-of-plane variable is fast, reflecting the abrupt change in charge density at the charge layer boundary. This separation enables us to asymptotically decouple the in-plane and out-of-plane degrees of freedom through a generalization of the adiabatic Born-Oppenheimer approximation. Specifically, we adapt the operator separation of variables method [56, 71] prior to introducing the semiclassical Ansatz. At the conclusion of this section, we obtain two ordinary differential equations in z derived from the Poisson equation: one of order $\hbar^0$ and another of order $\hbar^1$.

In the subsequent section, Sec. 2.1.2, we derive the effective classical Hamiltonian $\mathcal{H}_0$ for an arbitrary dielectric environment $\varepsilon(\mathbf{x}, z)$ by solving the $\hbar^0$-order out-of-plane differential equation resulting from the Poisson equation. Employing the method of variation of parameters [72], we construct a general solution that is independent of specific dielectric models. This solution yields the crucial result: an effective classical Hamiltonian for quantum plasmons in two-dimensional inhomogeneous systems. This classical Hamiltonian turns out to be given by the Lindhard expression for the two-dimensional dielectric function with a coordinate-dependent Fermi momentum and background dielectric constant. The Hamiltonian describes the classical dynamics of plasmons in phase space through the Hamilton-Jacobi equation, cf. Sec. 1.3.1.

In Sec. 2.1.3, the $\hbar^1$-order out-of-plane differential equation resulting from the Poisson equation is solved. The decoupling performed via the method of operator separation of variables simplifies the subsequent application of the semiclassical Ansatz, and the derivation of the transport equation. From the transport equation, we calculate the leading-order amplitude of the Ansatz, Eq. (1.53). Thus, we obtain a semiclassical expression for the full 3D electrostatic potential, which reveals the wave-like character of the plasmons and is related to the induced or Hartree potential via the self-consistency condition, as discussed in Sec. 1.3.2. The induced potential mimics the wavefunction for particles subject to the Schrödinger equation. We show in Sec. 2.1.4 that the induced potential can be related to the induced electron density and to the electrostatic energy density, which has a physical interpretation as a probability density. Until this point, our considerations are completely general, independent of the z dependence of the dielectric environment.

We consider the applicability of the semiclassical approximation in Sec. 2.1.5. Since $\hbar$ is a constant, it is strictly speaking not possible to use it as a small parameter in the series expansion. The main reason for this is that it is unclear with respect to which quantity it should be small. In order to resolve this issue, we identify the correct dimensionless semiclassical parameter for the series expansion. In this section, we consider the characteristic scales in the problem, and define dimensionless quantities. We also discuss the applicability regime of the semiclassical approximation, and give the dimensionless analytical formulas for the effective classical Hamiltonian at zero temperature and the analytical dispersion relation, relating it to the homogeneous results discussed in Sec. 1.2.3.

Finally, we consider two specific models for the dielectric environment $\varepsilon(\mathbf{x}, z)$ in Sec. 2.2. First, we examine the simplest model, where the dielectric environment is independent of z and characterized by constants ε_A and ε_B above and below the 2D layer, respectively. This leads to an effective dielectric function in the plasmon dispersion, which represents the average screening effects of the surrounding media. This simple model will be applied in the subsequent chapter, Ch. 3, on scattering of plasmons. Second, we explore a more complex model for layered materials, where a thin film with dielectric constant ε_{M} and finite thickness d is sandwiched between two semi-infinite layers with dielectric constants ε_{A} and ε_{B}. This model captures nonlocal Coulomb screening from the substrate [34, 39, 40] and has been shown to accurately describe electrons in materials like TMDCs [34, 35]. This more sophisticated model will be utilized in Chs. 4 and 5 discussing plasmonic waveguides and plasmonic crystals.

2.1.1 Separating in-plane and out-of-plane degrees of freedom in the Poisson equation

As discussed in Sec. 1.3.2, we consider electrons that are confined to a two-dimensional plane $\mathbf{x} = (x, y)$, whose dynamics are governed by the single-electron Hamiltonian $\hat{H}_0$. We assume that this single-electron Hamiltonian has the form $\hat{H}_0 = \hat{\mathbf{p}}^2/2m + U(\mathbf{x})$, where the potential $U(\mathbf{x})$ in this Hamiltonian can be related to a spatially varying electron density $n^{(0)}(\mathbf{x})$ using the Thomas-Fermi approximation [2, 28, 58]. A natural way to obtain a spatially varying $n^{(0)}(\mathbf{x})$ is by doping or combining different materials.

In equilibrium, the electrons have a certain distribution, that can be described by the equilibrium density operator $\hat{\rho}_0$. When a weak perturbation is applied to the system, this equilibrium distribution is modified. In turn, this new electron distribution gives rise to a potential, which can be computed through the Poisson equation. In this way, a system of equations arises, which has to be solved self-consistently [1, 2, 50]. Within this framework, the plasmons are the self-sustained oscillations that remain after the external perturbation is switched off.

The induced density $n(\mathbf{x})$ Eq. (1.60) in the layer at $z = 0$ gives rise to an electrostatic potential $\Phi(\mathbf{x}, z)$ through the Poisson equation. In two-dimensional materials, the strength of this electrostatic potential is strongly affected by the dielectric environment of the two-dimensional charge layer [1, 40], since this environment screens the Coulomb interaction between charges at different positions. This screening becomes more important when the distance between two charges increases, so it is especially important in the limit of small $\mathbf{q}$. We denote the dielectric environment of the surrounding media by $\varepsilon(\mathbf{x}, z)$, explicitly indicating that it can vary in both the in-plane and out-of-plane directions. The next step in the derivation, similar to the homogeneous case discussed in Sec. 1.2.3, is relating the induced potential to the induced density through the Poisson equation. As previously mentioned, the electric field $\boldsymbol{\mathcal{E}}$ is a three-dimensional quantity. The three-dimensional potential $V(\mathbf{x}, z, t)$, which is related to the electric field by $\boldsymbol{\mathcal{E}} = e^{-1}\nabla V$, satisfies the Poisson equation

$$\langle \nabla, \varepsilon(\mathbf{x}, z) \nabla \rangle \, V(\mathbf{x}, z) = -4\pi e^2 n(\mathbf{x}, z), \tag{2.1}$$

where $V(\mathbf{x}, z)$ is related to the electrostatic potential by $V = -e\Phi$, with e the elementary charge, and the induced density equals $n(\mathbf{x}, z) = n(\mathbf{x})\delta(z)$. In Eq. (2.1), $\nabla = (\partial/\partial\mathbf{x}, \partial/\partial z)$ denotes the three-dimensional gradient and $\langle a, b\rangle$ denotes the three-dimensional Cartesian inner product between the vectors a and b. Note that we do not consider external electric fields in the Poisson equation because we are interested in plasmons, which are self-sustained collective oscillations.

Since we consider electrons that are confined to the plane $z = 0$, the induced density can be written as

$$n(\mathbf{x}, z, t) = n(\mathbf{x}, t)\delta(z), \tag{2.2}$$

where $n(\mathbf{x}, t)$ only depends on the in-plane coordinates $\mathbf{x}$, see also expression (1.60). We consider a setup where the plane $z = 0$ is encapsulated by two arbitrary dielectric media. We allow the background dielectric constant $\varepsilon(\mathbf{x}, z)$ to depend on both $\mathbf{x}$ and z. The latter occurs when we have different dielectrics above and below the plane or even multiple layers of dielectrics. A dependence on $\mathbf{x}$ can for instance originate from a combination of different dielectrics below the plane [39, 40].

It is important to note that the Hartree potential $V_{\mathrm{pl}}(\mathbf{x})$ in the Hamiltonian Eq. (1.47) is caused by the potential $V(\mathbf{x}, z)$ obtained from Eq. (2.1). In order to ensure that our set of equations is self-consistent, we therefore impose the additional condition [1]

$$V(\mathbf{x}, z = 0) = V_{\mathrm{pl}}(\mathbf{x}). \tag{2.3}$$

In other words, the full three-dimensional potential should be equal to the in-plane potential (1.53), in the form of the semiclassical Ansatz, in the plane $z = 0$. Note that this condition is not necessary in this form for a three-dimensional electron density, since we automatically have V on both sides of the Poisson equation in that case [50]. Since we consider a time-independent Hamiltonian $\hat{H}_0$, the time dependence has the standard exponential form, see also expression (1.50), and we omit it from here on.

At this point, one can proceed in two different ways. In the first approach, we note that the self-consistency condition (2.3) implies that $V(\mathbf{x}, z)$ is proportional to $\exp(iS(\mathbf{x})/\hbar)$, as the semiclassical Ansatz Eq. (1.53). We may therefore write down an Ansatz for $V(\mathbf{x}, z)$ as

$$V(\mathbf{x}, z) = e^{iS(\mathbf{x})/\hbar} \left(V_0(\mathbf{x}, z) + \hbar V_1(\mathbf{x}, z) + \mathcal{O}(\hbar^2) \right), \tag{2.4}$$

and solve the Poisson equation order by order in $\hbar$. However, this approach mixes the separation of the in-plane and out-of-plane degrees of freedom with the application of the semiclassical Ansatz.

A more elegant approach, in which these two steps are separated, can be developed by modifying the operator separation of variables technique [56, 71].

In its original formulation, this technique can be seen as a generalization of the Born-Oppenheimer approximation [56]. It can be regarded as an adiabatic approximation, and can therefore be used when there are two different length scales. As noted in Ref. [1], the in-plane variables $\mathbf{x}$ can be regarded as "slow" variables, since the system parameters do not change significantly on the scale of the electron wavelength, as asserted by the semiclassical limit. The out-of-plane variables z can, instead, be considered "fast" variables, since the delta function causes a fast change in the out-of-plane coordinate z. The method then separates the motion in these fast and slow variables order by order in the small parameter $\hbar$. In what follows, we proceed with this second method. For completeness, we show the derivation of the first approach in appendix A, derived with the method of undetermined coefficients.

In the original formulation of the Born-Oppenheimer approximation, one employs an instantaneous eigenfunction that depends on the variables $\mathbf{x}$ and z. As explained in Ref. [56], this formulation does not suffice when one deals with a rapidly oscillating exponent (1.53). Instead, one has to consider a slightly more complicated form for the potential $V(\mathbf{x}, z)$, namely [1]

$$V(\mathbf{x}, z) = (\hat{\Gamma} V_{\mathrm{pl}})(\mathbf{x}, z), \tag{2.5}$$

where $\hat{\Gamma}$ is a pseudodifferential operator that transforms the in-plane potential $V_{\mathrm{pl}}(\mathbf{x})$, which is independent of z, into the full three-dimensional potential $V(\mathbf{x}, z)$. This Ansatz yields a generalized Born-Oppenheimer approximation, in which the instantaneous eigenfunction is replaced by an operator. In the traditional Born-Oppenheimer approximation, one would thus in this context consider a function $\Gamma(\mathbf{x}, z)$, instead of an operator $\hat{\Gamma}$. Here we have to use a full pseudodifferential operator, since V_{pl} is a rapidly oscillating exponential, as discussed in detail in Ref. [56].

Similar to $\sigma(\hat{\Pi})$ (Eq. (1.61)), the symbol $\sigma(\hat{\Gamma})$ has an asymptotic expansion in powers of $\hbar$, given by

$$\sigma(\hat{\Gamma}) = \Gamma_0(\mathbf{x}, \mathbf{q}, z) + \hbar \Gamma_1(\mathbf{x}, \mathbf{q}, z) + \mathcal{O}(\hbar^2), \tag{2.6}$$

cf. Eq. (1.56). Comparing these symbols, which depend on $\mathbf{x}$, $\mathbf{q}$ and z, to the instantaneous eigenfunctions in the original Born-Oppenheimer approach, we may loosely say that the generalized form (2.5) adds the momentum variable to the original Ansatz. As we see in section 2.1.2, this momentum variable will be replaced by $\partial S/\partial \mathbf{x}$ at the end of the calculation, similar to the WKB approximation discussed in Sec. 1.3.1.

Inserting the Ansatz (2.5) into the Poisson Eq. (2.1), and taking the induced density (1.60) and its confinement to the $z = 0$ plane (Eq. (2.2)) into account, we obtain

$$\left(\hat{F} \hat{\Gamma} + 4\pi e^2 \delta(z) \hat{\Pi} \right) V_{\mathrm{pl}}(\mathbf{x}) = 0, \tag{2.7}$$

where $\hat{F} \equiv \langle \nabla, \varepsilon(\mathbf{x}, z) \nabla \rangle$. Since we are looking for a plasmon, we require that $V_{\text{pl}}(\mathbf{x}) \neq 0$. We therefore obtain an operator equation, to wit,

$$\hat{F}\hat{\Gamma} + 4\pi e^2 \delta(z) \hat{\Pi} = 0. \tag{2.8}$$

In what follows, we solve this equation order by order in $\hbar$.

To this end, we first compute the symbol of the operator $\hat{F}$ in Eq. (2.8), taking only the slow variables into account [56]. In other words, we compute the symbol using expression (1.55), leaving the variable z out of the consideration. In this way, we obtain the operator-valued symbol

$$\sigma(\hat{F}) = F_0\left(\mathbf{x}, \mathbf{q}, z, \frac{\partial}{\partial z}\right) + \hbar F_1\left(\mathbf{x}, \mathbf{q}, z, \frac{\partial}{\partial z}\right) + \mathcal{O}(\hbar^2), \tag{2.9}$$

where

$$F_0\left(\mathbf{x}, \mathbf{q}, z, \frac{\partial}{\partial z}\right) = -\frac{1}{\hbar^2}\varepsilon(\mathbf{x}, z)|\mathbf{q}|^2 + \frac{\partial}{\partial z}\left(\varepsilon(\mathbf{x}, z)\frac{\partial}{\partial z}\right), \tag{2.10}$$

and

$$F_1\left(\mathbf{x}, \mathbf{q}, z, \frac{\partial}{\partial z}\right) = \frac{i}{\hbar^2}\left\langle \mathbf{q}, \frac{\partial \varepsilon}{\partial \mathbf{x}}(\mathbf{x}, z) \right\rangle. \tag{2.11}$$

Note that the derivative with respect to z is included in the principal symbol. Naively, one may say this is because the derivative gives rise to a factor $1/\hbar$, as we will see shortly. More fundamentally, we previously observed that changes in the out-of-plane coordinate z are fast, whereas changes in the in-plane coordinates $\mathbf{x}$ are slow. When one introduces proper dimensionless units, as we do in Sec. 2.1.5, one therefore sees that the combination $\hbar/z$ is of order one, cf. the discussion in Ref. [56].

We can now compute the symbol for all terms in Eq. (2.8), using the expression for the symbol of an operator product [49, 67], that is,

$$\sigma(\hat{a}\hat{b}) = \sigma(\hat{a})\sigma(\hat{b}) - i\hbar \left\langle \frac{\partial \sigma(\hat{a})}{\partial \mathbf{q}}, \frac{\partial \sigma(\hat{b})}{\partial \mathbf{x}} \right\rangle + \mathcal{O}(\hbar^2). \tag{2.12}$$

Inserting the symbols of the Ansatz (2.5) into the Poisson equation (2.1) and taking the induced electron density Eq. (1.60) into account, one can convert the Poisson equation into two ordinary differential equations for the principal and subprincipal symbols Γ_0 and Γ_1, respectively, of the operator $\hat{\Gamma}$. The terms of order $\hbar^0$ give

$$F_0\left(\mathbf{x}, \mathbf{q}, z, \frac{\partial}{\partial z}\right)\Gamma_0(\mathbf{x}, \mathbf{q}, z) = -4\pi e^2 \delta(z) \Pi_0(\mathbf{x}, \mathbf{q}), \tag{2.13}$$

whilst the terms of order $\hbar^1$ give

$$\begin{aligned} F_1\left(\mathbf{x},\mathbf{q},z,\frac{\partial}{\partial z}\right)\Gamma_0(\mathbf{x},\mathbf{q},z) + F_0\left(\mathbf{x},\mathbf{q},z,\frac{\partial}{\partial z}\right)\Gamma_1(\mathbf{x},\mathbf{q},z) \\ - i\left\langle \frac{\partial F_0}{\partial \mathbf{q}}\left(\mathbf{x},\mathbf{q},z,\frac{\partial}{\partial z}\right), \frac{\partial \Gamma_0}{\partial \mathbf{x}}(\mathbf{x},\mathbf{q},z)\right\rangle \\ = -4\pi e^2\delta(z)\Pi_1(\mathbf{x},\mathbf{q}). \end{aligned} \tag{2.14}$$

These expressions constitute two linear ordinary differential equations (ODE's) (with variable z) for the principal symbol Γ_0 and the subprincipal symbol Γ_1. Loosely speaking, one can think of the symbol $\sigma(\hat{\Gamma})(\mathbf{x},\mathbf{q},z)$ as a generalization of the instantaneous eigenfunction in the adiabatic approximation. As shown by Eqs. (2.13), (2.14) and (2.5), the principal symbol Γ_0 and subprincipal symbol Γ_1 express the z-dependence of the full potential for given values of $\mathbf{x}$ and $\mathbf{q}$. In the remainder of this section, we solve Eqs. (2.13) and (2.14) one by one and thereby construct the symbol $\sigma(\hat{\Gamma})$.

From the above considerations, we are ready to construct the effective classical Hamiltonian, without explicitly calculating the symbol $\Gamma_0(\mathbf{x},\mathbf{q},z)$. Inserting the semiclassical Ansatz (1.53) and the operator separation of variables equation (2.5) into the self-consistency condition (2.3), and taking the lowest order, one finds the secular equation $\mathcal{H}_0(\mathbf{x},\partial S/\partial\mathbf{x})\varphi_0(\mathbf{x})\exp(iS/\hbar) = 0$, where [1]

$$\mathcal{H}_0(\mathbf{x},\mathbf{q}) = 1 - \Gamma_0(\mathbf{x},\mathbf{q},z=0). \tag{2.15}$$

This secular equation is equivalent to the Hamilton-Jacobi equation $\mathcal{H}_0(\mathbf{x},\partial S/\partial\mathbf{x}) = 0$ for the action $S(\mathbf{x})$. This implies that $\mathcal{H}_0$ can be interpreted as the effective classical Hamiltonian that describes the dynamics of quantum plasmons.

We summarize this subsection by saying that one obtains an effective classical Hamiltonian for two-dimensional quantum plasmons by solving the ordinary differential equation (2.13) and inserting the solution into Eq. (2.15). In the next subsection, we study the structure of this ordinary differential equation in detail, and construct the effective classical Hamiltonian $\mathcal{H}_0(\mathbf{x},\mathbf{q})$ for a general class of models for $\varepsilon(\mathbf{x},z)$. In Sec. 2.1.3, we solve the second ordinary differential equation (2.14) in order to derive the transport equation, and an explicit expression for the first order of the three-dimensional induced potential Eq. (2.4).

2.1.2 General construction of the effective classical Hamiltonian

For an explicit expression of the effective classical Hamiltonian Eq. (2.15), we solve the ordinary differential equation (2.13) with the method of variation of parameters [72]. Let us therefore first consider the fundamental solutions of the homogeneous differential equation, i.e.,

$$\left(\frac{\partial}{\partial z}\left(\varepsilon(\mathbf{x},z)\frac{\partial}{\partial z}\right)-\frac{1}{\hbar^2}\varepsilon(\mathbf{x},z)|\mathbf{q}|^2\right)w_{\mathrm{H}}(z)=0. \tag{2.16}$$

This equation has two solutions, which we denote by w_1 and w_2, and whose specific form depends on the form of $\varepsilon(\mathbf{x},z)$. In order to make some progress, we assert that $\varepsilon(\mathbf{x},z)$ goes to a constant as $z\to\pm\infty$. This condition is physically very intuitive, and does not limit the practical applicability of our theory, since it does not dictate the precise shape of $\varepsilon(\mathbf{x},z)$. Because of our assertion, the fundamental solutions $w_{1,2}$ are asymptotically equivalent to a linear combination of the functions $\exp(\pm|\mathbf{q}|z/\hbar)$ in the limit $z\to\pm\infty$, as follows from Eq. (2.16). By making use of the freedom in our choice of the asymptotic solutions, we can then construct them in such a way that $w_1\to 0$ as $z\to\infty$, and $w_2\to 0$ as $z\to-\infty$.

The solution of the inhomogeneous differential equation (2.13) can subsequently be expressed as [72]

$$\Gamma_0(z)=c_1(z)w_1(z)+c_2(z)w_2(z). \tag{2.17}$$

We find the functions $c_1(z)$ and $c_2(z)$ by inserting Eq. (2.17) into the inhomogeneous differential equation. Using standard arguments from the method of variation of parameters [72], we then find a set of two differential equations, which can be combined into

$$W^T\begin{pmatrix}\frac{dc_1}{dz}\\ \frac{dc_2}{dz}\end{pmatrix}=\begin{pmatrix}0\\ f(z)\end{pmatrix}, \tag{2.18}$$

where W resembles the Wronskian matrix, and is given by

$$W=\begin{pmatrix}w_1 & \varepsilon\frac{dw_1}{dz}\\ w_2 & \varepsilon\frac{dw_2}{dz}\end{pmatrix}. \tag{2.19}$$

The function $f(z)=-4\pi e^2\Pi_0\delta(z)$ corresponds to the inhomogeneous term in Eq. (2.13). Because of the homogeneous differential equation (2.16), the derivative of the determinant $\det(W)$ with respect to z vanishes, which means that $\det(W)$ is constant. Moreover, $\det(W)$ is non-zero when the fundamental solutions $w_{1,2}$ are linearly independent, meaning that the matrix W is invertible.

On physical grounds, we require that the potential $V(\mathbf{x}, z)$, see Eqs. (2.1) and (2.5), goes to zero as $z \to \pm\infty$. In turn, this implies that $\Gamma_0 \to 0$ as $z \to \pm\infty$. Because of the way in which we constructed $w_{1,2}(z)$, this condition implies that $c_1 \to 0$ as $z \to -\infty$ and $c_2 \to 0$ as $z \to \infty$. We may therefore write

$$\begin{aligned} \int_{-\infty}^{z} \frac{dc_1}{dz'} dz' &= c_1(z) - c_1(-\infty) = c_1(z), \\ \int_{z}^{\infty} \frac{dc_2}{dz'} dz' &= c_2(\infty) - c_2(z) = -c_2(z). \end{aligned} \tag{2.20}$$

Rewriting the solution (2.17) for $\Gamma_0(z)$ using Eq. (2.20), and inserting expressions for the derivatives obtained from Eq. (2.18), we obtain, cf. Ref. [72],

$$\begin{aligned} \Gamma_0(z) &= \left(\int_{-\infty}^{z} \frac{dc_1}{dz'} dz' \right) w_1(z) - \left(\int_{z}^{\infty} \frac{dc_2}{dz'} dz' \right) w_2(z) \\ &= \int_{-\infty}^{\infty} g(z, z') f(z') dz', \end{aligned} \tag{2.21}$$

where

$$g(z, z') = \begin{cases} -\frac{1}{\det W} w_2(z') w_1(z), & -\infty < z' < z \\ -\frac{1}{\det W} w_1(z') w_2(z), & z < z' < \infty \end{cases}. \tag{2.22}$$

The function $g(z, z')$ is, of course, a Green's function.

Inserting our expression for the inhomogeneous term $f(z) = -4\pi e^2 \Pi_0 \delta(z)$, we now easily find that

$$\Gamma_0(z) = \begin{cases} \frac{2\pi e^2 \hbar}{\varepsilon_{\mathrm{eff}} |\mathbf{q}|} \Pi_0 \frac{w_1(z)}{w_1(0)}, & z > 0 \\ \frac{2\pi e^2 \hbar}{\varepsilon_{\mathrm{eff}} |\mathbf{q}|} \Pi_0 \frac{w_2(z)}{w_2(0)}, & z < 0 \end{cases}, \tag{2.23}$$

where we defined an effective dielectric function $\varepsilon_{\mathrm{eff}}$ as

$$\varepsilon_{\mathrm{eff}} = \frac{\hbar}{2|\mathbf{q}|} \frac{\det W}{w_1(0) w_2(0)}, \tag{2.24}$$

which captures the three-dimensional screening from the environment. Using the definition of W, this expression can also be written as

$$\varepsilon_{\mathrm{eff}} = \frac{\hbar}{2|\mathbf{q}|} \varepsilon(0) \left(\frac{1}{w_2(0)} \frac{dw_2}{dz}(0) - \frac{1}{w_1(0)} \frac{dw_1}{dz}(0) \right). \tag{2.25}$$

This dielectric function should not be confused with the longitudinal Lindhard (dielectric) function, which describes the dynamical screening of the electric field by electrons in the 2D electron gas. Instead, that dynamical screening is incorporated into our theory through the effective classical Hamiltonian, which we obtain by substituting the expression for $\Gamma_0(z)$ into the self-consistency condition Eq. (2.3). This yields

$$\mathcal{H}_0(\mathbf{x}, \mathbf{q}) = 1 - \frac{2\pi e^2 \hbar}{\varepsilon_{\text{eff}}(\mathbf{x}, \mathbf{q})|\mathbf{q}|} \Pi_0(\mathbf{x}, \mathbf{q}), \tag{2.26}$$

where we now explicitly indicated the dependence on $\mathbf{x}$ and $\mathbf{q}$. The Hamilton-Jacobi equation then reads $\mathcal{H}_0\left(\mathbf{x}, \partial S/\partial \mathbf{x}\right) = 0$.

At first glance, one may think that we have incorrectly matched the different orders of $\hbar$ in this expression, since Γ_0 contains a factor of $\hbar$. As we explain in detail in Sec. 2.1.5, this is however not the case. In short, the apparent contradiction can be resolved by noting that we should actually perform our semiclassical expansion using a dimensionless parameter, instead of $\hbar$. Introducing this dimensionless expansion parameter, one can show that Γ_0 is of order one. Alternatively, one may refer to the analogy with the homogeneous case, which was discussed in Sec. 1.2.3 and which implies that the terms in Eq. (2.26) are of the same order.

It is important to note that this classical Hamiltonian strongly resembles the Lindhard expression for the dielectric function of homogeneous two-dimensional materials Eq. (1.30), with two main differences. First of all, it includes a dependence on the coordinate $\mathbf{x}$, whereas the homogeneous dielectric function only depends on $\mathbf{q}$. The classical Hamiltonian $\mathcal{H}_0$ is therefore a function on classical phase space, rather than a function of the plasmon momentum only. As we discussed below Eq. (1.62), the position dependence of the polarization amounts to the replacement of the Fermi momentum by the local Fermi momentum $p_{\text{F}}(\mathbf{x})$. The second difference between the two-dimensional dielectric function and the effective Hamiltonian $\mathcal{H}_0$ is a factor of ε_{eff}, since the former should evaluate to ε_{eff} in the absence of a polarization.

To avoid misunderstanding, we again emphasize that $\mathcal{H}_0$ is not a classical Hamiltonian in the traditional sense, since it does not describe a classical plasma, as we just established. Instead, we use $\mathcal{H}_0$ to calculate the classical trajectories which form the basis for the construction of our asymptotic solution, as we discussed in the previous chapter. This process is analogous to the way in which one adds interference to the rays in geometrical optics. Although we do not consider it explicitly in this article, we remark that we can still make the transition to a classical plasma in the traditional sense by making two changes in expression (1.62) for the polarization. First, we should replace the Fermi-Dirac distribution by the Maxwell-Boltzmann distribution, or, equivalently, consider the regime where the temperature is much larger than the chemical potential. Second, we should only keep the linear term in the expansion of the denominator, that is, $H_0(\mathbf{x}, \mathbf{p}) - H_0(\mathbf{x}, \mathbf{p} + \mathbf{q}) \approx -\langle \mathbf{p}, \mathbf{q} \rangle / m$. Unfortunately, this transition to a classical plasma leads to practical issues with Landau damping, cf. the discussion in Ref. [50].

Let us take a step back and reflect on the physical implications of this result. Equation (2.26) shows that we can capture the effect of any dielectric environment by a single effective dielectric function $\varepsilon_{\text{eff}}(\mathbf{x}, \mathbf{q})$ in the effective classical Hamiltonian. This is an important generalization of Ref. [1], where this effective dielectric function was only computed for the simplest model of the dielectric environment. Here, we have instead considered an arbitrary shape of $\varepsilon(\mathbf{x}, z)$, only imposing the condition that it becomes constant as $z \to \pm\infty$. Moreover, we have expressed the effective dielectric function in terms of the value of the fundamental solutions to the homogeneous differential equation (2.16) at the point $z = 0$, which makes our result easily applicable to different dielectric environments.

We note that the specific model that is commonly used for layered structures with an encapsulated thin film with finite thickness d also leads to an effective dielectric function ε_{eff} in the Lindhard function [34, 40, 70]. As we already discussed, the effective classical Hamiltonian can be viewed as, and takes the form of, the Lindhard function with position-dependent parameters. The result (2.26) may therefore seem somewhat straightforward. However, we would like to stress two key findings. First of all, it does not depend on the specific model that is used. In some sense, one may therefore say that expression (2.26), and more specifically Eq. (2.25) also extends the known results for the homogeneous case. Second, our result is not just valid for the homogeneous case, but also, and especially, for the inhomogeneous case, as long as $\lambda_{\text{el}}/\ell \ll 1$ as discussed in the introduction.

The Hamilton-Jacobi equation determines the action $S(\mathbf{x})$ in our Ansatz for the induced potential $V_{\text{pl}}(\mathbf{x})$. As is well known from classical mechanics [73, 74], this Hamilton-Jacobi equation is equivalent to the system of Hamilton equations for the effective classical Hamiltonian $\mathcal{H}_0$, i.e.,

$$\frac{d\mathbf{x}}{d\tau} = \frac{\partial \mathcal{H}_0}{\partial \mathbf{q}}, \qquad \frac{d\mathbf{q}}{d\tau} = -\frac{\partial \mathcal{H}_0}{\partial \mathbf{x}}. \tag{2.27}$$

Of course, suitable initial conditions have to be specified, which we discuss in more detail in Ch. 3. For now, let us assume that these conditions can be parameterized by α and constitute a line, Λ^1. In a typical scattering setup, the parameter α corresponds to the coordinate along the initial wavefront. Under these assumptions, we can denote the solutions of Hamilton's equations by $(\mathbf{X}(\tau, \alpha), \mathbf{Q}(\tau, \alpha))$. Together, these solutions describe a two-dimensional plane Λ^2 in classical phase space, parameterized by (τ, α), which is known as a Lagrangian manifold [48, 55]. In general, open trajectories in phase space correspond to plasmon scattering (Ch. 3). Closed trajectories, on the other hand, correspond to bound states and plasmon quantization. We consider these bound states in detail in Ch. 4, in which we also analyze the classical Hamiltonian (2.26). The classical action can be written as

$$S(\mathbf{x}) = \int_{\mathbf{x}_0}^{\mathbf{x}} \langle \mathbf{Q}, d\mathbf{X} \rangle , \tag{2.28}$$

where we integrate from an initial point $\mathbf{x}_0$ to the point $\mathbf{x}$. In passing from the Hamilton-Jacobi equation to Hamilton's equations, we have lifted the problem from configuration space, on which we deal with the coordinate $\mathbf{x}$, to phase space, which involves the coordinates $(\mathbf{x}, \mathbf{q})$. The projection of the Lagrangian manifold Λ^2 onto the configuration space is (locally) invertible, as long as the Jacobian

$$J(\mathbf{x}) = \det\left(\frac{\partial \mathbf{X}}{\partial(\tau, \alpha)}\right) \tag{2.29}$$

is not equal to zero [48, 75, 76]. As we have seen in Sec. 1.3.1, this Jacobian plays an important role in the construction of the amplitude, which is discussed in the next section.

2.1.3 General construction of the induced potential

In the previous subsections, we discussed the derivation of the classical Hamiltonian $\mathcal{H}_0$. While performing this derivation, we also found the defining equation for the classical action $S(\mathbf{x})$ in the Ansatz (1.53) for the Hartree potential, namely the Hamilton-Jacobi equation. In this subsection, we consider the amplitude φ_0 in the Ansatz (1.53).

In the final step of the derivation of the classical Hamiltonian in Sec. 2.1.1, we inserted both the Ansatz (1.53) and expression (2.5) into the self-consistency condition (2.3). Gathering the terms of order $\hbar^0$ one finds that the amplitude φ_0 drops out, in the resulting Hamilton-Jacobi equation (cf. Eq. (2.15)). In order to obtain an equation for this amplitude, we should therefore consider the terms of order $\hbar^1$ in the self-consistency condition. All terms of the operator separations of variables equation (2.5) up to order $\hbar^1$ are given by

$$\begin{aligned}(\hat{\Gamma} V_{\mathrm{pl}})(\mathbf{x}, z) = e^{iS(\mathbf{x})/\hbar} \Bigg(& \Gamma_0 \varphi_0 + \hbar \Gamma_0 \varphi_1 - i\hbar \left\langle \frac{\partial \Gamma_0}{\partial \mathbf{q}}, \frac{\partial \varphi_0}{\partial \mathbf{x}} \right\rangle \\ & + \hbar \Gamma_1 \varphi_0 - \frac{i\hbar}{2} \sum_{j,k} \frac{\partial^2 \Gamma_0}{\partial q_j \partial q_k} \frac{\partial^2 S}{\partial x_j \partial x_k} \varphi_0 \Bigg) + \mathcal{O}(\hbar^2),\end{aligned} \tag{2.30}$$

where Γ_0 and Γ_1 are to be evaluated at the point $(\mathbf{x}, \partial S/\partial \mathbf{x}, z)$. Although we did not determine the quantity Γ_1 yet, we continue with this expression. We come back to the defining equations for Γ_1, also called the subprincipal symbol, at the end of this subsection.

Gathering all terms of order $\hbar^1$ in the self-consistency condition (2.3) with the help of Eq. (2.30), and using our previous definition of the effective classical Hamiltonian (2.15), we find that [1]

$$\begin{aligned}& \mathcal{H}_1\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}\right) \varphi_0 - i \left\langle \frac{\partial \mathcal{H}_0}{\partial \mathbf{q}}\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}\right), \frac{\partial \varphi_0}{\partial \mathbf{x}} \right\rangle \\ & - \frac{i}{2} \sum_{j,k} \frac{\partial^2 \mathcal{H}_0}{\partial q_j \partial q_k}\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}\right) \frac{\partial^2 S}{\partial x_k \partial x_j} \varphi_0 = 0,\end{aligned} \tag{2.31}$$

where we defined $\mathcal{H}_1 \equiv -\Gamma_1(\mathbf{x}, \partial S/\partial \mathbf{x}, z=0)$. This equation is known in the literature [1, 48] as the transport equation.

We now show how to solve the transport equation using standard semiclassical techniques, which gives us the amplitude $\varphi_0(\mathbf{x})$. First, we look at the time evolution of the Jacobian (2.29), which can be determined using the Liouville formula [1, 48, 50]

$$\frac{d}{dt}\log J = \sum_{j,k} \frac{\partial^2 \mathcal{H}_0}{\partial q_j \partial q_k} \frac{\partial^2 S}{\partial x_j \partial x_k} + \sum_j \frac{\partial^2 \mathcal{H}_0}{\partial x_j \partial q_j}. \tag{2.32}$$

Second, we use that the time evolution of the amplitude φ_0 along the trajectories of the Hamiltonian system (2.27) is given by

$$\frac{d\varphi_0}{dt} = \left\langle \frac{\partial \mathbf{x}}{\partial t}, \frac{\partial \varphi_0}{\partial \mathbf{x}} \right\rangle = \left\langle \frac{\partial \mathcal{H}_0}{\partial \mathbf{q}}, \frac{\partial \varphi_0}{\partial \mathbf{x}} \right\rangle. \tag{2.33}$$

Substitution of Eq. (2.32) and Eq. (2.33) in the transport equation (2.31) yields

$$\frac{d\varphi_0}{dt} = -i\mathcal{H}_1\varphi_0 - \frac{1}{2}\left(\frac{d}{dt}\log J - \sum_j \frac{\partial^2 \mathcal{H}_0}{\partial x_j \partial q_j}\right)\varphi_0. \tag{2.34}$$

We now define the quantity $A_0(\mathbf{x})$ by $A_0(\mathbf{x}) \equiv \varphi_0(\mathbf{x})\sqrt{J(\mathbf{x})}$, which transforms the above equation in an ordinary differential equation for $A_0(\mathbf{x})$ along the trajectories of the dynamical system, namely

$$\frac{dA_0}{dt} + i\left(\mathcal{H}_1 + \frac{i}{2}\sum_j \frac{\partial^2 \mathcal{H}_0}{\partial x_j \partial q_j}\right)A_0 = 0, \tag{2.35}$$

which has the straightforward solution

$$A_0(\mathbf{x}) = A_0^0 \exp\left(-i\int_0^t \mathcal{H}_1 + \frac{i}{2}\sum_j \frac{\partial^2 \mathcal{H}_0}{\partial x_j \partial q_j} dt'\right). \tag{2.36}$$

The integral in this expression is to be performed along the trajectories of the Hamiltonian system. Note that, strictly speaking, the above derivation is only valid when there is a one-to-one mapping of the trajectories onto configuration space $\mathbf{x}$, that is, in the absence of turning points. However, as argued in Refs. [1, 50], one can incorporate the presence of turning points in the description by introducing the Maslov index. We come back to this point in greater detail in Ch. 4.

Equation (2.36) gives us the solution for the amplitude $\varphi_0(\mathbf{x}) = A_0(\mathbf{x})/\sqrt{J(\mathbf{x})}$ in expression (1.53). However, as we already mentioned in the beginning of this subsection, this solution contains the so far undetermined quantity $\mathcal{H}_1(\mathbf{x}, \mathbf{q}) = -\Gamma_1(\mathbf{x}, \mathbf{q}, z=0)$. Therefore, we consider the second differential equation, given by Eq. (2.14), to find a solution for this quantity. The quantity Π_1 in Eq. (2.14) is the so-called subprincipal symbol of the polarization operator. Since $\hat{\Pi}$ is a Hermitian operator, the following relation between the principal and the subprincipal symbol always holds [49, 50]:

$$\operatorname{Im} \Pi_1(\mathbf{x}, \mathbf{q}) = -\frac{1}{2} \sum_j \frac{\partial \Pi_0}{\partial q_j \partial x_j}(\mathbf{x}, \mathbf{q}). \tag{2.37}$$

In the specific case that we consider in this thesis, where the electron Hamiltonian takes the form $\hat{H} = \hat{\mathbf{p}}^2/2m + U(\mathbf{x})$, we have an even simpler relation [50], namely

$$\Pi_1(\mathbf{x}, \mathbf{q}) = -\frac{i}{2} \sum_j \frac{\partial \Pi_0}{\partial q_j \partial x_j}(\mathbf{x}, \mathbf{q}). \tag{2.38}$$

In the remainder of this chapter, we assume that Eq. (2.38) holds. At the same time, we briefly explore the consequences of the relation (2.37), which is more general and also holds for other electron Hamiltonians. For a more detailed discussion on the background of Π_1, we refer to Ref. [50].

So far, we reviewed the derivation of the amplitude φ_0 and found that this amplitude contains the quantity $\mathcal{H}_1 = -\Gamma_1(z = 0)$. In turn, the quantity Γ_1 is the solution of the ordinary differential equation (2.14). Next, we construct an expression for Γ_1 using variation of parameters, in the same way as we constructed an expression for Γ_0 in Sec. 2.1.2. As before, this solution generalizes the discussion in Ref. [1], where Γ_1 was only constructed for the simplest case using the method of undetermined coefficients, in which the dielectric environment does not depend on z. Here, we consider an arbitrary $\varepsilon(\mathbf{x}, z)$, only assuming that it becomes constant as $z \to \pm\infty$. With our expression for Γ_1, we subsequently compute the amplitude φ_0 according to the theory reviewed in the previous section. We finally show that the electromagnetic energy density computed with the potential $V(\mathbf{x}, z)$ can be interpreted as a probability density in the semiclassical sense, again generalizing the results from Ref. [1].

Let us therefore consider the differential equation (2.14). We first note that the only difference between the differential equations for Γ_0 and Γ_1, Eqs. (2.13) and (2.14), is the inhomogeneous term. The homogeneous differential equations are exactly the same. We can therefore use our previous fundamental solutions $w_1(z)$ and $w_2(z)$ to construct an expression for Γ_1. Let us denote the inhomogeneous term in Eq. (2.14) by

$$f_1(z) = f_{1s}(z) - 4\pi e^2 \Pi_1 \delta(z), \tag{2.39}$$

where we omitted the dependence on $(\mathbf{x}, \mathbf{q})$ and

$$f_{1s}(z) = -F_1(z)\Gamma_0(z) + i \left\langle \frac{\partial F_0}{\partial \mathbf{q}}\left(z, \frac{\partial}{\partial z}\right), \frac{\partial \Gamma_0}{\partial \mathbf{x}}(z) \right\rangle. \tag{2.40}$$

Drawing on the discussion in Sec. 2.1.2, we can then write

$$\Gamma_1(z) = \int_{-\infty}^{\infty} g(z,z')f_1(z')dz', \tag{2.41}$$

cf. Eq. (2.21), where $g(z,z')$ is given by Eq. (2.22).

We now note that we are not interested in the full expression for Γ_1, but only in its value at $z = 0$, since this value enters the amplitude (2.36) through $\mathcal{H}_1 = -\Gamma_1(z=0)$. We can compute this value by performing the integration in Eq. (2.41). This integration is straightforward for the part with the delta function, but more involved for f_{1s}, which also contains our expression (2.23) for Γ_0. Nevertheless, the integral can be calculated explicitly, using integration by parts, as we show in appendix C.1.

The final result can be cast in the form

$$\begin{aligned} \mathcal{H}_1 + \frac{i}{2}\sum_j \frac{\partial^2 \mathcal{H}_0}{\partial x_j \partial q_j} = &-\frac{2\pi e^2 \hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\left(\Pi_1 + \frac{i}{2}\sum_j \frac{\partial^2 \Pi_0}{\partial x_j \partial q_j}\right) \\ &- \frac{i}{2}\left\{\ln \varepsilon_{\text{eff}}|\mathbf{q}|, \mathcal{H}_0\right\}, \end{aligned} \tag{2.42}$$

where $\{a, \mathcal{H}_0\}$ denotes the Poisson bracket, defined in Eq. (C.22), of a and the effective Hamiltonian $\mathcal{H}_0$. We note that we are free to add a function $g(\mathcal{H}_0(\mathbf{x},\mathbf{q}))$ to the first argument of the Poisson bracket for Eq. (2.42) to still be true. However, this does not influence the final result, since we are constrained to level set $\mathcal{H}_0 = 0$ by the Hamilton-Jacobi equation. Because of Hamilton's equations, this Poisson bracket can be written as a total derivative with respect to time [73, 74]. This greatly simplifies the integration in the expression (2.36) for the amplitude, cf. the discussion in Ref. [1], and leads to

$$A_0(\mathbf{x}) = \frac{A_0^0}{\sqrt{\varepsilon_{\text{eff}}(\mathbf{x}, \partial S/\partial \mathbf{x})|\partial S/\partial \mathbf{x}|}} e^{i\Phi_{\text{B}}(\mathbf{x})}, \tag{2.43}$$

where $\mathbf{q}$ has become $\partial S/\partial \mathbf{x}$ in the denominator because we integrate along the trajectories of the Hamiltonian system. The quantity Φ_{B} in Eq. (2.43) is the Berry phase, defined by

$$\Phi_{\text{B}}(\mathbf{x}) = \int_0^t \frac{2\pi e^2 \hbar}{\sqrt{\varepsilon_{\text{eff}}|\mathbf{q}|}}\left(\Pi_1 + \frac{i}{2}\sum_j \frac{\partial^2 \Pi_0}{\partial q_j \partial x_j}\right) dt. \tag{2.44}$$

Because of Eq. (2.38), the Berry phase is zero for a parabolic electron Hamiltonian $H_0 = \hat{\mathbf{p}}^2/(2m) + U(\mathbf{x})$. However, for more complicated electron Hamiltonians the term in parentheses need not be zero. Nevertheless, Eq. (2.37) indicates that Eq. (2.44) is purely real and therefore indeed a phase.

With the help of the operator separation of variables equation (2.5) and the semiclassical Ansatz (1.53), we now obtain the leading-order term of the full potential $V(\mathbf{x},z) = \hat{\Gamma} V_{\mathrm{pl}}$, namely

$$V(\mathbf{x},z) = \varphi_0(\mathbf{x})\Gamma_0\left(\mathbf{x},\frac{\partial S}{\partial \mathbf{x}},z\right)e^{iS(\mathbf{x})/\hbar} \tag{2.45}$$

$$= \frac{A_0^0 e^{i\Phi_{\mathrm{B}}(\mathbf{x})}e^{iS(\mathbf{x})/\hbar}}{\sqrt{J(\mathbf{x})\varepsilon_{\mathrm{eff}}(\mathbf{x},\partial S/\partial \mathbf{x})|\partial S/\partial \mathbf{x}|}}\Gamma_0\left(\mathbf{x},\frac{\partial S}{\partial \mathbf{x}},z\right), \tag{2.46}$$

where Γ_0 is given by Eq. (2.23). As discussed in Sec. 2.1.2, the action $S(\mathbf{x})$ in this expression can be calculated from the Hamilton-Jacobi equation, or, equivalently, from integrating Hamilton's equations. We have thus found an expression for the full potential in real space, which is not only valid at $z = 0$, where the charged layer is situated, but also outside of it. It is valid for arbitrary $\varepsilon(\mathbf{x},z)$, provided that this function goes to a constant as $z \to \pm\infty$. Note that this assumption ensures that the potential decays exponentially for $z \to \pm\infty$.

We immediately see that the asymptotic solution (2.46) diverges at the points where $J(\mathbf{x}) = 0$. In a one-dimensional setting, these divergences are the so-called classical turning points, and Eq. (2.46) cannot be used in their vicinity. This general fact is well known for the one-dimensional WKB approximation for the Schrödinger equation, as discussed in Sec. 1.3.1 [61], where $J(x) \propto k(x)$. In the vicinity of the turning points, one normally constructs an approximate solution in terms of Airy functions [48, 61], see also the discussion in Ref. [50]. The additional factor $(\varepsilon_{\mathrm{avg}}|\partial S/\partial \mathbf{x}|)^{-1/2}$, on the other hand, is not commonly present in WKB-type approximations. In the next section, we give a physical interpretation to it with the help of the energy density of electromagnetic fields.

2.1.4 Interpretation through the energy density

Let us take a closer look at the physical interpretation to Eq. (2.46). We first note that Eq. (2.46) differs from semiclassical approximations to the solution of the Schrödinger equation, which have the form $\psi = A_0^0 \exp(iS/\hbar)/\sqrt{J}$, as discussed in Sec. 1.3.1. The Jacobian J ensures that the expression for the probability to find a particle in a certain region γ takes the same form in all coordinate systems, namely $\int_\gamma d\mathbf{x}/|J|$. In the context of the Schrödinger equation, the Jacobian therefore ensures the invariance of the probability integral $\int_\gamma d\mathbf{x}/|J|$ under coordinate changes, reflecting the concept of a half-density [55, 67], which ensures that probability density $|\psi|^2 \propto 1/|J|$ remains invariant under coordinate transformations. Clearly, Eq. (2.46) cannot be interpreted in the same way, since V is a different quantity: it is not a solution to the Schrödinger equation. Instead, it is related to an electrostatic potential Φ through $V = -e\Phi$. We may therefore consider the energy density stored in the electromagnetic field coming from the potential $V(\mathbf{x},z)$. In Ref. [1], this energy density was computed for the case where the background dielectric does not depend on z. It was shown that the electromagnetic energy density, when integrated along z, is proportional to $1/|J|$, that is, has the same mathematical structure as $|\psi|^2$ for the Schrödinger equation. This means that the integrated energy density indeed behaves as a density, and provides an additional physical interpretation of the potential $V(\mathbf{x}, z)$. Let us check whether the same conclusion holds for an arbitrary $\varepsilon(\mathbf{x}, z)$.

Since $V(\mathbf{x}, z, t)$ is a potential that corresponds to an electric field, we cannot use the interpretation in terms of probabilities for our plasmons. If we want to consider an equivalent quantity, we should look at the energy density of the electromagnetic field, which can be derived from Poynting's theorem [25, 77]. For a complex electric field $\boldsymbol{\mathcal{E}}$ and displacement field $\boldsymbol{\mathcal{D}}$, one has

$$-\nabla \cdot \boldsymbol{\mathcal{S}} = \frac{1}{16\pi} \left(\boldsymbol{\mathcal{E}} \cdot \frac{\partial \boldsymbol{\mathcal{D}}^*}{\partial t} + \boldsymbol{\mathcal{E}}^* \cdot \frac{\partial \boldsymbol{\mathcal{D}}}{\partial t} \right), \tag{2.47}$$

where $\boldsymbol{\mathcal{S}}$ is the Poynting vector. Note that we omitted the terms associated with the magnetic field, since they do not play a role in our problem. Moreover, we left out the terms proportional to $\boldsymbol{\mathcal{E}} \cdot \boldsymbol{\mathcal{D}}$ and $\boldsymbol{\mathcal{E}}^* \cdot \boldsymbol{\mathcal{D}}^*$, since they become zero after averaging over time. The energy density $\mathcal{U}$ is obtained [25, 77] by writing the right-hand side as a time derivative $\partial \mathcal{U}/\partial t$.

In our problem, the energy density $\mathcal{U}(\mathbf{x}, z)$ depends on both the in-plane and out-of-plane coordinates. In order to study the dependence on the in-plane coordinate $\mathbf{x}$, we integrate over z. We obtain the leading-order term of the electric field in Eq. (2.47) using $\boldsymbol{\mathcal{E}} = e^{-1} \nabla V$, where V is given by Eq. (2.46). We have to be more careful for the displacement field, since the dielectric function changes as a function of z. However, above and below the plane $z = 0$, we have a dielectric environment without dispersion, characterized by real and positive dielectric constants. Consequently, the displacement field is there simply given by $\boldsymbol{\mathcal{D}}(\mathbf{x}, z) = \varepsilon(\mathbf{x}, z) \boldsymbol{\mathcal{E}}(\mathbf{x}, z)$. The displacement field in the plane $z = 0$ is, however, much more complicated. Not only is there dispersion, but the dielectric function is also an operator. In appendix B, we derive a general formula for the energy density in this plane and show that it is in our case generally zero. The principal reason for this is that we are dealing with an infinitesimal layer.

Now that we have obtained both the electric and displacement fields in terms of the induced potential for the regions above and below the plane, we turn back to Eq. (2.47). Taking out the time derivative on the right-hand side, we find that the energy density is given by

$$\mathcal{U}(\mathbf{x}, z) = \frac{\varepsilon(\mathbf{x}, z)}{16\pi} \boldsymbol{\mathcal{E}}(\mathbf{x}, z) \cdot \boldsymbol{\mathcal{E}}^*(\mathbf{x}, z) = \frac{\varepsilon(\mathbf{x}, z)}{16\pi e^2} |\nabla V(\mathbf{x}, z)|^2 \tag{2.48}$$

where the gradient is three-dimensional, meaning that one should take the derivative with respect to both $\mathbf{x}$ and z. The leading-order term of the gradient of the induced potential (2.46) reads

$$\nabla V = \frac{A_0^0\, e^{i(S(\mathbf{x})-Et)/\hbar}}{\sqrt{J}\sqrt{\varepsilon_{\text{eff}}\left|\frac{\partial S}{\partial \mathbf{x}}\right|}} e^{-\frac{|z|}{\hbar}\left|\frac{\partial S}{\partial \mathbf{x}}\right|} \left(\frac{i}{\hbar}\frac{\partial S}{\partial \mathbf{x}}, -\frac{1}{\hbar}\left|\frac{\partial S}{\partial \mathbf{x}}\right|\right), \tag{2.49}$$

where the quantity in parentheses should be interpreted as a vector in the space $(\mathbf{x}, z)$. We remark that the term proportional to $|z|/\hbar$, coming from the derivative of the exponential factor with respect to $\mathbf{x}$ is not of leading order. As we discuss in the next subsection, the combination $z/\hbar$ is of order one when we introduce proper dimensionless units, see also Sec. 2.1.5.

In appendix C.2, we compute the leading-order term of the energy density (2.48) for the potential (2.46) and show that it satisfies

$$\mathcal{U}_{\text{I}}(\mathbf{x}) = \int_{-\infty}^{\infty} \mathcal{U}(\mathbf{x}, z) dz = \frac{1}{8\pi e^2} \frac{|A_0^0|^2}{|J(\mathbf{x})|}. \tag{2.50}$$

We observe that this expression does not depend on the momentum $|\partial S/\partial \mathbf{x}|$, and is also independent of the various ε. In fact, it is proportional to $1/|J|$, just as the probability density for the Schrödinger equation. We can therefore interpret the additional factor $(\varepsilon_{\text{avg}}|\partial S/\partial \mathbf{x}|)^{-1/2}$ in $V(\mathbf{x}, z, t)$ as follows: it ensures that the energy density has the correct mathematical structure of a density [55, 67] discussed at the beginning of this section. This result provides an additional physical interpretation of the potential $V(\mathbf{x}, z)$.

2.1.5 Dimensionless parameters and applicability of the semiclassical approximation

So far, we have constructed our semiclassical theory for plasmons in inhomogeneous media as a power series in $\hbar$. However, since $\hbar$ is a constant, it is, strictly speaking, not possible to use it as a small parameter in the series expansion. The main reason for this is that it is unclear with respect to which other quantity it should be small. This led to the apparent contradiction in the result of Sec. 2.1.2, namely the effective classical Hamiltonian (2.26), where we combined terms with different powers of $\hbar$ in $\mathcal{H}_0$. In order to resolve these issues, we should identify the correct dimensionless semiclassical parameter for the series expansion. In this section, we consider the characteristic scales in the problem, and define dimensionless quantities. We also discuss the applicability regime of the semiclassical approximation.

We start by identifying the characteristic length scales involved. Since we consider an inhomogeneous medium, the first scale is the characteristic length ℓ over which the inhomogeneity changes. The second length scale is the electron wavelength λ_{el}, since plasmons are collective excitations of electrons. It is given by $\lambda_{\text{el}} = 2\pi\hbar/p_0$, where p_0 is a typical value of the Fermi momentum in the system, cf. Eq. (1.49). Our semiclassical theory is applicable when $\lambda_{\mathrm{el}} \ll \ell$, which implies that the potential is locally almost homogeneous for the electrons [48, 50, 61]. We therefore introduce the dimensionless semiclassical parameter h, given by

$$h = \frac{\hbar}{p_0 \ell} = \frac{1}{2\pi} \frac{\lambda_{\mathrm{el}}}{\ell}. \tag{2.51}$$

The applicability criterion is then given by $h \ll 1$.

The third length scale is the Thomas-Fermi screening length [28, 33]. It characterizes the screening of the electric field of an electron by other electrons, and is given by $\lambda_{\mathrm{TF}} = \hbar^2 \varepsilon_{\mathrm{c}} / 2 g_{\mathrm{s}} m e^2$, where ε_{c} is a typical value of the effective background dielectric function $\varepsilon_{\mathrm{eff}}(\mathbf{x})$. Our second applicability criterion is that the characteristic length ℓ is much larger than this screening length [50]. We therefore define a second dimensionless parameter κ, given by

$$\kappa = \frac{\lambda_{\mathrm{TF}}}{\ell}. \tag{2.52}$$

We also require that $\kappa \ll 1$.

Now that we have identified the characteristic length scales, we introduce the dimensionless coordinates $\tilde{\mathbf{x}} = \mathbf{x}/\ell$, and momenta $\tilde{q} = |\mathbf{q}|/p_0$. Moreover, we also define the dimensionless Fermi momentum $\tilde{p}_{\mathrm{F}}(\tilde{\mathbf{x}}) = p_{\mathrm{F}}(\mathbf{x})/p_0$, background dielectric constant $\tilde{\varepsilon}(\tilde{\mathbf{x}}) = \varepsilon_{\mathrm{eff}}(\mathbf{x})/\varepsilon_{\mathrm{c}}$, and energy $\tilde{E} = 2mE/p_0^2$. We substitute the dimensionless parameters in our effective Hamiltonian Eq. (2.26), and calculate the integral of the polarization at zero temperature analogously to the homogeneous case in Sec. 1.2.3 [28]. This yields the dimensionless effective Hamiltonian

$$\begin{aligned}\tilde{\mathcal{H}}_0(\tilde{\mathbf{x}}, \tilde{q}) = 1 - \frac{h}{2\tilde{\varepsilon}\kappa} \frac{\tilde{p}_{\mathrm{F}}}{\tilde{q}^2} \Bigg(-\frac{\tilde{q}}{\tilde{p}_{\mathrm{F}}} - \operatorname{sign}\left(\operatorname{Re}(\tilde{\nu}_-)\right) \sqrt{\tilde{\nu}_-^2 - 1} \\ + \operatorname{sign}\left(\operatorname{Re}(\tilde{\nu}_+)\right) \sqrt{\tilde{\nu}_+^2 - 1} \Bigg),\end{aligned} \tag{2.53}$$

where $\tilde{\nu}_\pm = \tilde{E}/2\tilde{q}\,\tilde{p}_{\mathrm{F}} \pm \tilde{q}/2\tilde{p}_{\mathrm{F}}$, Looking at the analytic result of the polarization Eq. (1.33) in the homogeneous case, we see that the effective classical Hamiltonian is the same as the Lindhard dielectric function (1.30), but with spatially varying parameters. Furthermore, comparing this expression to Eq. (2.26), we see that it contains the ratio of the two small parameters h and κ. In contrast to the parameters themselves, this ratio is typically not small. This resolves the apparent contradiction, encountered in Sec. 2.1.2, regarding the different orders of $\hbar$. In Ch. 3, we calculate the ratio h/κ and show that it is of order one for the systems we consider. Note that the above parameters are in Gaussian units. We can convert them to S.I. units, by making the substitution $e^2 \to e^2/4\pi\varepsilon_0$.

So far, we only discussed the length scales related to the in-plane coordinates $\mathbf{x}$, which we called the slow variables in Sec. 2.1.1. It does not seem very sensible to make the out-of-plane coordinate z, previously called the fast variable, dimensionless using the length scale ℓ. After all, changes in the electron density $n(\mathbf{x}, z)$ happen over a characteristic length ℓ_z, which is much smaller than ℓ, as we discussed in the introduction. In our setup, one may even say that this length scale is infinitely small, since we are considering a delta function. However, in a more realistic setup, ℓ_z would correspond to the thickness of the two-dimensional electron gas, cf. Refs. [34, 78]. When introducing dimensionless units, we have to look at the combination $\hbar/(p_0 z)$, cf. Eq. (2.10) and Sec. 2.1.1. Rewriting this quotient as $\hbar/(p_0 \ell_z) \cdot \ell_z/z$, we see that both terms in the latter expression are of order one. In particular, the dimensionless parameter $\hbar/(p_0 \ell_z)$ is not small, because $\ell_z \ll \ell$. This is the reason why we included the derivatives with respect to z in the principal symbol in Sec. 2.1.1, and also why we used the ratio $z/\hbar$ in our Ansatz for Γ_1. Using the dimensionless quantities that we defined in this section, one can show that our semiclassical approximation is an asymptotic series expansion in powers of the dimensionless semiclassical parameter h, cf. the discussion in Ref. [50].

We end this subsection with a short analysis of the semiclassical parameter h, to get a better understanding for which electron densities the theory is applicable. Since the criterion for h reads $h = \hbar/p_0\ell \ll 1$, we can consider for which numerical values we have $h = 0.1$. As mentioned before, the parameter p_0 is directly related to the electron density by Eq. (1.49). We readily see that for a typical metal with electron density $n = 10^{14}$ cm^{-2}, we have $h = 0.1$ for a characteristic length of $\ell = 4$ nm. When we go to smaller densities, and for example consider a typical semiconductor with electron density around $n = 10^{11}$ cm^{-2}, we have $h = 0.1$ for a characteristic length scale of about 120 nm. Whilst the above parameters give an idea of the applicability regime of the theory, we emphasize that the semiclassical approximation generally gives good results outside of this regime as well, as discussed in the introduction.

2.2 Application of the theory to two distinct models

The general theoretical framework developed in the preceding sections is applicable to any static dielectric environment $\varepsilon(\mathbf{x}, z)$ that asymptotically approaches a constant as $z \to \pm\infty$, rendering it broadly applicable. In this section, we apply this general theory to two specific, yet distinct, models of the dielectric environment. First, we examine a simplified model, characterized by two uniform dielectric constants above and below the electron layer. While primarily serving as a demonstration of the theory's capabilities, this simplified model will also be employed in Ch. 3 to describe plasmon scattering. Second, we consider a more complex layered structure, where a thin film with dielectric constant ε_{M} and finite thickness d is sandwiched between two semi-infinite layers with dielectric constants ε_{A} and ε_{B}. This model accounts for nonlocal Coulomb screening from the substrate [34, 39, 40] and has been shown to accurately represent electrons in materials like TMDCs [34, 35]. The latter model will be employed in Chs. 4 and 5 to describe plasmon waveguides and plasmonic crystals.

We derive the effective classical Hamiltonian and the effective dielectric function ε_{eff} for both models. For the second, more complex model, we demonstrate that the derived dielectric function matches the established expression found in the literature [34, 40, 70], but with position-dependent parameters. Our objective extends beyond consistency checks, i.e., verifying that our formalism reproduces known results in homogeneous systems. More importantly, we aim to generalize this well-established result to inhomogeneous environments. Specifically, provided $\lambda_{\text{el}}/\ell \ll 1$, we can obtain the effective classical Hamiltonian $\mathcal{H}_0$ for inhomogeneous systems directly from the Lindhard dielectric function for homogeneous systems by substituting position-dependent variables for the constant parameters.

From the derivation in Sec. 2.1.2, we observe that the principal symbol $\Gamma_0(z)$ directly determines the effective classical Hamiltonian $\mathcal{H}_0$. In turn, $\Gamma_0(z)$ is constructed from the fundamental solutions w_1 and w_2 of the homogeneous differential equation (2.16). Utilizing the principal symbol $\Gamma_0(z)$, we can also compute the full leading-order potential $V(\mathbf{x}, z)$ and the energy density $\mathcal{U}_{\text{I}}(\mathbf{x})$, as detailed in Secs. 2.1.3 and 2.1.4.

2.2.1 Plasmons in a simple dielectric environment

We solve the homogeneous differential equation Eq. (2.16) to obtain an expression for the effective dielectric function ε_{eff}, and consequently the effective classical Hamiltonian. As we said before, we assume that our two-dimensional electron gas is located in the plane $z = 0$. For the first model, we consider the simple variation of the dielectric constant $\varepsilon(\mathbf{x}, z)$ that does not depend on z above (A) and below (B) this plane. However, the dielectric constant on either side might be different. We therefore have

$$\varepsilon(\mathbf{x}, z) = \varepsilon_i(\mathbf{x}) = \begin{cases} \varepsilon_{\text{A}}(\mathbf{x}), & z > 0 \\ \varepsilon_{\text{B}}(\mathbf{x}), & z < 0. \end{cases} \tag{2.54}$$

where the transition at $z = 0$ is instantaneous and behaves like a step function. We call the ε_i dielectric constants, since we assume that each layer has a locally well-defined static dielectric constant, which does not depend on momentum or energy. However, these dielectric constants can still vary spatially across $\mathbf{x}$, for instance because of a variation of different materials in each layer. Since we assume that the spatial scale ℓ of variations in the ε_i is much larger than the electron wavelength, these dielectric constants are well defined on the scale of the electron wavelength.

The two independent solutions to the homogeneous differential equation have the same form, namely

$$w_i(z) = \begin{cases} c_{A+}e^{|\mathbf{q}|z/\hbar} + c_{A-}e^{-|\mathbf{q}|z/\hbar}, & z > 0 \\ c_{B+}e^{|\mathbf{q}|z/\hbar} + c_{B-}e^{-|\mathbf{q}|z/\hbar}, & z < 0 \end{cases}, \tag{2.55}$$

where the constants are determined from the boundary conditions. We obtain these conditions directly from the differential equation (2.16): both w_i and $\varepsilon(\mathbf{x}, z)\partial w_i/\partial z$ have to be continuous at the boundaries $z = 0$. These boundary conditions are a reflection of the electrostatic boundary conditions [25] The first reflects that the potential is continuous on a boundary. The second condition is reminiscent of the electrostatic boundary condition that relates the discontinuity of the displacement field $\boldsymbol{\mathcal{D}}$ to a surface charge. However, with this method, the surface charge is taken into account separately via the method of variation of parameters in Eq. (2.18). In Ref. [1], this reflection is directly visible, where the system was solved with a different method, namely with the method of undetermined coefficients.

Since w_1 decays as $z \to \infty$, we have $c_{A+} = 0$ in Eq. (2.55). The remaining constants are determined by the boundary conditions. We note that, in principle, the fundamental solutions are not uniquely defined, since any linear combination of two fundamental solutions is again a fundamental solution. By demanding that w_1 (w_2) decays for $z \to \infty$ ($z \to -\infty$), we determine it uniquely up to normalization. We obtain

$$w_1(z) = \begin{cases} e^{-|\mathbf{q}|z/\hbar}, & z > 0 \\ \left(1 - \frac{1}{2}\frac{\varepsilon_A+\varepsilon_B}{\varepsilon_B}\right) e^{|\mathbf{q}|z/\hbar} + \frac{1}{2}\frac{\varepsilon_A+\varepsilon_B}{\varepsilon_B} e^{-|\mathbf{q}|z/\hbar}, & z < 0 \end{cases}. \tag{2.56}$$

For w_2, we have $c_{B-} = 0$, since it decays as $z \to -\infty$, which yields

$$w_2(z) = \begin{cases} \left(1 - \frac{1}{2}\frac{\varepsilon_A+\varepsilon_B}{\varepsilon_A}\right) e^{-|\mathbf{q}|z/\hbar} + \frac{1}{2}\frac{\varepsilon_A+\varepsilon_B}{\varepsilon_A} e^{|\mathbf{q}|z/\hbar}, & z > 0 \\ e^{|\mathbf{q}|z/\hbar}, & z < 0 \end{cases}. \tag{2.57}$$

Following the general theory discussed in Sec. 2.1.2, we find the principal symbol Γ_0 from Eq. (2.23), which simply yields

$$\Gamma_0(z) = \frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\Pi_0 \begin{cases} e^{-\frac{z}{\hbar}|\mathbf{q}|}, & z > 0 \\ e^{\frac{z}{\hbar}|\mathbf{q}|}, & z < 0 \end{cases}, \tag{2.58}$$

where

$$\varepsilon_{\text{eff}}(\mathbf{x}) = \frac{\varepsilon_A(\mathbf{x}) + \varepsilon_B(\mathbf{x})}{2}, \tag{2.59}$$

as defined by Eq. (2.24). We explicitly stated that the effective dielectric constant might depend on $(\mathbf{x})$, through variations in the substrate. Notice that $\varepsilon_{\text{eff}}(\mathbf{x})$, contrary to $\varepsilon(\mathbf{x}, z)$, does not depend on z. Having computed the effective dielectric function $\varepsilon_{\text{eff}}(\mathbf{x})$, we immediately obtain the effective classical Hamiltonian (2.26) and the full potential, given by Eq. (2.46).

In Ch. 3, we study the Hamiltonian, derived for this simple model of the dielectric environment in more detail, in order to get a physical understanding of this model. More specifically, we calculate the classical trajectories through Hamilton's equations for a plasmon scattered on a radially symmetric inhomogeneity.

2.2.2 Plasmons in layered structures with effective height *d*

For the second, more complicated model, we again consider a system where the electrons are bound to the 2D plane at $z = 0$. However, for this model the total dielectric environment as function of z takes the form

$$\varepsilon(\mathbf{x}, z) = \varepsilon_i(\mathbf{x}) = \begin{cases} \varepsilon_{\text{A}}(\mathbf{x}), & z > d/2 \\ \varepsilon_{\text{M}}(\mathbf{x}), & d/2 > z > -d/2\,, \\ \varepsilon_{\text{B}}(\mathbf{x}), & z < -d/2 \end{cases} \tag{2.60}$$

where now the transition at ±d is instantaneous and behaves like a step function.

The two independent solutions to the homogeneous differential equation have the same, more complicated, form, namely

$$w_i(z) = \begin{cases} c_{\text{A}+} e^{|\mathbf{q}|z/\hbar} + c_{\text{A}-} e^{-|\mathbf{q}|z/\hbar}, & z > d/2 \\ c_{\text{M}+} e^{|\mathbf{q}|z/\hbar} + c_{\text{M}-} e^{-|\mathbf{q}|z/\hbar}, & d/2 > z > -d/2\,, \\ c_{\text{B}+} e^{|\mathbf{q}|z/\hbar} + c_{\text{B}-} e^{-|\mathbf{q}|z/\hbar}, & z < -d/2 \end{cases} \tag{2.61}$$

where the constants are again determined from the boundary conditions: both w_i and $\varepsilon(\mathbf{x}, z) dw_i/dz$ have to be continuous at the boundaries $\pm d$.

Similar to the simple model, we construct w_1 explicitly. Since w_1 decays as $z \to \infty$, we have $c_{\text{A}+} = 0$ in Eq. (2.61). The remaining constants are determined by the boundary conditions. We obtain

$$\begin{aligned} w_1(z) = & \frac{1}{\left(1 + \tilde{\varepsilon}_{\text{B}} e^{-2|\mathbf{q}|d/\hbar}\right)\left(1 + \tilde{\varepsilon}_{\text{A}} e^{-2|\mathbf{q}|d/\hbar}\right)} \\ & \times \begin{cases} \left(1 + \tilde{\varepsilon}_{\text{A}}\right) e^{-|\mathbf{q}|z/\hbar}, & z > d/2 \\ \tilde{\varepsilon}_{\text{A}} e^{-2|\mathbf{q}|d/\hbar} e^{|\mathbf{q}|z/\hbar} + e^{-|\mathbf{q}|z/\hbar}, & d/2 > z > -d/2, \\ \frac{\tilde{\varepsilon}_{\text{A}} e^{-2|\mathbf{q}|d/\hbar} + \tilde{\varepsilon}_{\text{B}} e^{2|\mathbf{q}|d/\hbar}}{1 + \tilde{\varepsilon}_{\text{B}}} e^{|\mathbf{q}|z/\hbar} + \frac{1 + \tilde{\varepsilon}_{\text{A}} \tilde{\varepsilon}_{\text{B}} e^{-4|\mathbf{q}|d/\hbar}}{1 + \tilde{\varepsilon}_{\text{B}}} e^{-|\mathbf{q}|z/\hbar}, & z < -d/2 \end{cases} \end{aligned} \tag{2.62}$$

where $\tilde{\varepsilon}_i = (\varepsilon_{\text{M}} - \varepsilon_i)/(\varepsilon_{\text{M}} + \varepsilon_i)$. Note that the normalization factor, which is arbitrary, has no physical consequences. Looking at our expression (2.25) for ε_{eff}, we clearly see that it is divided out. Because of the symmetry, the second solution w_2 is then easily found by changing z to $-z$, and interchanging ε_{A} and ε_{B}.

Following the general theory discussed in Sec. 2.1.2, we find the principal symbol Γ_0 from Eq. (2.23), which yields

$$\Gamma_0(z)=\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\Pi_0\begin{cases}\frac{1+\tilde{\varepsilon}_{\text{A}}}{1+\tilde{\varepsilon}_{\text{A}}e^{-2\frac{d}{\hbar}|\mathbf{q}|}}e^{-\frac{z}{\hbar}|\mathbf{q}|}, & z>d/2\\ \frac{e^{-\frac{z}{\hbar}|\mathbf{q}|}+\tilde{\varepsilon}_{\text{A}}e^{-2\frac{d}{\hbar}|\mathbf{q}|}e^{\frac{z}{\hbar}|\mathbf{q}|}}{1+\tilde{\varepsilon}_{\text{A}}e^{-2\frac{d}{\hbar}|\mathbf{q}|}}, & <z<d/2\\ \frac{\tilde{\varepsilon}_{\text{B}}e^{-2\frac{d}{\hbar}|\mathbf{q}|}e^{-\frac{z}{\hbar}|\mathbf{q}|}+e^{\frac{z}{\hbar}|\mathbf{q}|}}{1+\tilde{\varepsilon}_{\text{B}}e^{-2\frac{d}{\hbar}|\mathbf{q}|}}, & -d/2<z<0\\ \frac{1+\tilde{\varepsilon}_{\text{B}}}{1+\tilde{\varepsilon}_{\text{B}}e^{-2\frac{d}{\hbar}|\mathbf{q}|}}e^{-\frac{z}{\hbar}|\mathbf{q}|}, & z<-d/2\end{cases}, \tag{2.63}$$

where

$$\varepsilon_{\text{eff}}(\mathbf{x})=\varepsilon_{\text{M}}(\mathbf{x})\frac{1-\tilde{\varepsilon}_{\text{A}}(\mathbf{x})\tilde{\varepsilon}_{\text{B}}(\mathbf{x})e^{-2|\mathbf{q}|d/\hbar}}{1+\left(\tilde{\varepsilon}_{\text{A}}(\mathbf{x})+\tilde{\varepsilon}_{\text{B}}(\mathbf{x})\right)e^{-|\mathbf{q}|d/\hbar}+\tilde{\varepsilon}_{\text{A}}(\mathbf{x})\tilde{\varepsilon}_{\text{B}}(\mathbf{x})e^{-2|\mathbf{q}|d/\hbar}}, \tag{2.64}$$

as defined by Eq. (2.24). Having computed the effective dielectric function ε_{eff}, we immediately obtain the effective classical Hamiltonian (2.26) and the full potential, given by Eq. (2.46). Compared to the simple model discussed in Sec. 2.2.1, where the effective dielectric function was simply the average of the dielectric constants above and below the 2D layer, this expression is more complex. Notably, it exhibits a dependence on the momentum $\mathbf{q}$, which becomes increasingly significant for small values of $|\mathbf{q}|$. This momentum dependence introduces nonlocal screening effects into the plasmon dynamics, reflecting the influence of the dielectric environment at distances comparable to the plasmon wavelength [34, 39, 40]

Equation (2.64) indeed corresponds to the well-known expression from the literature [34, 40, 70], but with position-dependent parameters. We have therefore extended this result to the inhomogeneous case. As previously discussed, the effective classical Hamiltonian can be viewed as the analog of the Lindhard function, with the parameters corresponding to their values at a given point $\mathbf{x}$. In the next section, we study this Hamiltonian in more detail, in order to get a physical understanding of this model. More specifically, we look at in-plane variations of various parameters, and show that these variations allow for the existence of plasmonic bound states.

2.3 Conclusion

In this chapter, we developed a comprehensive semiclassical theory for plasmons in spatially inhomogeneous two-dimensional media. We extended the three-dimensional formalism of Ref. [50] to two dimensions by employing a careful separation of in-plane and out-of-plane coordinates prior to applying the semiclassical Ansatz. This approach allows for the calculation of the full quantum plasmon dispersion within the random phase approximation, without restrictions on momentum or system size. The central assumption of our formalism is that the electron wavelength is significantly smaller than the length scale of the inhomogeneity, enabling the introduction of a dimensionless semiclassical parameter $h \ll 1$. As discussed in Sec. 2.1.5, this condition requires inhomogeneity length scales on the order of a few nanometers for metallic systems and a hundred nanometers for semiconductor systems, values readily achievable in experiments.

Building upon the general approach outlined above, we developed a detailed semiclassical theory for plasmons in inhomogeneous two-dimensional systems embedded within an arbitrary three-dimensional dielectric environment. The in-plane dynamics were treated semiclassically, subject to the limit of small h (for a comprehensive discussion of the semiclassical approximation and its applicability to plasmonic systems, refer to Refs. [1, 50]). In Sec. 2.1.1, we accomplished this by adiabatically decoupling the in-plane and out-of-plane variables using the method of operator separation of variables [56], and solving the resulting systems independently. Following the imposition of a self-consistency condition, we derived an expression for the effective classical Hamiltonian Eq. (2.15).

In Sec. 2.1.2, we demonstrated that the effective classical Hamiltonian incorporates the effective dielectric function Eq. (2.24) through the determinant of the matrix W, which bears resemblance to the Wronskian determinant. This effective dielectric function, which can be momentum-dependent, accounts for screening effects arising from the three-dimensional dielectric environment. The classical Hamiltonian is connected to the Hamilton-Jacobi equation, governing the dynamics of quantum plasmons in classical phase space. Consequently, the classical trajectories of plasmons in phase space can be derived from Hamiltonian's equations. Furthermore, it determines the classical action, $S(\mathbf{x})$, through the Hamilton-Jacobi equation, which in turn governs the phase of the induced potential, V_{pl}.

Subsequently, in Sec. 2.1.3, we derived the leading-order term (in the semiclassical parameter) of the induced potential Eq. (2.46) by solving the transport equation (2.31). This induced potential reveals the wave-like character of plasmons, and its derivative is proportional to the electric field. We utilized this relationship in Sec. 2.1.4 to compute the energy density, demonstrating its natural interpretation within the framework of the semiclassical approximation.

We applied this theory to two distinct models of the dielectric environment. First, we considered a simplified model characterized by uniform dielectric constants above and below the electron layer, primarily serving as a demonstration of our theory's capabilities. Second, we examined a widely used model for layered structures, as described in Refs. [34, 40, 70]. This model represents a thin film of thickness d and dielectric constant ε_M, sandwiched between two semi-infinite dielectric substrates. In both models, an infinitely thin 2D free electron layer was positioned at $z = 0$. Within our framework, the dielectric constants, film thickness, and electron density can all be treated as spatially varying parameters.

In conclusion, this chapter presented a semiclassical theory that offers a valuable new tool for investigating plasmons in spatially inhomogeneous two-dimensional media. We believe this framework can be effectively extended to model realistic experimental setups, as demonstrated in the following chapters, where we apply this theory to scattering, waveguiding, and periodic potentials.

3

Scattering on radially symmetric inhomogeneities

We demonstrate the applicability of the general theory developed in the previous chapter by considering the scattering of plasmons by a radially symmetric inhomogeneity. Specifically, we use the simple model for the dielectric environment (cf. Sec. 2.2.1) to illustrate how the theory can be applied to plasmon scattering. The effective classical Hamiltonian (2.26) is employed to calculate the classical trajectories of plasmons, which illustrates the general behavior of the plasmons in an intuitive way. Afterwards, the wave-like character of plasmons is incorporated through the induced potential (2.46), specifically via a phase shift that occurs for scattered plasmons. From this, we derive and compute numerical values for the total and differential scattering cross sections for a Gaussian-shaped inhomogeneity in the local electron density. We analyze how variations in the length scale ℓ and peak electron density of this inhomogeneity affect the scattering patterns, revealing phenomena such as forward scattering, interference effects, and backscattering.

This chapter details the full research and analysis from the second part of the paper published in Ref. [1]. Copyright by the American Physical Society.

3.1 Introduction to scattering

In the previous chapter, we developed a general semiclassical theory for plasmons in two-dimensional inhomogeneous media. In this chapter, we illustrate this theory by studying a scattering experiment. Specifically, we consider the simple model discussed in Sec. 2.2.1, namely a two-dimensional electron gas at $z = 0$, encapsulated by two homogeneous dielectric layers. The electron density in the $z = 0$ plane is also homogeneous, except near the $\mathbf{x}$-origin, where a radially symmetric inhomogeneity is placed, locally increasing or decreasing the electron density. We now proceed to demonstrate how our theory can be used by considering the scattering of plasmons by this radially symmetric inhomogeneity in the local charge density, which can be caused by defects, doping, or local gating. We consider the situation where a plasmon, represented as a plane wave, comes in from the left-hand side ($x \to -\infty$), and is scattered by the inhomogeneity. A preliminary understanding of the scattering can be gained by plotting the classical trajectories of this system, shown in Fig. 3.1, which are computed using the effective classical Hamiltonian (2.26). In this chapter, we incorporate the wave-like behavior through a scattering phase shift, to gain a deeper physical understanding of the scattering process.

In this chapter, we develop our scattering theory for plasmons, structured as follows. We begin by reviewing the relationship between the scattering cross section and the phase shift in Sec. 3.2, drawing from a quantum scattering theory perspective. This allows us to establish how experimentally measurable scattering cross sections can be expressed in terms of the phase shift. We then identify the phase shift of the scattered plasmons, crucial for calculating the total and differential scattering cross sections. In Sec. 3.3, we derive a semiclassical expression for the phase shift from our induced potential, constructing the formula for this phase shift within the semiclassical approximation. This derivation makes extensive use of the concept of a Lagrangian manifold [48, 55, 76, 79, 80], positioning it as a practical application of the abstract formalism of the Maslov canonical operator [48]. In Sec. 3.3.1, we recast this expression into a form suitable for numerical computations. After obtaining an expression for the scattering cross section, we present numerical values for plasmon scattering by a Gaussian bump or well in the charge density distribution of a metallic system. We compare these results with classical trajectories, highlighting the role of interference between different trajectories. Our numerical implementation is detailed in Sec. 3.4, where we present total and differential scattering cross sections for parameters indicative of a metallic system. We explore how variations in the length scale ℓ and peak electron density of this inhomogeneity lead to diverse scattering phenomena, including forward scattering, interference effects, and backscattering, providing a comprehensive understanding of plasmon scattering in radially symmetric systems.

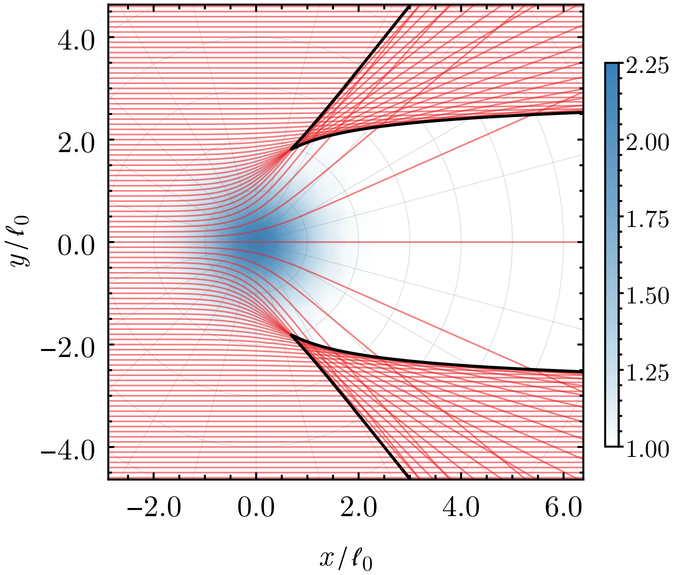


Figure 3.1: Classical trajectories of a plasmon, coming in from $x \to -\infty$. The maximum change in local density is $\delta n = +1.25$. The blue shading indicates the region where the electron density is increased. The solid black lines are the caustics.

Our description is based on the in-plane induced potential V_{pl}, which we denote by V throughout this section in order to simplify the notation. This notational simplification is justified and should not cause confusion because, as established in the previous chapter, $V(z=0) = V_{\text{pl}}$, and for scattering phenomena, we are primarily concerned with the plasmon behavior within the $z = 0$ plane. We note that an alternative method to describe plasmon scattering has been developed in Ref. [47]. This approach leverages the analogy between the RPA equations for the electrostatic potential and the Lippmann-Schwinger formalism, a cornerstone of conventional scattering theory [81]. While the formalism presented in Ref. [47] offers a broader scope of applicability, our approach is specifically tailored to facilitate the construction of a semiclassical description. By focusing on the classical trajectories and incorporating quantum effects through the phase shift, we gain an intuitive understanding of plasmon scattering dynamics, particularly in the context of radially symmetric inhomogeneities. This allows us to directly connect the classical paths to observable scattering patterns, providing a clear physical interpretation of the scattering process.

3.2 Scattering cross sections and phase shifts

Before we define the phase shift, we first have to carefully set the scene. In order to obtain the relation between the velocity and the momentum, we consider the Hamiltonian system Eqs. (2.27). A short calculation yields

$$\mathcal{H}_0 \approx 1 - \frac{g_s e^2 p_F^2 |\mathbf{q}|}{2m\varepsilon_{\text{avg}}\hbar E^2}, \quad \frac{d\mathbf{x}}{d\tau} = \frac{\partial \mathcal{H}_0}{\partial \mathbf{q}} \approx -\frac{g_s e^2 p_F^2}{2m\varepsilon_{\text{avg}}\hbar E^2}\frac{\mathbf{q}}{|\mathbf{q}|}, \tag{3.1}$$

where we considered the limit of small $\mathbf{q}$. This counter-intuitive result shows that the (classical) velocity of the plasmon is in the opposite direction of the momentum $\mathbf{q}$. Note that it is sufficient to calculate the velocity in the limit of small $\mathbf{q}$, since it is not expected to change sign for larger $\mathbf{q}$. A right-moving plasmon, propagating along the x axis from negative infinity, thus corresponds to a plane wave with momentum $q_x = -|\mathbf{q}| < 0$ parallel to the x axis, see also Fig. 3.1.

In the remainder of this section, we derive the formulas that express the total and differential scattering cross section in terms of the phase shift. Readers who are familiar with this subject may directly skip to the result Eq. (3.8).

We first take a closer look at the incoming plasmon, which is described by a plane wave, namely,

$$V(r,\theta) = A e^{-i\frac{|\mathbf{q}|x}{\hbar}} = A e^{-i\frac{|\mathbf{q}|r}{\hbar}\cos\theta}, \tag{3.2}$$

where the amplitude A is constant, we used that $q_x = -|\mathbf{q}|$ and introduced polar coordinates in the last step. Because the inhomogeneity is radially symmetric, we expand this incoming plane wave in radial waves. Using the Fourier-Bessel series [82], we obtain

$$V(r,\theta) = A \sum_{m=-\infty}^{\infty} (-1)^m i^m J_m\left(\frac{|\mathbf{q}|r}{\hbar}\right) e^{im\theta}, \tag{3.3}$$

where J_m is the Bessel function of the first kind and where the factor $(-1)^m$ arises because the argument of the exponent is negative. We subsequently rewrite each Bessel function as the sum of two Hankel functions, which represent incoming and outgoing waves. Because we are interested in the far-field regime $r/\ell \gg 1$, far away from the inhomogeneity, we can use the asymptotic expansions of the Hankel functions [82] to obtain

$$V(r,\theta) = A\sqrt{\frac{\hbar}{2\pi|\mathbf{q}|r}} \sum_{m=-\infty}^{\infty} e^{-im\frac{\pi}{2}-i\frac{\pi}{2}} e^{im\theta}\left(e^{i\frac{|\mathbf{q}|r}{\hbar}+i\frac{\pi}{4}-im\frac{\pi}{2}} - e^{-i\frac{|\mathbf{q}|r}{\hbar}-i\frac{\pi}{4}+im\frac{\pi}{2}}\right), \tag{3.4}$$

where the first and second terms represent radially incoming and outgoing waves, respectively.

Our next step is to take a closer look at the outgoing wave. Far away from the inhomogeneity ($r/\ell \gg 1$), the scattered plasmon is described by

$$V^{\text{scat}}(r,\theta) \approx A \left(e^{-i\frac{|\mathbf{q}|r}{\hbar}\cos\theta} + g(\theta)\frac{e^{-i\frac{|\mathbf{q}|r}{\hbar}}}{\sqrt{r}} \right). \tag{3.5}$$

The first term in this expression comes from the incoming wave Eq. (3.2), and the second term represents the outgoing radial wave in the far field. This asymptotic form can be understood as a extension of the Sommerfeld radiation condition for the Helmholtz equation [83–85]. Our goal is to find the scattering amplitude $g(\theta)$, which expresses the amplitude of the scattered wave scattered in every direction θ.

It is common to express the effect of the inhomogeneity using a phase shift δ_m [61]. By definition, twice this phase shift is added to the outgoing radial components in expression (3.4). Here we use a minus sign, mainly because it simplifies our results. On a more fundamental level, one may think that we need this minus sign because $\exp(-i|\mathbf{q}|r/\hbar)$ is the outgoing wave in our problem, instead of $\exp(i|\mathbf{q}|r/\hbar)$. We therefore define

$$\begin{aligned} V^{\text{scat}}(r,\theta) = A\sqrt{\frac{\hbar}{2\pi|\mathbf{q}|r}} \sum_{m=-\infty}^{\infty} e^{-i\frac{m\pi}{2}-i\frac{\pi}{2}} e^{im\theta} \\ \times \left(e^{i\frac{|\mathbf{q}|r}{\hbar}+i\frac{\pi}{4}-i\frac{m\pi}{2}} - e^{-i\frac{|\mathbf{q}|r}{\hbar}-i\frac{\pi}{4}+i\frac{m\pi}{2}-2i\delta_m} \right). \end{aligned} \tag{3.6}$$

This definition shows that the phase shift is defined up to an integer multiple of π. For future reference, we note that each of the terms in the series (3.6) can be rewritten in terms of a sine. We have

$$V_m^{\text{scat}}(r,\theta) \propto \sin\left(\frac{|\mathbf{q}|r}{\hbar} + \delta_m - \frac{m\pi}{2} + \frac{\pi}{4} \right), \tag{3.7}$$

where we left out the factors of proportionality.

We remark that in our scattering setup both the incoming wave and the inhomogeneity possess mirror symmetry around the $y = 0$ axis. The scattered wave (3.6) should therefore have the same symmetry. Requiring that $V(r,\theta) = V(r,-\theta)$ and reversing the summation order ($m \to -m$) on the right hand side, we find that the phase shifts with opposite m have to be related by $\delta_{-m} = \delta_m + n_m\pi$, where n_m is an arbitrary integer. Since the phase shift is only defined up to an integer multiple of π, we may set $\delta_{-m} = \delta_m$.

Comparing expressions (3.5) and (3.6), we can now obtain an expression for the scattering amplitude $g(\theta)$. Since the differential cross section is equal to the square of the absolute value of this scattering amplitude [61], we find

$$\frac{d\sigma}{d\theta} = |g(\theta)|^2 = \left| e^{-i\frac{\pi}{4}} \frac{\sqrt{2\hbar}}{\sqrt{\pi|\mathbf{q}|}} \sum_{m=-\infty}^{\infty} \sin\delta_m e^{im\theta - i\delta_m} \right|^2 . \tag{3.8}$$

Integrating over all angles θ gives the total scattering cross section, specifically

$$\sigma = \int |g(\theta)|^2 d\theta = \frac{4\hbar}{|\mathbf{q}|} \sum_{m=-\infty}^{\infty} \sin^2\delta_m, \tag{3.9}$$

where we used the orthogonality of the azimuthal terms.

At the end of this section, we briefly come back to the dimensionless parameters we introduced in the previous section. Since the scattering cross section has the dimensions of length, we can make it dimensionless by dividing by the characteristic length scale, i.e., $\tilde{\sigma} = \sigma/\ell$. This yields

$$\tilde{\sigma} = \frac{4h}{\tilde{q}} \sum_{m=-\infty}^{\infty} \sin^2\delta_m, \tag{3.10}$$

where we used the definitions for h and $\tilde{q}$ from Sec. 2.1.5.

3.3 Derivation of the semiclassical phase shift

In the previous subsection, we showed that the phase shift is sufficient to calculate both the differential and the total scattering cross section. In this subsection, we obtain a semiclassical expression for this phase shift. To this end, we first construct an asymptotic solution using the semiclassical Ansatz, and subsequently compare it to the solution (3.6), which was derived using general quantum mechanical principles.

Our starting point is the asymptotic solution (2.46), which is based on the classical trajectories. In the construction of this expression, we implicitly assumed that to each point $\mathbf{x}$ corresponds a single trajectory. However, this is generally not the case, because multiple electron trajectories may arrive at the same point, cf. Fig. 3.1. Physically, this is nothing but the well-known phenomenon of interference. We obtain the full asymptotic solution by adding the contributions of the individual trajectories [48, 55], that is,

$$V^{\mathrm{SC}}(\mathbf{x}) = \sum_j \frac{A_0^0}{\sqrt{\varepsilon_{\mathrm{avg}} \left| \frac{\partial S_j}{\partial \mathbf{x}} \right|}} \frac{1}{\sqrt{|J_j(\mathbf{x})|}} e^{-i\frac{\pi}{2}\mu_j} e^{\frac{i}{\hbar}S_j(\mathbf{x})}, \tag{3.11}$$

where S_j is the action along the j-th trajectory. The object μ_j is the Maslov index [48, 55, 73], which expresses the complex phase of the Jacobian J_j and will be discussed in more detail later on. Formally, this solution can be obtained with the so-called Maslov canonical operator [48].

As previously mentioned, Fig. 3.1 shows the classical trajectories that arise from the incoming plane wave (3.2). In principle, we could use Eq. (3.11) to construct an asymptotic solution for our scattering problem based on these trajectories, see Ref. [79] for an example. However, as clearly explained in Ref. [86], it is not at all straightforward to relate this asymptotic solution to Eq. (3.6), which takes the form of a series of radially incoming and outgoing waves labeled by the index m.

We therefore construct the full asymptotic solution in a different way. Keeping the asymptotic expression (3.6) in mind, we start by considering radially symmetric incoming plane waves. In order to compute the classical trajectories, we first write down the initial conditions, which form a one-dimensional surface Λ^1 in phase space. Since we consider radially symmetric waves, Λ^1 is a circle:

$$\Lambda^1 = \{r = r_0,\ \theta = \alpha,\ q_r = q_{r,0},\ q_\theta = q_{\theta,0}\}, \tag{3.12}$$

where the parameter $\alpha \in [0, 2\pi)$ corresponds to the angle of incidence. Although one should theoretically consider $r_0 \to \infty$, a numerical computation requires a finite r_0, and we assert that $r_0/\ell \gg 1$. Moreover, note that $q_r = q_{r,0} > 0$ gives rise to an incoming wave, cf. Eq. (3.1). The total momentum is given by $|\mathbf{q}_0|^2 = (q_{r,0})^2 + (q_{\theta,0}/r_0)^2$, and is equal to $q_{r,0}$ at large distances.

Next, we consider the time evolution of each point on Λ^1, shown as a black circle in Fig. 3.2, by the Hamiltonian system $d\mathbf{x}/dt = \partial\mathcal{H}_0/\partial\mathbf{p}$, $d\mathbf{p}/dt = -\partial\mathcal{H}_0/\partial\mathbf{x}$ with the effective classical Hamiltonian $\mathcal{H}_0$, see Eq. (2.26). This time evolution gives rise to the two-dimensional surface Λ^2, shown in orange in Fig. 3.2. Each point on this surface can be parameterized by the angle of incidence α and the time τ, i.e.,

$$\begin{aligned}\Lambda^2 = \{&r = r(\tau,\alpha), \theta = \theta(\tau,\alpha), q_r = q_r(\tau,\alpha),\\ &q_\theta = q_\theta(\tau,\alpha) \mid \alpha \in [0, 2\pi), \tau \in [0,\infty)\}.\end{aligned} \tag{3.13}$$

With a more detailed analysis, one can show that this surface is a so-called Lagrangian manifold [48, 55, 76, 79, 80]. An important property of such a manifold is that the action integral (2.28) is path independent, that is, its outcome only depends on the initial and final points on Λ^2, and not on the specific path that is used to compute the integral.

Looking at Fig. 3.2, one sees that Λ^2 consists of two distinct leaves. In other words, when one projects Λ^2 onto the coordinate plane, each point on the coordinate plane corresponds to two points on Λ^2. The upper leaf corresponds to the incoming wave, since it consists of points with radial momentum $q_r^{\mathrm{inc}} > 0$. Similarly, the lower leaf corresponds to the outgoing wave, since it consists of points with $q_r^{\mathrm{out}} < 0$. The two leaves join at the point $r_{\mathrm{c}}(q_\theta)$, where $q_r = 0$. This point is a classical turning point, as can be inferred

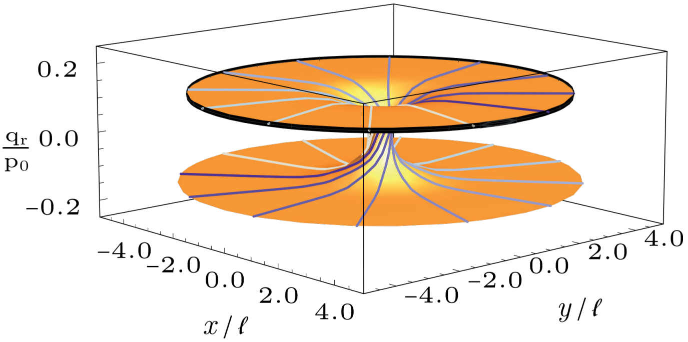


Figure 3.2: Schematic representation of the Lagrangian manifold Λ^2 in orange. The surface Λ^1 is depicted by the black circle. The purple, blue and grey colored lines represent plasmon trajectories, coming in from Λ^1 with momentum q_r^0.

from Eq. (3.1). Moreover, one can show that $q_r^2 \propto r - r_\mathrm{c}$ in the vicinity of r_c, which implies that we are dealing with a simple turning point [48, 75, 76, 86].

The proof of this statement is simplest when there is no inhomogeneity at all, in other words, when the system is homogeneous, since in this case both q_θ and $|\mathbf{q}|^2 = q_r^2(r) + q_\theta^2/r^2$ are constants of motion. Provided that $q_\theta \neq 0$, one has $r_\mathrm{c} = |q_\theta|/|\mathbf{q}|$, and a Taylor expansion of $|\mathbf{q}|^2 = q_r^2(r) + q_\theta^2/r^2$ around r_c readily gives the result. When there is a radially symmetric inhomogeneity present, q_θ is still a constant of motion, since the effective classical Hamiltonian $\mathcal{H}_0$ does not depend on θ. When q_θ is small, one can consider the Taylor expansion (3.1). Solving the equation $\mathcal{H}_0 = 0$ for the energy E leads to the familiar square-root dispersion relation

$$E \approx \sqrt{\frac{g_\mathrm{s} e^2 p_\mathrm{F}(r)^2}{2m\varepsilon_\mathrm{avg}\hbar}|\mathbf{q}|} = \sqrt{\frac{2\pi e^2 \hbar n(r)}{m\varepsilon_\mathrm{avg}}|\mathbf{q}|}, \tag{3.14}$$

where $p_\mathrm{F}(r)$ and $n(r)$ depend on the radial coordinate, and we used that $n(r) = g_\mathrm{s} p_\mathrm{F}^2(r)/(4\pi\hbar^2)$ in the last equality. The turning point is subsequently defined by the relation $E^2 = 2\pi e^2 \hbar n(r)|q_\theta|/(m\varepsilon_\mathrm{avg} r_\mathrm{c})$, since it corresponds to $q_r = 0$. Performing a Taylor expansion in r around r_c then gives

$$q_r^2 \approx \frac{2q_\theta^2}{r_\mathrm{c}^2}\left(\frac{1}{r_\mathrm{c}} - \frac{n'(r_\mathrm{c})}{n(r_\mathrm{c})}\right)(r - r_\mathrm{c}), \tag{3.15}$$

which indeed shows that $q_r^2 \propto r - r_\mathrm{c}$. For larger values of $|q_\theta|$, numerical methods confirm that this proportionality remains valid. A more general proof of this result is provided in appendix D. Furthermore, this result is consistent with the general theory of caustics and singularities for effectively one-dimensional geometries [75, 76].

Now that we have studied the classical trajectories and the Lagrangian manifold Λ^2, we can construct the full asymptotic solution. Since we started with radially symmetric incoming plane waves, we parameterize the Cartesian plane with polar coordinates, i.e., $\mathbf{x} = \mathbf{x}(r, \theta)$. As expressed by Eq. (3.11) and mentioned previously, we obtain the induced potential at the point $\mathbf{x}(r, \theta)$ by adding the contributions of the two points on Λ^2 that are projected onto $\mathbf{x}(r, \theta)$. Note that this construction cannot be used in the vicinity of the turning point, which is a singular point of the projection. We therefore limit ourselves to the far-field regime $r/\ell \gg 1$.

We first consider the action (2.28), Since it is given by a line integral over a path on Λ^2, we can consider the action as a function of the coordinates (τ, α) on Λ^2. The action $S_j(r, \theta)$ on a given leaf follows from this more general quantity by projection. Switching to polar coordinates, and using the definitions of the polar momenta, i.e., $q_x = \cos\theta\, q_r - \frac{\sin\theta}{r} q_\theta$, and $q_y = \sin\theta\, q_r + \frac{\cos\theta}{r} q_\theta$, cf. Refs. [74, 76], we have

$$S(\tau, \alpha) = \int_{\mathcal{C}} \langle q_{\mathbf{x}}, d\mathbf{x} \rangle = \int_{\mathcal{C}} q_r dr + q_\theta d\theta. \tag{3.16}$$

The integral is taken over the line $\mathcal{C}$ with starting point, the so-called central point, (τ_0, α_0) [48] and end point (τ, α). For convenience, we set $\tau_0 = \alpha_0 = 0$ from here on.

As we previously mentioned, the integral (3.16) is independent of the specific path, since Λ^2 is a Lagrangian manifold [48, 55, 76, 79, 80]. Given a point (r, θ), we therefore split the integration into two parts: we first integrate along Λ^1 to the initial point of the trajectory on which the point (r, θ) lies, and then proceed the integration along the trajectory. Figure 3.2 shows that the two trajectories that contain the point (r, θ), corresponding to the two different leaves of Λ^2, and hence to the incoming and the outgoing waves, originate from two different angles $\theta_1 = \alpha_1$ and $\theta_2 = \alpha_2$. We therefore compute the integral separately for both trajectories. For the trajectory corresponding to the incoming wave, we have

$$\begin{aligned} S_1(\tau, \alpha) &= \int_{(0,0)}^{(0,\alpha_1)} (q_r dr + q_\theta d\theta) + \int_{(0,\alpha_1)}^{(\tau,\alpha)} (q_r dr + q_\theta d\theta), \\ S_1(r, \theta) &= q_\theta \theta_1 + q_\theta(\theta - \theta_1) + \int_{r_0}^{r} q_r^{\text{inc}}(r') dr', \end{aligned} \tag{3.17}$$

where r_0 was defined in Eq. (3.12). We used that q_θ is constant because $\mathcal{H}_0$ is independent of θ due to the radial symmetry, and immediately see that the terms containing θ_1 drop out.

In a similar way, we can compute the action for the trajectory corresponding to the outgoing wave. Since it has passed the turning point r_c, there is an additional radial contribution. We find

$$S_2(r, \theta) = q_\theta \theta_2 + q_\theta(\theta - \theta_2) + \int_{r_0}^{r_c} q_r^{\text{inc}}(r') dr' + \int_{r_c}^{r} q_r^{\text{out}}(r') dr', \tag{3.18}$$

where the terms containing θ_2 cancel as before.

At this point, we have to consider an additional constraint: the action $S(\tau, \alpha)$ should be single-valued, since Λ^2 is a Lagrangian manifold. When we consider the action along a circular path that goes around the central hole in Λ^2, it should therefore equal a multiple of 2π. This is nothing but an expression of the Bohr-Sommerfeld quantization condition [48], see also Refs. [50, 87] for similar applications. We have

$$\int_0^{2\pi} q_\theta d\theta = 2\pi m\hbar, \tag{3.19}$$

where m is the (integer) azimuthal quantum number. This quantization condition determines the values which q_θ can take, and shows that $q_\theta = m\hbar$. Inserting this result into the expressions for S_1 and S_2, and considering Eq. (3.11), we see that the angular dependence of our asymptotic solution is given by $\exp(im\theta)$, in accordance with the quantum mechanical results discussed in Sec. 3.2.

Our next step is to determine the Jacobian in Eq. (3.11). In general, it is given by

$$J = \det \begin{pmatrix} \frac{\partial x}{\partial \tau} & \frac{\partial x}{\partial \alpha} \\ \frac{\partial y}{\partial \tau} & \frac{\partial y}{\partial \alpha} \end{pmatrix} = \det \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} \det \begin{pmatrix} \frac{\partial r}{\partial \tau} & \frac{\partial r}{\partial \alpha} \\ \frac{\partial \theta}{\partial \tau} & \frac{\partial \theta}{\partial \alpha} \end{pmatrix}, \tag{3.20}$$

where the second equality follows from our parametrization in polar coordinates. The first determinant on the right-hand side is the usual Jacobian associated with the transformation to polar coordinates and equals r. In order to compute the second Jacobian, we first consider its value on Λ^1. Since $\partial r/\partial\alpha = 0$ and $\partial\theta/\partial\alpha = 1$ on Λ^1, we have

$$J = r\frac{\partial r}{\partial \tau}. \tag{3.21}$$

Using the variational system for Hamilton's equations [80], see also Refs. [50, 79], one can show that the time derivatives of $\partial r/\partial\alpha$ and $\partial\theta/\partial\alpha$ equal zero on Λ^2. The result (3.21) is therefore valid on all of Λ^2. Using Hamilton's equations, we see that the Jacobian can also be written as

$$J = r\frac{\partial \mathcal{H}_0}{\partial q_r} = r\frac{\partial \mathcal{H}_0}{\partial |\mathbf{q}|}\frac{q_r}{|\mathbf{q}|}, \tag{3.22}$$

where we have used that the effective Hamiltonian $\mathcal{H}_0$ is a function of $|\mathbf{q}|$ only, and not of its components.

At a given point (r, θ), the incoming and outgoing waves have opposite radial momenta, $q_r^{\text{inc}}(r) = -q_r^{\text{out}}(r)$, see also Fig. 3.2. Equation (3.22) then shows that $J_1 = -J_2$. Since Eq. (3.11) contains the absolute value of the Jacobian, we obtain the same factor for the contribution of each leaf. The sign of the Jacobian is, nevertheless, taken into account through the Maslov index [48, 55, 80, 88]. On the upper leaf, which corresponds to the incoming waves, the sign of the Jacobian is equal to its sign on Λ^1, and we set $\mu_1 = 0$. On the lower leaf, which corresponds to outgoing waves, the Jacobian has the opposite sign. The Maslov index now regulates the analytic continuation of the square root of this Jacobian. Computations performed explicitly in Ref. [50], cf. Ref. [80], show that $\mu_2 = -1$ for points on the lower leaf.

Lastly, we calculate the factor $|\partial S/\partial \mathbf{x}|$ in Eq. (3.11), which is part of the amplitude. In polar coordinates, we have

$$\left|\frac{\partial S}{\partial \mathbf{x}}\right| = \sqrt{\left(\frac{\partial S}{\partial r}\right)^2 + \left(\frac{1}{r}\frac{\partial S}{\partial \theta}\right)^2} = \sqrt{q_r^2(r) + \left(\frac{q_\theta}{r}\right)^2}, \tag{3.23}$$

where the latter expression is to be understood on the Lagrangian manifold Λ^2. Since $q_r^{\mathrm{inc}}(r) = -q_r^{\mathrm{out}}(r)$, this factor is equal for both contributions to the sum (3.11).

We are now ready to combine all ingredients and compute the asymptotic solution (3.11). As a final preparatory step, we rewrite the integral from 0 to r in our expression (3.17) for S_1 as the sum of an integral from 0 to r_{c} and an integral from r_{c} to r. Putting everything together, we obtain

$$\begin{aligned} V_m^{\mathrm{SC}}(r,\theta) = {} & \frac{A^0_{0,m}}{\varepsilon_{\mathrm{avg}}^{1/2}\left((q_r^{\mathrm{inc}})^2 + (\frac{q_\theta}{r})^2\right)^{1/4}} \frac{1}{\sqrt{r\left|\frac{\partial \mathcal{H}_0}{\partial |\mathbf{q}|}\frac{q_r^{\mathrm{inc}}}{|\mathbf{q}|}\right|}} \\ & \times \left(e^{\frac{i}{\hbar}\int_{r_{\mathrm{c}}}^{r} q_r^{\mathrm{inc}} dr' + i\frac{\pi}{4}} - e^{-\frac{i}{\hbar}\int_{r_{\mathrm{c}}}^{r} q_r^{\mathrm{inc}} dr' - i\frac{\pi}{4}}\right) \\ & \times e^{im\theta} e^{\frac{i}{\hbar}\int_{r_0}^{r_{\mathrm{c}}} q_r^{\mathrm{inc}} dr' - i\frac{\pi}{4}}, \end{aligned} \tag{3.24}$$

where we omitted the arguments of q_r and used that $q_r^{\mathrm{out}} = -q_r^{\mathrm{inc}}$. We also added an index m, because we constructed an asymptotic solution for a radially incoming wave with angular momentum $q_\theta = m\hbar$.

Comparing the asymptotic solution (3.24) with the m-th component of the general solution (3.6), we observe that they exhibit the same asymptotic behavior. First of all, their angular dependence is the same, as they are both proportional to $\exp(im\theta)$. Second, they both decay as $1/\sqrt{r}$ in the far field, since q_r becomes constant for $r/\ell \gg 1$.

At this point, we can obtain an asymptotic solution for the original plane wave, see Fig. 3.1, by considering a series of incoming plane waves and matching the coefficients $A^0_{0,m}$ with the constants in front of the series in Eq. (3.6). However, this is not at all necessary, since we previously established that we can express the scattering cross section in terms of the phase shift δ_m. A semiclassical expression for this phase shift can be directly determined by rewriting the asymptotic solution (3.24) in the form of a sine, namely,

$$V_m^{\mathrm{SC}}(r,\theta) \propto \sin\left(\frac{1}{\hbar}\int_{r_{\mathrm{c}}}^{r} q_r^{\mathrm{inc}}(r')dr' + \frac{\pi}{4}\right). \tag{3.25}$$

Comparing this result with Eq. (3.7), we immediately obtain

$$\delta_m = \lim_{r\to\infty}\left(\int_{r_{\mathrm{c}}}^{r}\frac{q_r^{\mathrm{inc}}(r')}{\hbar}dr' - \frac{q_\infty r}{\hbar}\right) + \frac{m\pi}{2}, \tag{3.26}$$

where we used q_∞ to denote the value of $|\mathbf{q}|$ at infinity where the presence of the inhomogeneity is no longer felt. By taking the limit, we can rewrite this expression as

$$\delta_m = \int_{r_{\mathrm{c}}}^{\infty}\left(\frac{q_r^{\mathrm{inc}}(r')}{\hbar} - \frac{q_\infty}{\hbar}\right)dr' - \frac{q_\infty r_{\mathrm{c}}}{\hbar} + \frac{m\pi}{2}. \tag{3.27}$$

Although this result for the semiclassical phase shift has the same form as most results in scientific literature, see e.g. Ref. [86], there is an important difference. Most semiclassical expressions for the phase shift that are commonly found in the literature are derived by first performing separation of variables in the two-dimensional differential equation and subsequently constructing an asymptotic solution for the remaining one-dimensional equation. On the other hand, we constructed an asymptotic solution for a two-dimensional pseudodifferential equation, where we carefully accounted for the contributions of the different trajectories on the Lagrangian manifold Λ^2. An added advantage of our approach is that there is no need to perform an explicit Langer substitution Ref. [63, 89]. Instead, the correct expression naturally arises from the quantization of the azimuthal variable, cf. Ref. [50].

3.3.1 Alternative expression for the phase shift

Unfortunately, expression (3.26) for the scattering phase is not very convenient when one wants to compute the phase shift for a given system. If $|\mathbf{q}|$ is the solution of $\mathcal{H}_0(\mathbf{x},\mathbf{q}) = 0$, then $q_r^{\mathrm{inc}}(r) = (|\mathbf{q}|^2(r) - m^2\hbar^2/r^2)^{1/2}$ for a given value of m. The integral in expression (3.26) therefore always converges very slowly (because of the second term in q_r^{inc}), even when $|\mathbf{q}|(r)$ rapidly becomes constant. Hence, we need to use a large cutoff radius in a numerical implementation. In this subsection, we use a trick from Ref. [64] to obtain an expression for δ_m that is more suitable for practical calculations.

To this end, we consider a second system, for which $n^{(0)}$ and $\varepsilon_{\mathrm{avg}}$ are constant, and correspond to the values far away from the inhomogeneity in the first system. In this second system the momentum q is constant, and equal to $q_\infty > 0$ which was defined in the previous subsection. The radial momentum of the incoming wave is then given by $q_r^0(r) = (q_\infty^2 - m^2\hbar^2/r^2)^{1/2}$ and vanishes at the classical turning point $r_{\mathrm{c}}^0 = |m|\hbar/q_\infty$. We now rewrite expression (3.26) as

$$\delta_m = \lim_{r\to\infty}\left(\int_{r_{\mathrm{c}}}^{r}\frac{q_r^{\mathrm{inc}}(r')}{\hbar}dr' - \int_{r_c^0}^{r}\frac{q_r^0(r')}{\hbar}dr'\right) + \lim_{r\to\infty}\left(\int_{r_c^0}^{r}\frac{q_r^0(r')}{\hbar}dr' - \frac{q_\infty r}{\hbar}\right) + \frac{m\pi}{2}, \quad (3.28)$$

where we were allowed to split the limit into two because both converge.

Using our expression for $q_r^0(r)$, the integral in the second limit can be computed explicitly [64]. Taking the limit $r \to \infty$ in the result, we find

$$\lim_{r\to\infty}\left(\int_{r_c^0}^{r}\frac{q_r^0(r')}{\hbar}dr' - \frac{q_\infty r}{\hbar}\right) = -\frac{\pi}{2}\frac{q_\infty r_{\mathrm{c}}^0}{\hbar} = -\frac{|m|\pi}{2}. \quad (3.29)$$

Inserting this result into expression (3.28), we obtain

$$\delta_m = \lim_{r\to\infty}\left(\int_{r_{\mathrm{c}}}^{r}\frac{q_r^{\mathrm{inc}}(r')}{\hbar}dr' - \int_{r_c^0}^{r}\frac{q_r^0(r')}{\hbar}dr'\right) + \frac{\pi}{2}\left(m - |m|\right). \quad (3.30)$$

The last part of this expression equals zero for positive m, and $-|m|\pi$ for negative m. In section 3.2, we noted that the phase shift is only defined up to an integer multiple of π. We can thus freely add a multiple of π to expression (3.30) without changing the physical result. We therefore write

$$\delta_m = \lim_{r\to\infty}\left(\int_{r_{\mathrm{c}}}^{r}\frac{q_r^{\mathrm{inc}}(r')}{\hbar}dr' - \int_{r_c^0}^{r}\frac{q_r^0(r')}{\hbar}dr'\right), \quad (3.31)$$

which has the property that all phase shifts vanish in the absence of an inhomogeneity, cf. Refs. [64, 86]. It also satisfies the symmetry that we previously derived, $\delta_m = \delta_{-m}$.

Finally, we argue that expression (3.31) is more suitable for practical calculations than expression (3.26). Let us first consider an inhomogeneity with a finite range, such that $|\mathbf{q}|(r) = q_\infty$ for $r > R$. We may then split the limit and obtain

$$\delta_m = \int_{r_{\mathrm{c}}}^{R}\frac{q_r^{\mathrm{inc}}(r')}{\hbar}dr' - \int_{r_c^0}^{R}\frac{q_r^0(r')}{\hbar}dr' + \lim_{r\to\infty}\int_{R}^{r}\frac{q_r^{\mathrm{inc}}(r') - q_r^0(r')}{\hbar}dr'. \quad (3.32)$$

The last part of this expression vanishes, because the behavior of the two integrands is identical for $R < r < \infty$. More colloquially, the (large r) tails of the integrands cancel, and as a result we only have to integrate over a finite interval. Moreover, we directly see that δ_m vanishes when both $r_{\mathrm{c}} > R$ and $r_{\mathrm{c}}^0 > R$. Comparing this to Eq. (3.26), we see that the integral in the latter expression converges very slowly in terms of r, because of the slow decay of $q_r^{\mathrm{inc}}(r) = (|\mathbf{q}|^2(r) - m^2\hbar^2/r^2)^{1/2}$. We therefore have to integrate over a much larger interval in order to obtain an accurate result, which is computationally more demanding.

When the inhomogeneity does not have a finite range, the difference between the two expressions is less clear cut. Nevertheless, one can use a similar argument to show that expression (3.31) converges faster, in terms of r, than expression (3.26) whenever $|\mathbf{q}|(r)$ decays faster than $1/r$. We therefore consider expression (3.31) more suitable for our numerical computations in the next section.

3.4 Numerical results for the scattering cross section

In this section, we apply our scattering theory to a specific example. First, we introduce an explicit shape of the inhomogeneity, i.e. a change in local electron density $n^{(0)}$. Subsequently, we numerically evaluate the semiclassical phase shift (3.31) with Wolfram Mathematica [90] and discuss the total and differential scattering cross sections, given by Eqs. (3.10), and (3.8), respectively. We show their dependence on three parameters: the change in local electron density, plasmon energy, and decay length of the inhomogeneity. In order to gain a better understanding of the system, we also discuss the classical trajectories associated with the plasmon scattering, similar to those shown in Fig. 3.1.

We can compute the semiclassical phase shift with different approaches. In the first approach, we numerically determine the root of the effective classical Hamiltonian $\mathcal{H}_0$ for a given coordinate r, to obtain the radial momentum q_r. Following the discussion in Sec. 3.3.1, the phase shift can then be computed efficiently using Eq. (3.31). In the second approach, we numerically solve Hamilton's equations (2.27), supplemented with a differential equation for the action S based on Eq. (2.28). From this we determine the phase shift directly, by noting that both terms in Eq. (3.31) are the actions of the plasmon in the two systems discussed in Sec. 3.3.1. Unfortunately, the latter approach is computationally expensive, because it requires small time steps in the numerical integration. We therefore use the first approach in our computation.

In practical applications of Eq. (3.31), we have to choose a cut-off radius R for the integration. We pick R in such a way that the last term in Eq. (3.32) is negligible. This choice of R is, naturally, highly affected by the spatial decay of the inhomogeneity. In order to ensure rapid convergence, we therefore consider a Gaussian inhomogeneity. However, the theory is not limited to this form. Specifically, we describe the local Fermi momentum as

$$p_{\mathrm{F}}(r) = p_0 \left(1 + \delta p_{\mathrm{F}}\, e^{-r^2/\ell^2}\right), \tag{3.33}$$

where p_0 is the Fermi momentum far away from the inhomogeneity. In this model for the inhomogeneity, we have two independent parameters, namely the maximum change in local Fermi momentum δp_{F}, and the decay length ℓ. Following Eq. (1.49), the change in Fermi momentum can be directly

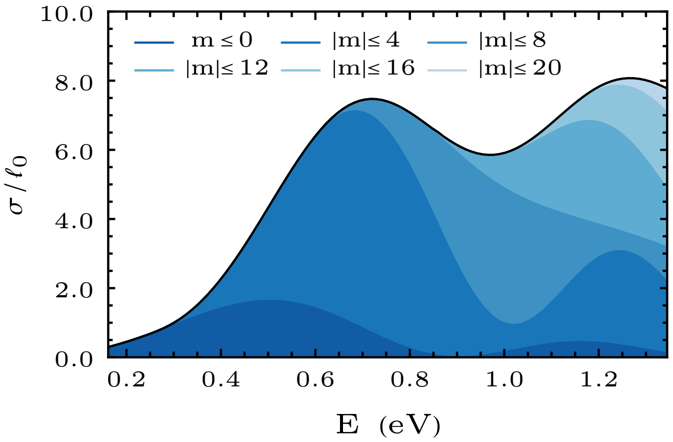


Figure 3.3: Numerically evaluated total scattering cross section σ/ℓ_0 as function of the plasmon energy E. In this plot, the local density change $\delta n = 1.25$, and the decay length is $\ell_0 = 13.2$ nm. The shades of blue indicate the partial waves which have been taken into account, whilst the solid black line indicates the cross section for $|m| \leq 28$. The scattering cross section increases monotonously up to around $E = 0.7$ eV, after which it starts to oscillate.

related to the change in electron density by $\delta n = g_s p_F^2 \left(2\delta p_F + \delta p_F^2\right)/4\pi\hbar^2 n^{(0)}$. Throughout this section, we consider changes in the local electron density instead of the Fermi momentum, since they can be more easily compared to experiments.

In our specific example, we consider a free-electron density of $n^{(0)} = 2.25 \times 10^{15}$ cm^{-2}, indicative of a metallic system. The active layer is encapsulated by two dielectric materials, with $\varepsilon_{\text{avg}} = 10\varepsilon_0$. The effective mass of the electrons is taken as $m_{\text{eff}} = m_e$. Unless stated otherwise, we consider a plasmon with an energy of $E = 0.54$ eV, a local increase in electron density $\delta n = 1.25$, and a decay length of $\ell = \ell_0 = 13.2$ nm. For these values, we have $h = 0.0064$ and $\kappa = 0.0100$, and the ratio h/κ is of order one. We therefore satisfy the requirements given in Sec. 2.1.5. For these parameters, the classical trajectories corresponding to a scattered plasmon are plotted in Fig. 3.1.

We first computed the total cross section as function of the plasmon energy, which is shown in Fig. 3.3. Intuitively, the total cross section increases with increasing plasmon energy or momentum. We recognize this for low energies, but for higher energies the total cross section σ/ℓ_0 starts to oscillate. We attribute this oscillation to interference between different overlapping waves, as we discuss shortly.

Looking at the dispersion relation in Fig. 3.4, we observe that the plasmon dispersion enters a shaded region at a certain energy. This shaded region corresponds to the region of Landau damping [2, 28], where the collective electron excitation transfers energy to incoherent electron-hole pairs. Mathematically,

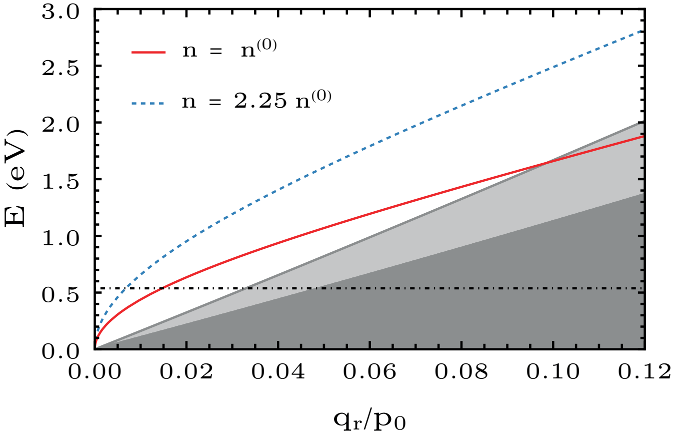
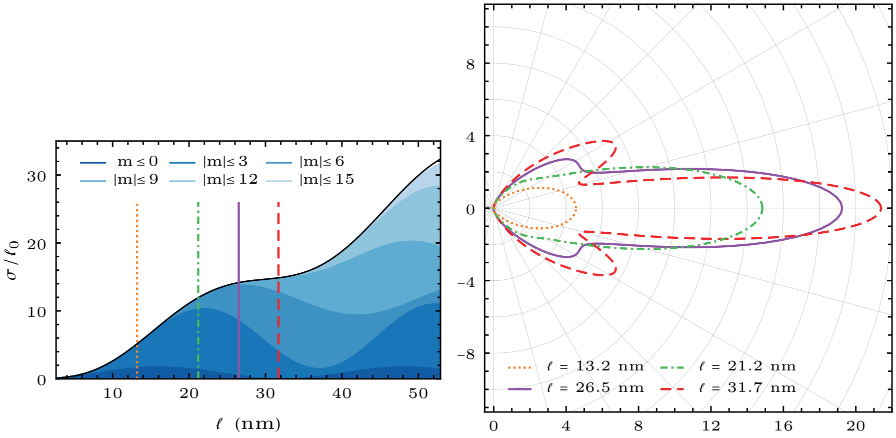


Figure 3.4: The dispersion relation for plasmons in a two-dimensional system for two different electron densities, namely $n^{(0)} = 2.25 \times 10^{15}$ cm^{-2} (solid red line) and $n^{(0)} = 5.06 \times 10^{15}$ cm^{-2} (dashed blue line). The black dash-dotted line indicates a constant energy. One can see that the plasmon momentum corresponding to a given energy decreases when the electron density increases. The light and dark gray area depict the Landau damped regions for the higher and lower electron density, respectively.

this corresponds to the point where $\mathcal{H}_0$, see Eq. (2.53), becomes complex, which leads to many complications in the application of the semiclassical approximation [50]. When plotting the cross section in Fig. 3.3, we therefore made sure that we stayed well outside the region of Landau damping, which starts at $E = 5.7$ eV for the given parameters.

Let us now examine the interference, by taking a closer look at the classical trajectories in Fig. 3.1. Since the trajectories emerge from a classical picture, they do not take the wave-like character into account and therefore do not show the interference. However, we can determine the regions in which interference takes place, by looking at points that are reached by more than one trajectory. The black lines in Fig. 3.1, which are known as caustics [75, 76], separate the regions where each point lies on a single trajectory from the regions where each point lies on multiple trajectories. It is precisely in the latter regions where interference takes place. The Jacobian (2.29) vanishes on the caustics, and therefore our expressions for the induced potential and the energy density diverge. This indicates that we cannot use our expressions in the vicinity of the caustic, and implies that the energy density is larger in this region. The interference is visible in the total cross section shown in Fig. 3.3, since our semiclassical expression takes the wave-like character of the plasmons into account.

A similar, but slightly different, oscillatory behavior is found in Fig. 3.5a, where the total scattering cross section is plotted for different values of the

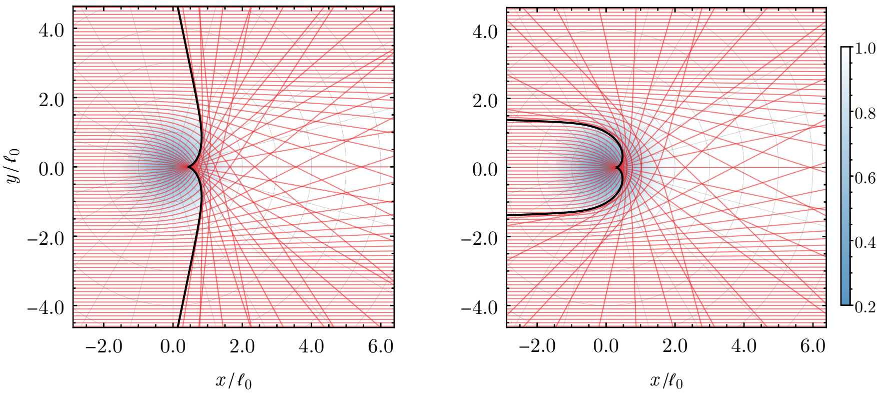


(a) Total scattering cross section (b) Differential scattering cross section

Figure 3.5: Numerical evaluation of the scattering cross section as a function of ℓ, with parameters: local density change $\delta n = 1.25$, reference decay length $\ell_0 = 13.2$ nm, and plasmon energy $E = 0.54$ eV. (a) Total scattering cross section σ/ℓ_0 as a function of the decay length ℓ. The shades of blue indicate the partial waves included in the calculation, with the solid black line representing the cross section for $|m| \leq 23$. The vertical colored lines mark the decay lengths used for the differential cross sections in (b). (b) Differential scattering cross section for various decay lengths ℓ. The $\ell = 13.2$ nm (dotted orange) and $\ell = 21.2$ nm (dash-dotted green) curves show predominantly forward plasmon scattering. The minima and maxima in the $\ell = 26.5$ nm (solid purple) and $\ell = 31.7$ nm (dashed red) curves indicate interference effects. Parameters are consistent with (a), and partial waves up to $|m| \leq 15$ were considered. Notice the minima around $\pm\pi/12$ and maxima around $\pm\pi/6$ for $\ell = 26.5$ nm and $\ell = 31.7$ nm.

decay length ℓ, not to be confused with the constant ℓ_0. The total cross section increases monotonously as a function of the decay length, however, with oscillating slope. We again attribute this behavior to interference between the overlapping plasmon trajectories. We can confirm again that this effect cannot be explained by the classical picture, by noting that the classical trajectories do not depend on the decay length ℓ once proper dimensionless parameters have been introduced, see Sec. 2.1.5.

To further investigate the interference, we plot the differential cross section $d\sigma/d\theta$, given by Eq. (3.8), which indicates the angular dependence of the plasmon scattering, for the lengths corresponding to the colored vertical lines in Fig. 3.5a. We divide it by ℓ_0 to make it a dimensionless quantity. Figure 3.5b shows that there is strong forward scattering for all decay lengths. For larger decay lengths, in other words, for smaller values of the semiclassical parameter h, two additional scattering peaks appear. The new local maxima



(a) $\delta n = -0.58$ (b) $\delta n = -0.70$

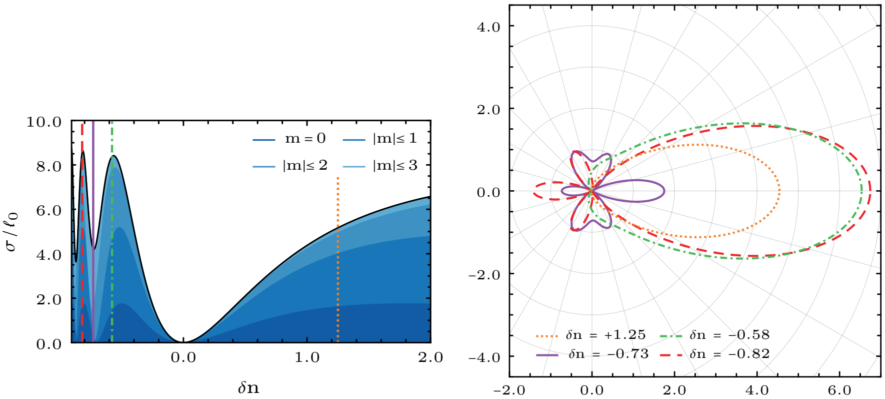
Figure 3.6: Classical trajectories of a plasmon, coming in from $x \to -\infty$. The maximum change in local density for the left plot is $\delta n = -0.58$, which corresponds to the first maximum in σ/ℓ_0 in Fig. 3.7a. The right plot has a maximum change in local density of $\delta n = -0.70$, which is just before the local minimum in the total scattering cross section. The solid black lines are the caustics.

around $\pm\pi/6$ originate from constructive interference, and similarly two local minima arise around $\pm\pi/12$ due to destructive interference. Comparing Figs. 3.5b and 3.1, we observe that these additional extrema are indeed located in the regions where the trajectories overlap. The interference effects thus become more clearly visible for smaller values of h, in other words, in the deep semiclassical limit. We see similar results when the differential cross section is plotted for increasing energies.

Lastly, we look at the effect of a change in the local electron density δn on the plasmon scattering. In Fig. 3.4, we see that, for a given energy E, an increase in the local electron density decreases the momentum q. In this sense, an increase in local electron density ($\delta n > 0$) can be considered a repelling potential, as shown in Fig. 3.1. In contrast, a decreasing local density ($\delta n < 0$) attracts the plasmon, as shown in Fig. 3.6.

Note that decreasing the local electron density also lowers the energy at which the Landau damped region is reached, as can be seen in Fig. 3.4. This means that for a certain energy there is a maximum to the decrease in local electron density. For $E = 0.54$ eV, the maximum becomes $\delta n_c = -0.99$. For the total scattering cross section shown in Fig. 3.7a, we stay well below this limit.

As expected, we see in Fig. 3.7a that the total scattering cross section σ/ℓ_0 increases for larger increases in the local density $|\delta n|$. The scattering cross section is asymmetric around $\delta n = 0$, which is expected since there is a maximum decrease in local electron density for the existence of plasmons, but no limit for positive values. Furthermore, the small $\mathbf{q}$ expansion given in



(a) Total scattering cross section

(b) Differential scattering cross section

Figure 3.7: Numerical evaluation of the scattering cross section as a function of δn, with parameters: plasmon energy $E = 0.54$ eV and decay length $\ell_0 = 13.2$ nm. (a) Total scattering cross section σ/ℓ_0 as a function of the maximum change in local electron density δn. The shades of blue indicate the partial waves included in the calculation, with the solid black line representing the cross section for $|m| \leq 6$. Note the oscillatory behavior for negative δn. The vertical colored lines mark the values of δn used for the differential cross sections in (b). (b) Differential scattering cross section for various values of δn. The $\delta n = +1.25$ (dotted orange) curve is the reference, while $\delta n = -0.58$ (dash-dotted green), $\delta n = -0.75$ (solid purple), and $\delta n = -0.82$ (dashed red) correspond to the first maximum, minimum, and next maximum in σ/ℓ_0, respectively. Note the extra minima, maxima, and the backscattering peak for $\delta n = -0.75$ and $\delta n = -0.82$. Partial waves up to $|m| \leq 8$ were considered.

Eq. (3.1) for $\mathcal{H}_0 = 0$, yields the proportionality $\mathbf{q} \propto 1/(n^{(0)} + \delta n)$, which is also not symmetric around $\delta n = 0$.

We observe in Fig. 3.7a that the total scattering cross section increases monotonically for $\delta n > 0$. Looking at the dotted orange line in Fig. 3.7b, we see that the plasmon is mainly scattered forward for $\delta n = 1.25$, without any additional maxima and minima. Note that this differential cross section is the same as the dotted orange differential cross section in Fig. 3.5b. On the contrary, for negative values of δn, we see that the total cross section shows maxima at approximately $\delta n = -0.58$ and $\delta n = -0.82$ and minima at $\delta n = -0.75$, which we attribute to interference.

In Fig. 3.7b, we take a closer look at these extrema by considering the differential cross section for the densities indicated by the colored vertical lines in Fig. 3.7a. We see additional scattering directions emerging for negative δn, including a sharp backscattering peak for $\delta n = -0.75$ and $\delta n = -0.82$. The backscattering comes from plasmons that are attracted and deflected by the lower local density, as can be seen from Fig. 3.6, where the classical trajectories are plotted for $\delta n = -0.58$ and $\delta n = -0.70$. Lowering the local density to $\delta n = -0.70$, which is just before the local minimum in the total scattering cross section, gradually deflects the plasmons 180^{o} degrees, and ergo explains the backscattering. Lowering δn further causes the caustics to intersect, and consequently points are reached by even more trajectories in that region. This in turn increases the interference effects, which is evident form the rough changes below $\delta n = -0.85$ in the total scattering cross section shown in Fig. 3.7a.

3.5 Conclusion

In this chapter, we have developed a semiclassical theory for plasmon scattering on radially symmetric inhomogeneities. We have applied this theory to a simple model for the dielectric environment, consisting of a two-dimensional electron gas at $z = 0$ encapsulated by two dielectric layers (cf. Sec. 2.2.1). The electron density in the $z = 0$ plane is homogeneous, except near the $\mathbf{x}$-origin, where a radially symmetric inhomogeneity is placed. We have considered the situation where a plasmon, represented as a plane wave, comes in from the left-hand side ($x \to -\infty$), and is scattered by the inhomogeneity.

In Sec. 3.2, we laid the groundwork for our scattering analysis by reviewing the relationship between differential cross sections and phase shifts within the framework of quantum scattering theory. To bridge the gap between theoretical constructs and experimental observables, we demonstrated how scattering cross sections, readily measurable in experiments, can be expressed in terms of phase shifts. Specifically, we decomposed the incoming plane wave into radial waves using a Fourier-Bessel series expansion. By focusing on the far-field regime, where the medium asymptotically approaches homogeneity, and by imposing the Sommerfeld radiation condition, we isolated the scattered waves. Leveraging the radial symmetry of the system, we derived expressions for the differential and total scattering cross sections, Eqs. (3.8) and (3.9), respectively. These expressions establish a direct link between the phase shift, which we expressed in terms of our semiclassical framework in the subsequent section.

Building upon the groundwork laid in the previous section, in Sec. 3.3, we derived a semiclassical expression for the phase shift, Eq. (3.31), directly from our induced potential. This involved carefully tracing and matching incoming trajectories with scattered trajectories within the framework of the Lagrangian manifold. This manifold provides the natural setting for incorporating the wave-like nature of plasmons. By making extensive use of the concept of a Lagrangian manifold, we effectively positioned this derivation as a practical application of the abstract formalism of the Maslov canonical operator. Following this matching, we constructed the full induced potential, accounting for the Jacobian, the other amplitude terms, and the phase, which is the classical action. This led to a complete semiclassical expression for the scattered plasmon, which we linked to the general solution from the quantum scattering approach, thus revealing an expression for the phase shift. Finally, in Sec. 3.3.1, we recast this expression into the numerically favorable form given by Eq. (3.31).

In Sec. 3.4, we discussed our numerical implementation of the theory. We calculated the total and differential scattering cross sections, given by Eqs. (3.9) and (3.8), respectively, for plasmon scattering by a Gaussian bump or well in the charge density distribution of a metallic system. We considered the dependence of the total cross section on the plasmon energy E, the spatial scale ℓ of the inhomogeneity, and the change in local density δn. We compared these results with classical trajectories, computed using the effective classical Hamiltonian (2.26), highlighting the role of interference between different trajectories. This interference was clearly observed in the differential cross section for different ℓ and δn, and was more pronounced in the deep semiclassical limit (smaller h). For a depletion in the local density, $\delta n < 0$, we observed backscattering in the differential cross section, consistent with the classical trajectories. We also observed that the total scattering cross section shows maxima and minima for negative values of δn, which we attributed to interference. We further analyzed these extrema by examining the differential cross section for specific density values, revealing additional scattering directions and a sharp backscattering peak. We explained the backscattering by considering the classical trajectories for negative δn, showing how plasmons are attracted and deflected by the lower local density. These results demonstrate that the developed semiclassical theory can effectively describe plasmon scattering phenomena, providing a robust framework for interpreting experimental observations and predicting scattering patterns in realistic systems.

4

(Quasi-)localized states for waveguiding

Plasmons, due to their strong light interaction and subwavelength confinement, are crucial for integrated photonic circuits. Plasmonic waveguides, essential for guiding and controlling these plasmons, can be fabricated through material patterning or dielectric structuring. In this chapter, we develop a semi-analytical theory for plasmonic waveguides, employing the effective classical Hamiltonian Eq. (2.26) to compute plasmonic trajectories, which, as will be shown, can under certain conditions be periodic. Specifically, we utilize the second model for the dielectric environment discussed in Ch. 2.2.2, which accounts for a finite height and incorporates nonlocal effects. This chapter therefore also serves as an example of the employment of this model. By incorporating the wave-like nature of plasmons, we identify two distinct waveguiding mechanisms: one based on total internal reflection with a quantization condition (cf. Sec. 1.3.1), and the other on varying dielectric environments that induce quasi-localized states via local amplitude changes, providing a foundation for understanding previous numerical studies.

This chapter details the full research and analysis from the paper published in Ref. [69].

4.1 Introduction to plasmonic waveguiding

In this chapter, we illustrate the theory developed in Ch. 2 by studying plasmonic waveguides. Specifically, we consider the more complex model discussed in Sec. 2.2.2, which describes a two-dimensional electron gas at $z = 0$ encapsulated by a dielectric layer with finite thickness d and dielectric constant ε_{M}. This layer is then sandwiched between background dielectrics above and below, similar to the simple model. We then proceed to demonstrate how our theory can be used to describe plasmonic waveguides by patterning quasi-one-dimensional systems, varying parameters such as the background dielectric constant, energy density, or the effective height d. We find two distinct mechanisms for plasmonic waveguiding: one based on plasmon localization through bound states, and another based on quasi-localized states induced by variations in the dielectric environment.

The effective classical Hamiltonian, which describes the dynamics of quantum plasmons in phase space, generates classical plasmon trajectories analogous to rays in geometrical optics. By incorporating the wave-like character of plasmons into these trajectories, we identify the first type of localized plasmonic state, based on total internal reflection, similar to photonic waveguides. Here, a classically forbidden region arises due to momentum along the waveguide direction [91]. This leads to periodic trajectories, allowing us to apply the quantization condition as outlined in Sec. 1.3.1. We demonstrate that by varying the dielectric substrate, electron density, or effective height of the middle layer, plasmonic bound states can be created. We then systematically study the conditions for the appearance of these bound states and compute their dispersion.

The second type of (quasi-)localization relies on a varying dielectric substrate that locally modifies the screening of the electrons. This, in turn, locally increases or decreases the amplitude of the plasmonic excitation, thereby altering the induced electron density. We discuss how the localization of the plasmon depends on parameters like the substrate dielectric constant, the energy of the excited plasmon, and the momentum along the direction of the waveguide. These findings provide a solid basis to understand previous numerical results from Ref. [40], and also demonstrate that it is possible to localize the plasmon in regions with higher screening.

This chapter is structured as follows. In Sec. 4.2, we conduct an in-depth study of bound states supported by the effective classical Hamiltonian. We elucidate their physical origin and investigate how the spectrum is influenced by spatial variations in different parameters. This is achieved by numerically solving the quantization condition, yielding a dispersion relation that connects the energy, the momentum along the waveguide direction, and a quantum number (cf. Sec. 1.3.1). In Sec. 4.3, we explore a second type of (quasi-)localized states. We demonstrate how their properties depend on the screening provided by the substrate, as well as on the energy and momentum of the excited plasmon, by numerically computing the real-space electron density using Eq. (1.60).

4.2 Plasmon localization through bound states

In this section, we analyze the effective classical Hamiltonian (2.26) at zero temperature and show that it allows for the formation of bound states. These states arise when a classically allowed region lies in between two regions where plasmon propagation is classically forbidden. In a plasmonic waveguide, these classically forbidden regions emerge through the presence of a momentum along the direction of propagation in the waveguide.

Throughout this section, we consider the model discussed in Sec. 2.2.2, where the substrate layer above (A) and below (B) have the same background (b) dielectric constant, that is, $\varepsilon_{\mathrm{A}}(\mathbf{x}) = \varepsilon_{\mathrm{B}}(\mathbf{x}) = \varepsilon_{\mathrm{b}}(\mathbf{x})$ in Eq. (2.24). We use the term dielectric constant to indicate that this quantity is static and does not depend on the out-of-plane coordinate z. However, it still depends on the in-plane coordinates x. At the same time, the length scale of variations in this direction is large compared to the electron wavelength, which justifies the term dielectric constant.

In Sec. 4.2.1, we analyze the classical Hamiltonian for this model. We show that plasmonic bound states can arise in a waveguide geometry by spatially varying the dielectric constant, but only when the momentum along the propagation direction of the waveguide is non-zero. In Sec. 4.2.2, we subsequently implement this setup in Wolfram Mathematica [90], and compute the bound-state spectrum. We not only consider variations in the dielectric constant, but also in the electron density $n^{(0)}$ and the effective height d, introduced in the previous section.

4.2.1 General analysis of the effective classical Hamiltonian: formation of bound states

We first demonstrate that the effective classical Hamiltonian supports plasmonic bound states. We study a system in which the parameters vary only in the x-direction, while the system is translationally invariant in the y-direction. This implies that $\mathcal{H}_0(x, \mathbf{q}, E)$ does not explicitly depend on y, which means $dq_y/d\tau = -\partial\mathcal{H}_0/\partial y$ due to Hamilton's equations. In other words, q_y is conserved and thus serves as a good quantum number. We refer to the y-direction as the propagation direction of the waveguide. Notably, throughout this chapter, we explicitly include the energy E dependence of the effective classical Hamiltonian, as this is important for our explanation of plasmonic bound state formation.

As previously mentioned, we can interpret the classical Hamiltonian as a spatially varying analog of the conventional Lindhard function, where the parameters take their local values at position x. A plasmon mode can exist at a given position x_i if there is a real momentum $\mathbf{q}$ satisfying $\mathcal{H}_0(x_i, \mathbf{q}, E) = 0$, for a given energy E. From the relation $\mathcal{H}_0(x,\mathbf{q},E)=0$, we can compute the local plasmon dispersion $E(x;\mathbf{q})$, treating x as a parameter. When there is no real momentum $\mathbf{q}$ satisfying this relation, we speak of a classically forbidden region.

At first sight, it may seem strange that a 2D plasmonic waveguide exhibits classically forbidden regions, since the 2D plasmon spectrum is gapless as $\mathbf{q}\to 0$. This is in contrast to the 3D plasmon spectrum, which has a cutoff at the plasma frequency ω_p. This energy gap, which depends on the dielectric constant and the electron density, results in classically forbidden regions which can give rise to bound states [50]. Although the 2D plasmon spectrum is gapless as a function of $|\mathbf{q}|$, the presence of a finite momentum q_y along the propagation direction creates an effective gap for propagation in the x-direction, as we discuss shortly. This mechanism is analogous to total internal reflection in photonic waveguides, where a critical angle determines whether a photon is completely reflected [91]. From this analogy, it is clear that a non-zero momentum in the direction of propagation is required.

Figure 4.1 shows the plasmon dispersion for two different substrate dielectric constants as a function of the total momentum $|\mathbf{q}|/\hbar$. In orange, denoted by x_1, the plasmon dispersion is given for a system where the dielectric constant equal to $\varepsilon_\mathrm{b}=1$. In blue, denoted by x_2, the plasmon dispersion is given for $\varepsilon_\mathrm{b}=9$. In general, the energies of the plasmon mode are pushed toward the electron-hole continuum (Landau damped region) for higher values of ε_b, because of an increased screening by the substrate. For the following discussion, it is insightful to split the momentum $|\mathbf{q}|$ into two components, namely $|\mathbf{q}|^2=q_x^2+q_y^2$, where, as discussed before, q_y is the momentum along the direction of the waveguide, and a constant of motion.

In the case of photonic waveguides, it is convenient to describe different regions as "faster" or "slower" based on the refractive index, as the linear dispersion allows for a well-defined group velocity and consequently a unique critical angle. However, for plasmons, the highly nonlinear dispersion complicates the definition of "fast" and "slow" regions, making it difficult to define a unique critical angle. Instead, we adopt an energy-gap perspective. When treating q_y as a parameter, we define the gap energy as $E_\mathrm{g}(x,q_y)=E(x,q_y;q_x=0)$, which represents the lowest plasmon energy at a given x for fixed q_y. The inset of Fig. 4.1 illustrates this concept by plotting $E(x,q_y;q_x)$ as a function of q_x for finite $q_y/\hbar=0.08$ Å^{-1}, at the two specific points discussed above. The gap energy can be extracted from this figure by looking at the limit $q_x\to 0$.

The existence of this finite energy gap allows us to define classically allowed and forbidden regions, analogous to the 3D case, as follows. Suppose we excite a plasmon with energy $E_\mathrm{pl}=0.8$ eV (i.e. the horizontal dashed line in Fig. 4.1) and a (constant) finite momentum $q_y/\hbar=0.08$ Å^{-1} (i.e. the vertical dashed line), and we look at a point, x_1, where the dielectric constant is equal to $\varepsilon_\mathrm{b}=1$. We observe that plasmons are not allowed to propagate at this point x_1, since they satisfy $E_\mathrm{pl}=E(x_1;|\mathbf{q}|)$ only when $|\mathbf{q}|<|q_y|$, meaning that q_x

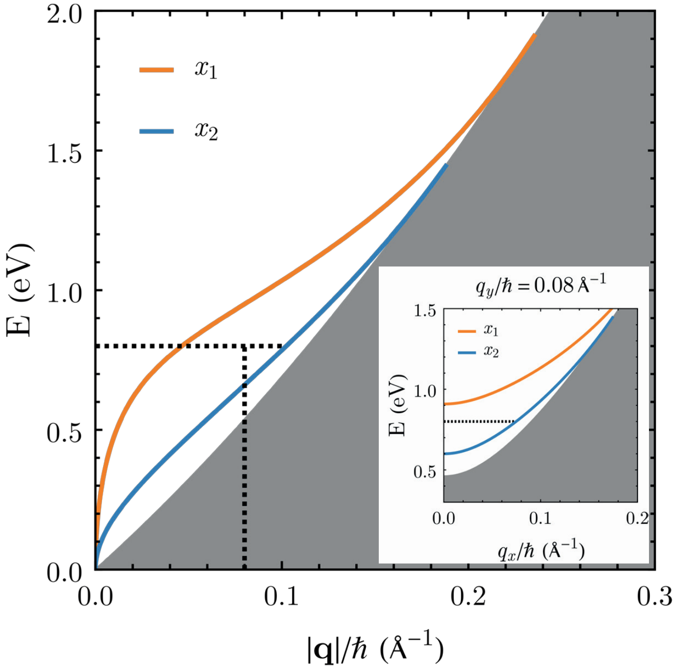


Figure 4.1: The local dispersion relation for plasmons in two-dimensional systems at two points with different values of the substrate dielectric constant, namely $\varepsilon_b = 1$ at point x_1, and $\varepsilon_b = 9$ at point x_2. The horizontal dashed black line indicates a constant energy $E_{pl} = 0.8$ eV, and the vertical dashed black line indicate a momentum $q_y/\hbar = 0.08$ Å^{-1}. For this momentum, the dispersion $E(q_x)$ is plotted in the inset, where we see a gap opening up for $q_x = 0$. In the inset, the energy $E_{pl} = 0.8$ eV lies below the dispersion for $\varepsilon_b = 1$, and therefore in the classically forbidden region. For the region where $\varepsilon_b = 9$, a state exists for this energy. The gray area in both plots depicts the Landau damped region or the particle hole continuum.

has to be imaginary. We therefore have exponentially damped waves, meaning that x_1 lies in a classically forbidden region. At the same time, plasmons can propagate at the point x_2, where $\varepsilon_b = 9$, since they satisfy $E_{pl} = E(x_2; |\mathbf{q}|)$ for $|\mathbf{q}| > |q_y|$, meaning that q_x is real. The latter leads to the traveling waves, meaning that x_2 lies in a classically allowed region.

So far, we have seen that it is possible to create classically forbidden and allowed regions for specific energies E_{pl} and momenta q_y. Let us now consider the quasi-one-dimensional setup shown in Fig. 4.2(a). Ignoring the exact spatial details for the present discussion, we can clearly distinguish three different regions: on the left and the right we have $\varepsilon_b = 1$ (e.g. at the point x_1), while $\varepsilon_b = 9$ in the middle (e.g. at x_2). The background dielectric constant enters the classical Hamiltonian through the effective dielectric function ε_{eff},

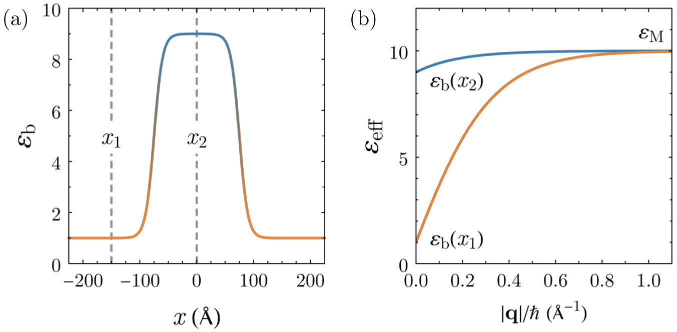


Figure 4.2: (a) Spatial variation of the substrate dielectric constant $\varepsilon_\mathrm{b}(x)$. On the left-hand and right-hand side the dielectric constant tends to $\varepsilon_\mathrm{b} = 1$, and in the middle it goes to $\varepsilon_\mathrm{b} = 9$. From Fig. 4.1, we see that the dispersion relation in the middle region, with higher dielectric constant, is pushed toward the particle hole continuum. (b) The effective dielectric function plotted as function of $|\mathbf{q}|/\hbar$. For small $|\mathbf{q}|/\hbar$, i.e. the long wavelength limit, the effective dielectric function goes to $\varepsilon_\mathrm{b}(x_i)$. In the opposite limit, for large $|\mathbf{q}|/\hbar$, it goes to ε_M.

which is depicted in Fig. 4.2(b) as function of $|\mathbf{q}|/\hbar$ for the two values of the substrate dielectric constant ε_b. For high values of $\mathbf{q}/\hbar$, the effective dielectric function tends to ε_M, for both values of ε_b, and the screening becomes equivalent in all regions. Comparing Figs. 4.1 and 4.2(b), we observe that, for these specific parameters, only the lower momenta in Fig. 4.2(b), are relevant for the plasmon dispersion, which means that the screening varies substantially between different regions.

Applying the logic of the previous paragraphs to the spatial variation $\varepsilon_\mathrm{b}(x)$ shown in Fig. 4.2(a), we conclude that, for certain energies E_pl and momenta q_y, plasmons are only allowed to propagate in the middle region. The presence of a classically allowed region between two classically forbidden regions leads to periodic trajectories in classical phase space. These periodic trajectories are shown in Fig. 4.3(a), for three different values of the momentum q_y. Note that the size of the classically allowed region increases when q_y decreases, meaning that the turning points move to larger values of x. Below a certain value of q_y, both regions are classically allowed, leading to the open trajectory that is also shown in Fig. 4.3(a).

In appendix D, we briefly show that the turning points on the periodic trajectories, that is, the points where $q_x \to 0$, are so-called simple turning points. This means that $q_x^2 \propto x$ in the vicinity of the turning point, and holds regardless of the value of the parameter q_y.

Figure 4.3(b) can help us to gain a better understanding of the energies of which periodic trajectories occur. It shows the gap energy $E_\mathrm{g}(x, q_y)$ as function of x for a specific value of $q_y/\hbar$. For this value of q_y, periodic trajectories can exist in the valley between $E_\mathrm{g}(x_2, q_y) < E_\mathrm{pl} < E_\mathrm{g}(x_1, q_y)$. When we increase q_y, we, at some point, reach a value at which the gap energy reaches the solid red line $E_{\mathrm{L}}(x)$. This corresponds to the point where the plasmon mode reaches the Landau damped region (gray area in Fig. 4.1) Therefore, the maximum energy for which periodic trajectories can exist, is determined by the lowest Landau energy, which is for our system given by $E_{\mathrm{L,min}} = E_{\mathrm{L}}(x_2)$ (red dotted line in Fig. 4.3(b)), cf. Ref. [50].

Note that, according to Ref. [50], so-called Landau turning points can exist for energies between $E_{\mathrm{L,min}} < E_{\mathrm{pl}} < E_{\mathrm{L,max}}$, where $E_{\mathrm{L,max}} = E_{\mathrm{L}}(x_1)$. However, these turning points and the subsequent periodic trajectories are not discussed in this thesis, because this region is relatively small and close to the particle-hole continuum (Landau damped region), as can be seen in Fig. 4.1.

So far, we have established the existence of classically allowed and forbidden regions, and we have discussed the conditions for periodic trajectories in phase space to arise. However, not al periodic trajectories correspond to bound states. Specifically, only periodic trajectories for which the classical action fulfills the Einstein-Brillouin-Keller quantization condition [48, 55, 86] lead to bound states. This condition can be stated as

$$\frac{S_{\mathrm{tot}}}{2\hbar} = \frac{1}{\hbar}\left| \int_{x_{c_1}}^{x_{c_2}} q_x(x)dx \right| = \left(m + \frac{1}{2}\right)\pi, \tag{4.1}$$

where x_{c_1} and x_{c_2} are the classical turning points, and m is a (non-negative) integer. We can intuitively understand this condition from the requirement that the induced potential should be single-valued when we move along the periodic trajectory in phase space. After one full revolution in phase space, see Fig. 4.3(a), the action should have increased by a multiple of 2π, which makes the potential (1.53) single-valued because it is invariant under phase differences of 2π.

The factor $\pi/2$ in Eq. (4.1) accounts for the phase shift of the solution (1.53) at a simple turning point, which can be formalized through the so-called Maslov index [48, 50, 55]. This phase shift arises because the asymptotic solution (1.53) breaks down at a turning point, since the Jacobian vanishes. From a practical perspective, the Maslov index ensures the correct phase evolution as the plasmon passes through a turning point, compensating for the sign change in the Jacobian in the amplitude, see Eq. (2.46).

The quantization condition (4.1) determines the spectrum of the plasmonic waveguide. It defines a one-to-one relation between the energy and the transverse momentum q_y, for a given m. In the next subsection, we numerically implement the waveguide discussed here and compute the spectra for waveguides with variations in different parameters.

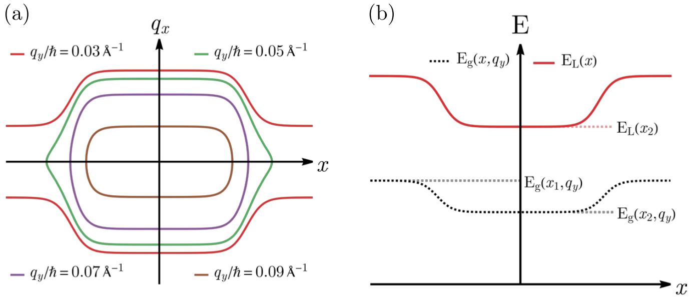

(a)
q_x
$q_y/\hbar = 0.03\,\text{Å}^{-1}$
$q_y/\hbar = 0.05\,\text{Å}^{-1}$
x
$q_y/\hbar = 0.07\,\text{Å}^{-1}$
$q_y/\hbar = 0.09\,\text{Å}^{-1}$
(b)
E
$\mathrm{E_g}(x, q_y)$
$\mathrm{E_L}(x)$
$\mathrm{E_L}(x_2)$
$\mathrm{E_g}(x_1, q_y)$
$\mathrm{E_g}(x_2, q_y)$
x

Figure 4.3: (a) Schematic representation of plasmonic trajectories in phase space (x, q_x) for four values of the momentum q_y. The energy for all bound states is constant and set to $E_{\text{pl}} = 0.8$ eV. For certain values of q_y, the phase space trajectories are periodic and are confined to the middle region. For the lowest value of q_y, we have an open trajectory, which also pervades the regions on the left and the right. (b) The spatial dependence of the gap energy $E_{\text{g}}(x, q_y)$ (dotted black line) for constant $q_y = 0.08$ Å^{-1}, and the energy $E_{\text{L}}(x)$ for which the Landau damped region is reached (solid red line). For this specific q_y, bound plasmonic states can exist in the middle valley of $E_{\text{g}}(x, q_y)$. For higher values of the perpendicular momentum q_y, the energy $E_{\text{g}}(x, q_y)$ increases. Above the solid red line, plasmons do not exists in either spatial region, because the dispersion has crossed the particle-hole continuum.

4.2.2 Numerical implementation semiclassical bound states

In this subsection, we numerically demonstrate the formation of plasmonic bound states in quasi-one-dimensional systems with spatially varying parameters. We investigate three distinct scenarios: variations in the substrate dielectric constant $\varepsilon_{\text{b}}(x)$, variations in both the dielectric constant and the electron density $n^{(0)}(x)$, and variations in the dielectric constant, electron density, and the effective height of the thin film $d(x)$. These scenarios explore how different physical mechanisms can be utilized to engineer plasmonic waveguides, i.e. invasively or non-invasively. We show that each additional degree of freedom allows for more precise control over the bound state spectrum.

For all three scenarios, we consider a quasi-one-dimensional geometry with three distinct regions: two outer regions with identical characteristics (denoted by subscript 1) and a central region with different properties (denoted by subscript 2). The spatial variation of the parameters along the x-direction is modeled using hyperbolic tangent functions ($\propto \pm \tanh\left[x/\ell \pm \ell_{\text{w}}/(2\ell)\right]$), ensuring a smooth transition between the regions. A specific example with varying the dielectric constant is given in Fig. 4.2(a), where significant variations happen over length scales $2\ell = 3$ nm, and $\ell_{\text{w}} = 15$ nm is the width of the middle part. The system is translationally invariant in the y-direction, which defines the propagation direction of the waveguide.

Throughout this section, we adopt parameters resembling a metallic system with a parabolic electronic dispersion and an effective electron mass

$m_{\mathrm{eff}} = 0.423 m_{\mathrm{e}}$. The 2D material has a background electron density $n^{(0)} = 1.8 \times 10^{14}\ \mathrm{cm}^{-2}$, surrounded by a thin film with an effective height $d_0 = 0.576$ nm and a dielectric constant $\varepsilon_{\mathrm{M}} = 10$, consistent with values reported in Ref. [40]. The width of the central region is set to $\ell_{\mathrm{w}} = 15$ nm, and the characteristic length of the boundary between regions is $2\ell = 3$ nm. These parameters yield a small dimensionless parameter $h = \hbar/(2\ell p_{\mathrm{F}}) = 0.1$, satisfying the criteria for the approximation as discussed in Refs. [1, 50].

Varying dielectric constant

We first consider a system with spatial variations in the substrate dielectric constant $\varepsilon_{\mathrm{b}}(x)$. The dielectric constant is varied between $\varepsilon_{\mathrm{b}}(x_1) = 1$ in the outer regions and $\varepsilon_{\mathrm{b}}(x_2) = 9$ in the central region, as described by

$$\begin{aligned} \varepsilon_{\mathrm{b}}(x) = \varepsilon_{\mathrm{b}}(x_1) &- \frac{\varepsilon_{\mathrm{b}}(x_1) - \varepsilon_{\mathrm{b}}(x_2)}{2} \tanh\left[\frac{x}{\ell} + \frac{\ell_{\mathrm{w}}}{2\ell}\right] \\ &+ \frac{\varepsilon_{\mathrm{b}}(x_1) - \varepsilon_{\mathrm{b}}(x_2)}{2} \tanh\left[\frac{x}{\ell} - \frac{\ell_{\mathrm{w}}}{2\ell}\right], \end{aligned} \tag{4.2}$$

and shown in Fig. 4.2(a). This variation can be achieved non-invasively by patterning the substrate. While the hyperbolic tangent function used to model $\varepsilon_{\mathrm{b}}(x)$ formally only reaches its maximum value at infinity, the spatial separation of the points is sufficient for the dielectric constant to effectively reach its constant asymptotic value, ensuring that $\varepsilon_{\mathrm{b}}(x)$ is locally constant.

The quantization condition Eq. (4.1) defines a unique relation between the bound state energy E_{bound} and the momentum q_y for a given quantum number m. In Fig. 4.4, we show this bound state spectrum for the dielectric substrate (4.2). The green lines represent the allowed plasmon energies, with the lowest line corresponding to $m = 0$. The spectrum is bounded by the gap energies $E_{\mathrm{g}}(x_1, q_y)$ and $E_{\mathrm{g}}(x_2, q_y)$, indicated by the dashed orange and blue lines, respectively. Plasmons are classically forbidden in all regions below the dashed blue curve, whilst above the dashed orange curve plasmons are classically allowed in both regions. As a consequence of the latter, plasmons above $E_{\mathrm{g}}(x_1, q_y)$ do not have simple turning points and are therefore in a continuum of allowed states. Bound states close to this continuum of states or to the electron-hole continuum (in gray) will probably not be measurable as localized states at finite temperatures, due to broadening of the modes into the respective continuum [28, 29].

This demonstrates the formation of plasmonic bound states by solely (non-invasively) manipulating the dielectric environment, which has a large effect on the plasmon dispersion. Plasmons are localized in regions of higher substrate dielectric constant. This may seem counterintuitive, as the potential (Eq. (2.46)) naively suggests a decrease in amplitude due to increased screening. However, this screening effect does not create the classically allowed and forbidden regions necessary for bound states. We further explore the effect of screening on the amplitude in Sec. 4.3.

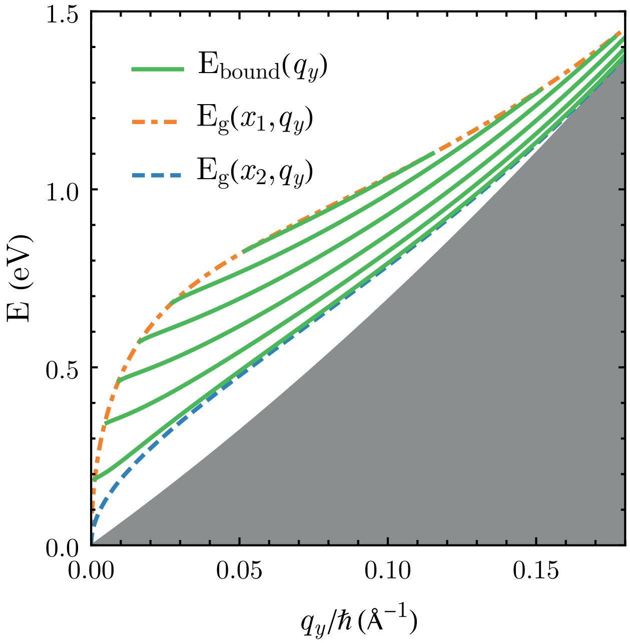


Figure 4.4: Spectrum of plasmonic bound states in a two-dimensional waveguide, with a spatially varying dielectric constant of the substrate, as function of the perpendicular momentum q_y. The dot-dashed orange and dashed blue lines indicate $E_g(x_1, q_y)$ and $E_g(x_2, q_y)$, respectively: they represent the upper and lower energy boundaries of the bound state spectrum. Above the energy $E_g(x_1, q_y)$, the plasmons are classically allowed in both spatial regions, but, therefore, no turning points can be formed, and hence no bound states exist.

Varying dielectric constant and electron density

We now investigate the combined effect of spatially varying both the dielectric constant $\varepsilon_b(x)$ and the electron density $n^{(0)}(x)$. Besides the dielectric variation described in the previous subsection, we introduce a 15% increase in $n^{(0)}$ in the central region; while this specific value is illustrative, the qualitative effects of increased carrier density are the focus of this investigation. This increase can, for example, be achieved non-invasively through local gating or invasively through doping.

The electron density is parameterized via the Thomas-Fermi approximation, $p_F(\mathbf{x}) = \hbar\sqrt{4\pi n^{(0)}(\mathbf{x})/g_s}$, and its spatial variation follows the same hyperbolic tangent profile as the dielectric constant (Eq. (4.2)), where we parameterize the electron density in the middle part with a relative increase δn compared to the background electron density, $n_0^{(0)}$ in the outer regions.

While increasing the dielectric constant tends to lower the plasmon dispersion, increasing the electron density has the opposite effect [1]. These two

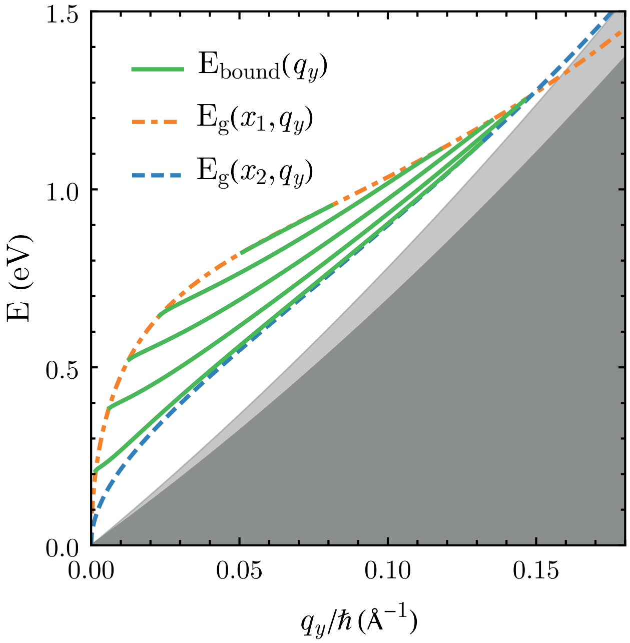


Figure 4.5: Spectrum of plasmonic bound states in a two-dimensional waveguide, with a spatially varying dielectric constant and electron density, as function of the perpendicular momentum q_y. The dot-dashed orange and dashed blue lines indicate $E_g(x_1, q_y)$ and $E_g(x_2, q_y)$, respectively: they represent the upper and lower energy boundaries of the bound state spectrum. Around $q_y/\hbar = 0.15$ Å^{-1}, the two gap energies cross, due to a combination of the increased electron density and dielectric constant in the waveguide channel. Before this crossing point, we have the bound state spectrum given by the green lines, and above the energy $E_g(x_1, q_y)$, we again have a continuum of plasmon states.

competing effects operate over different momentum ranges. Consequently, we expect a crossover regime where the influence of increased electron density outweighs the increased screening from the dielectric constant.

The resulting bound state spectrum is shown in Fig. 4.5. The interplay between $\varepsilon_b(x)$ and $n^{(0)}(x)$ leads to a crossing of the gap energies $E_g(x_1, q_y)$ and $E_g(x_2, q_y)$ at higher q_y. This crossing point defines the boundary of the allowed energy and momentum ranges for the bound states and can be tuned by adjusting the magnitude of the variations in both parameters. This demonstrates the enhanced control over the bound state spectrum achieved by incorporating electron density variations.

All bound states, except for the lowest state ($m = 0$), begin and end at the continuum edge defined by $E_g(x_1, q_y)$ (dashed orange line). The $m = 0$ state ends at $E_g(x_2, q_y)$ (dashed blue line), seemingly in a classically forbidden region. However, due to the opposing influences of the varying dielectric constant and electron density, $E_{\mathrm{g}}(x, q_y)$ locally dips below $E_{\mathrm{g}}(x_2, q_y)$ in the boundary region between the central and outer regions. This local decrease arises from the distinct momentum dependence of the two effects. While this local minimum in $E_{\mathrm{g}}(x, q_y)$ could theoretically support a bound state localized at the boundary, we verified, through numerical estimates, that this is not the case in our setups. We therefore do not further explore this possibility in this thesis.

Increasing $n^{(0)}$ not only increases the dispersion energy $E(x; |\mathbf{q}|)$ but also raises the energy of the particle-hole continuum (Landau damped region), as shown by the lighter and darker gray areas in Fig. 4.5, corresponding to the Landau damped regions at x_2 and x_1, respectively. One could theoretically consider reversing the parameter variations, placing the higher dielectric constant and electron density regions on the outside. This might lead to a lower bound on the allowed energies and momenta. However, the Landau damped region must be carefully considered in such a scenario, as for our parameters, the energy $E_{\mathrm{g}}(x_1, q_y)$ crosses into the particle-hole continuum of the new outer regions, resulting in damping for those q_y values.

In summary, the interplay of the varying dielectric constant and electron density creates a crossing point in the gap energies, providing control over the allowed energy and momentum ranges for the bound states. This highlights the increased flexibility in engineering the bound state spectrum by incorporating electron density variations. Both variations in the dielectric constant and the electron density can be done non-invasively. On the contrary, when the electron layer itself is varied, the characteristic properties and therefore the parameters change, e.g., the dielectric constant ε_{M} or the parameter d for the effective height of the thin film [35, 40].

Varying dielectric constant, electron density, and the effective height d

Here, we briefly discuss the effect of varying the effective height $d(x)$, which influences the plasmon dispersion through $\exp(-|\mathbf{q}|d/\hbar)$ in $\varepsilon_{\mathrm{eff}}$, Eq. (2.24). However, while variations in the dielectric constant and electron density offer significant control over the bound state spectrum, the impact of varying the effective height is generally smaller. As documented in Refs. [35, 40], typical variations in d are on the order of 10%, resulting in negligible changes to the bound state energies. Therefore, to illustrate the qualitative effects of varying d, we consider significantly larger, and potentially unrealistic, variations.

While not physically realistic at present, these large variations serve to illustrate the sensitivity of the bound state spectrum to changes in the effective height. Such variations could become relevant in systems with significant material or structural changes in the thin film, or potentially through substrate modifications (e.g., doping) that influence the out-of-plane penetration of the electron wavefunction.

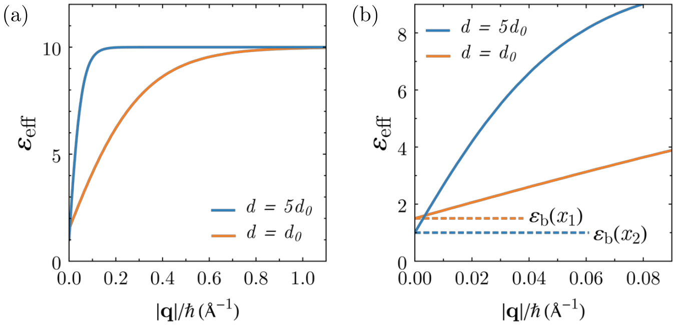


Figure 4.6: (a) The effective dielectric function ε_{eff} plotted as function of the momentum $|\mathbf{q}|/\hbar$. In the middle of the waveguide the effective thickness is equal to $d = 5d_0$ (depicted in blue), whereas on the outer regions, the effective thickness is given by $d = d_0$ (depicted in orange) . For small $|\mathbf{q}|/\hbar$, or in the long wavelength limit, the effective dielectric function goes to $\varepsilon_{\text{b}}(x_1) = 1.5$ and $\varepsilon_{\text{b}}(x_2) = 1$. In the opposite limit, for large $|\mathbf{q}|/\hbar$, it goes to ε_{M}. The two lines cross for small $|\mathbf{q}|/\hbar$, which can be seen clearly in (b).

Varying the effective height alters the effective dielectric function ε_{eff} Eq. (2.24), influencing the screening of the plasmon. Increasing d causes ε_{eff} to approach ε_{M} more rapidly as a function of momentum $|\mathbf{q}|$, as shown in Fig. 4.6(a). This effect is, therefore, most pronounced at lower momenta. Furthermore, the impact of varying d is greater when the difference between ε_{b} and ε_{M} is larger. Consequently, varying d in regions where ε_{b} is already close to ε_{M} (e.g., $\varepsilon_{\text{b}} = 9$ as in previous subsections) has a limited effect on the dispersion relation.

As mentioned above, realistic variations in d (e.g., 10% as discussed in Ref. [35]) result in minimal changes to the bound state spectrum. For instance, a 10% increase in d leads to less than a 2.5% change in energy eigenvalues where $\varepsilon_{\text{b}} = 1$ and only 0.1% where $\varepsilon_{\text{b}} = 9$.

However, to illustrate the potential impact of larger variations in d, we consider an unrealistic scenario where d is increased by a factor of 5 in the central region. The effective height is parameterized in the same way as the Fermi momentum, namely with three regions where the middle layer has a relative change in height δd and the boundary is described by a hyperbolic tangent. We set $\varepsilon_{\text{b}} = 1$ in the central region and $\varepsilon_{\text{b}} = 1.5$ in the outer regions. This large increase in d significantly alters ε_{eff} at low momenta, depicted in Fig. 4.6(b). For these low momenta, the outer regions experience stronger screening, while the central region is more screened at higher momenta. This leads to a crossing point in the gap energies $E_{\text{g}}(x_i, q_y)$ of the two regions at low momenta, as shown in Fig. 4.7.

Beside these variations, we increase the electron density by 15% in the central region, as in the previous subsection. This further modifies the dispersion relation, leading to another crossing point in the gap energies at higher momenta, as can be seen in Fig. 4.7. These two crossing points define the upper and lower limits for the existence of bound states, which are again depicted by green in the figure.

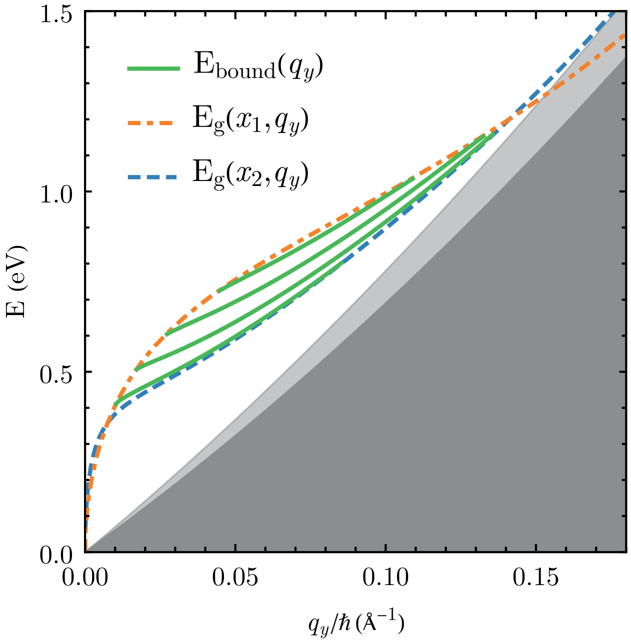


Figure 4.7: Spectrum of plasmonic bound states in a two-dimensional waveguide,with a spatially varying dielectric constant, electron density and effective height, as function of the perpendicular momentum q_y. The dot-dashed orange and dashed blue lines indicate $E_g(x_1, q_y)$ and $E_g(x_2, q_y)$, respectively: they represent the upper and lower energy boundaries of the bound state spectrum. The two gap energies have two crossing points, due to the interplay of the three varying parameters. The first crossing is at low momentum around $q_y/\hbar = 0.01$ Å^{-1}, whereas the second crossing lays at higher momentum around $q_y/\hbar = 0.14$ Å^{-1}. In between the crossing points, we have the bound states given by the green lines, and above the energy $E_g(x_1, q_y)$, we again have a continuum of plasmon states.

In summary, while realistic variations in the effective height d have a negligible impact on the bound state spectrum, we have shown that unrealistically large variations can lead to significant changes, creating additional crossing points in the gap energies and thus further modifying the allowed energy and momentum ranges for bound states. This illustrates the potential, albeit in an unrealistic regime, for controlling the bound state spectrum through variations in the thin film's effective height.

4.3 (Quasi-)localization of plasmons through local screening

In the previous section, we demonstrated that the effective classical Hamiltonian supports classically allowed and forbidden regions, leading to the formation of semiclassical bound states. In these states, plasmons are strictly confined within the waveguide channel and cannot propagate outside. However, effective plasmon localization can also arise through a different mechanism: variations in the local screening environment, which alters the amplitude. As explored in Ref. [40], plasmons can be (quasi-)localized even when they are classically allowed in all regions, due to variations in dielectric screening.

In this section, we investigate screening-induced plasmon (quasi-)localization within our theoretical framework. We demonstrate that while the plasmon remains classically allowed to propagate throughout the structure, variations in the dielectric environment can significantly modulate the plasmon amplitude, leading to a form of quasi-localization. Specifically, we show how the interplay of competing screening effects can be exploited to engineer plasmonic waveguides.

First, we discuss the specific setup considered throughout the section, and we present a detailed analysis of the plasmon amplitude, highlighting the different contributions to screening and their influence on the plasmon excitation. Next, we compare our results with the numerical calculations presented in Ref. [40], demonstrating the connection between our theoretical framework and previous work. Finally, we explore potential waveguide applications, showcasing how variations in the dielectric environment can be used to control plasmon propagation through competing screening effects on the plasmon amplitude.

4.3.1 The effect of screening on the amplitude

In order to use the analytical theory developed in this thesis, we examine the amplitude of the induced potential Eq. (2.46), and how it is influenced by inhomogeneities. The amplitude Eq. (2.43) has three contributing factors (not considering the Berry phase) which can depend on the position, namely: the total momentum, the effective dielectric function, and the Jacobian. We analyze the influence of these factors separately in this subsection. However, to make an accurate comparison of our results with the numerical findings in Ref. [40], we first define a specific setup.

Throughout this section, we consider variations in the substrate dielectric constant only. Specifically, we analyze a system with hard-wall boundary conditions at x_0 and x_w, with a total width of 240 Å. The hard-wall boundary conditions force the induced potential vanish at the boundary, leading to an Einstein-Brillouin-Keller quantization condition for the action.

Before we consider the quantization condition, it is more insightful to construct the induced semiclassical potential, as was done in Ref. [1, 50]. Considering plasmons with energy E_{pl} and transversal momentum q_y, and setting the reference point of the action to the left wall at x_0, we have contributions from we have contributions from both right- and left-moving plasmons. The left-moving component arises from scattering at the right wall. The full induced potential can then be written as

$$V_{\mathrm{pl}} = \varphi_0(x) e^{\frac{i}{\hbar}(y-y_0)q_y} \left(e^{-\frac{i}{\hbar}\int_{x_0}^{x} q_x(x')dx'} + e^{\frac{i}{\hbar}\int_{x_0}^{x} q_x(x')dx' - i\pi - i\Phi_{\mathrm{tot}}} \right), \tag{4.3}$$

where the action is decomposed in Cartesian components and integrated along the plasmon trajectories. We set the arbitrary reference point $y_0 = 0$. The phase $-i\pi$ accounts for the reflection at the hard wall, and $\Phi_{\mathrm{tot}} = S_{\mathrm{tot}}/\hbar$ represents the accumulated phase after one full revolution in phase space. Note that the right moving exponent is defined with a negative sign, because of the opposite direction of the momentum and velocity, as discussed in Ref. [1].

If we considered a system with hard-wall boundary conditions in both x and y, as in Ref. [40], then we would, from a semiclassical point of view, be considering a integrable square billiard [92]. This leads to either ergodic (with incommensurate wavevector components) or periodic trajectories (commensurate components). The latter can lead to bound states, for which both q_x and q_y are quantized. Instead, we only impose hard-wall boundary conditions in the x-direction. This means that we consider a physically accurate model for a waveguide, in which the momentum along the propagation direction is not quantized.

The quantization condition [48, 55, 86] imposed by the hard-wall is given by

$$\frac{S_{\mathrm{tot}}}{2\hbar} = \frac{1}{\hbar}\left| \int_{x_0}^{x_{\mathrm{w}}} q_x(x)dx \right| = (m+1)\,\pi, \tag{4.4}$$

where the Maslov index contributes an additional phase π. In contrast to simple turning points, where the classical motion is smoothly reversed, since q_x smoothly goes to zero, a hard-wall boundary imposes an abrupt momentum reversal, leading to a discontinuous phase shift and a different Maslov index. The number m can again be any non-negative integer, and represents the number of nodes in the plasmonic bound state along the x-direction. Alternatively, we could have derived this quantization condition from the induced potential (4.3), by requiring that it vanishes at x_0. This shows that Φ_{tot} is a multiple of 2π, which is equivalent to Eq. (4.4). Since we assume the plasmon to be classically allowed in all regions, the values of m are bounded from above by the lowest Landau energy $E_{\mathrm{L,min}}$.

Note that multiple reflections from excitations far below y_0 should, in principle, be included. However, these reflections do not introduce new phase contributions due to the quantization condition; they only lead to a multiplicative factor in the amplitude, which contributes to the normalization.

When the plasmon is classically allowed in all regions, q_x and q_y are strictly real, and the exponents in Eq. (4.3) represent plane waves. The amplitude of these plane waves is determined by $\varphi_0(x)$ Eq. (2.43), which captures the screening effects by the dielectric. Spatial variations in the substrate lead to a local change in screening of the electrons and therefore alter the amplitude of the excited plasmon in various ways (e.g. the inverse dependence on the effective dielectric function in Eq. (2.43)).

Besides the direct dependence on the effective dielectric function, the Jacobian and the total momentum are also altered by the change in dielectric environment. For the total momentum, this is evident from the dispersion relation in either region in Fig. 4.1, where we see that for constant energy, the momentum increases for higher screening. On the other hand, the dependence of the Jacobian is more complex. It is given by

$$J = \det \begin{pmatrix} \frac{\partial x}{\partial \tau} & \frac{\partial x}{\partial \alpha} \\ \frac{\partial y}{\partial \tau} & \frac{\partial y}{\partial \alpha} \end{pmatrix}, \tag{4.5}$$

where (τ, α) are parameters representing, respectively, the time evolution and initial conditions that determine the phase space trajectories governed by the Hamiltonian system defined by Eq. (2.26) (see Ref. [1] for a full discussion for a circular symmetric problem). Following the discussion in Refs. [1, 50], we parameterize the trajectories by τ and their y-coordinate at the point x_0. One can show [1, 50] that this results in $\partial x/\partial \alpha = 0$ and $\partial y/\partial \alpha = 1$. These relations not only hold at the initial point where $\tau = 0$, but at all points, as follows by analyzing the variational system [1, 50].

Thus, $J = |\partial x/\partial \tau|$, which is the group velocity in the x-direction. Using Hamilton's equations, the Jacobian can be written as

$$J = \left| \frac{\partial \mathcal{H}_0}{\partial q_x} \left(x, \frac{\partial S}{\partial x}, E \right) \right|, \tag{4.6}$$

evaluated along the classical trajectories. This shows that the dependence of the Jacobian on local screening is complex, as it emerges through the derivative of the effective classical Hamiltonian with respect to q_x. Although this dependence is difficult to analyze analytically, it can be calculated numerically.

So far, we have discussed the setup which will be considered in the following sections, and the various components of the amplitude and their influence on the plasmon excitation in different environments. The amplitude is related to both the electrostatic energy density (Eq. (2.50)) and the leading order of the induced electron density (Eq. (1.60)), given by $n_0(x) = \Pi_0(x) V_{\mathrm{pl}}(x)$. In the following subsections, we consider the latter, similar to Ref. [40], and numerically analyze systems with varying dielectric constants. For this analysis, we use the real part of $V_{\mathrm{pl}}(x)$ to calculate the induced electron density.

4.3.2 The effect of local screening on the localization of the plasmonic excitation

In the previous subsection, we discussed the various ways the plasmonic state depends on the local dielectric screening. Here, we investigate this dependence more directly by analyzing three distinct systems with varying dielectric constants on the right side, while keeping the left-side dielectric constant fixed at $\varepsilon_{\mathrm{b}} = 1$. The dispersion relation for a system with this dielectric constant is shown in Fig. 4.1, and closely matches the dispersion in Ref. [40] (likewise for $\varepsilon_{\mathrm{b}} = 9$). The hard-wall boundary conditions quantize the allowed momenta, establishing a unique relationship between energy and momentum q_y for a given mode number m. For consistency, we fix the energy at $E_{\mathrm{pl}} = 1.2$ eV and the mode number at $m = 3$ throughout this subsection, allowing q_y to vary between systems.

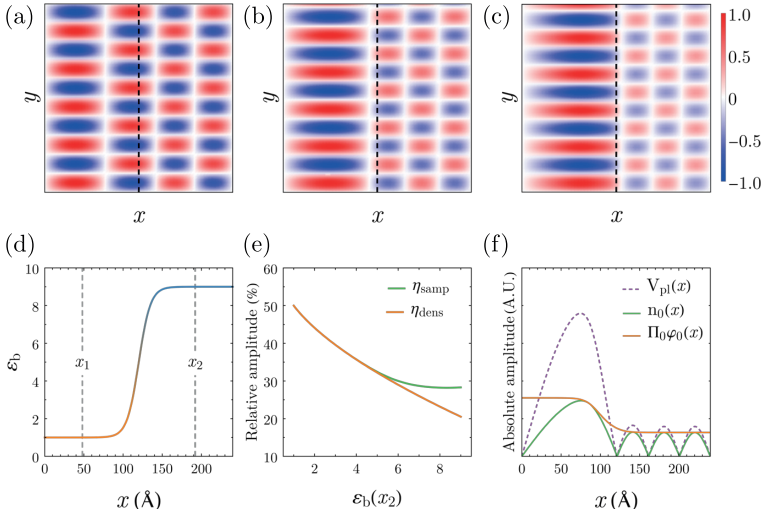


Figure 4.8: Localization of the plasmon in regions with lower screening. (a)-(c) Real-space induced electron density $n_0(x, y)$, plotted for different values of the substrate dielectric constant on the right side. The left side has a dielectric constant of $\varepsilon_{\mathrm{b}}(x_1) = 1$ for all three plots, and the right side is varied, namely (a) $\varepsilon_{\mathrm{b}}(x_2) = 2$, (b) $\varepsilon_{\mathrm{b}}(x_2) = 4$, and (c) $\varepsilon_{\mathrm{b}}(x_2) = 9$. (d) Spatial variation of the dielectric constant $\varepsilon_{\mathrm{b}}(x)$, for $\varepsilon_{\mathrm{b}}(x_1) = 1$ and $\varepsilon_{\mathrm{b}}(x_2) = 9$. (e) Relative amplitude η_{dens} (orange) and η_{samp} (green) of the plasmon excitation on the right. (f) Absolute value of the induced plasmon potential, induced electron density and the amplitude of the electron density over x, for $\varepsilon_{\mathrm{b}}(x_2) = 6$. The orange line corresponds to the amplitude taken for η_{dens}, and the green line for η_{samp}.

Figures 4.8(a)-4.8(c) show the real-space induced electron densities n_0, for the profile of the dielectric constant shown in Fig. 4.8(d). The dielectric constant on the right side takes the values $\varepsilon_{\mathrm{b}}(x_2) = (2, 4, 9)$, from (a) to (c). The transition between the dielectric constants is modeled as a hyperbolic tangent, i.e. $\propto \tanh[x/\ell]$, similar to the boundary described in Sec. 4.2.

The electron density is clearly more localized on the left side of the system, where the dielectric constant is lower. To quantify this localization, we compare the electron density amplitude in the two regions. We define a relative amplitude as

$$\eta_{\mathrm{dens}} = \frac{\Pi_0(x_2)\varphi_0(x_2)}{\Pi_0(x_1)\varphi_0(x_1) + \Pi_0(x_2)\varphi_0(x_2)}, \tag{4.7}$$

which is plotted in orange in Fig 4.8(e), as function of $\varepsilon_{\mathrm{b}}(x_2)$. Note that this relative amplitude does not include the wave-like nature of the plasmon, as it only considers the amplitude.

Figure 4.8(f), shows that the theoretical maximum of the induced electron density, $\Pi_0(x)\varphi_0(x)$, is not always attained due to the finite system size and boundary conditions. Here, the absolute value of the induced potential $V_{\mathrm{pl}}(x)$ (dashed purple) and the electron density $n_0(x)$ (solid green) are plotted for $\varepsilon_{\mathrm{b}}(x_2) = 6$. The solid orange line represents the theoretical maximum $\Pi_0(x)\varphi_0(x)$, which does not account for the wave-like features of $V_{\mathrm{pl}}(x)$. The maximum of the attained electron density is often located near the boundary between the two regions, particularly for higher values of $\varepsilon_{\mathrm{b}}(x_2)$, where half of the wavelength of the plasmon (in the x-direction) becomes comparable to or larger than the size of the left region.

On that note, we can also define a relative amplitude based on the maximum of the attained density in the sample, which is given by

$$\eta_{\mathrm{samp}} = \frac{n_{0,\mathrm{max}}(x_{\mathrm{left}})}{n_{0,\mathrm{max}}(x_{\mathrm{right}}) + n_{0,\mathrm{max}}(x_{\mathrm{left}})}, \tag{4.8}$$

where the maximum of n_0 on the left side is taken over the left and the entire boundary region, while the region on the right side starts where the difference between $\varepsilon_{\mathrm{b}}(x)$ and $\varepsilon_{\mathrm{b}}(x_2)$ is less than 2%. This relative amplitude is plotted in green in Fig. 4.8(e), and shows a clear deviation from the theoretical relative amplitude η_{dens} from roughly $\varepsilon_{\mathrm{b}}(x_2) = 6$.

In summary, this subsection has demonstrated that the localization of the plasmonic excitation is significantly influenced by the local dielectric screening. By analyzing systems with varying dielectric constants, we observed that the electron density tends to localize more on the side with the lower dielectric constant, indicating that (quasi-)localization of the plasmonic excitation is indeed possible under these conditions. This provides valuable insights into the behavior of plasmonic states in heterogeneous dielectric environments. We finish this discussion by noting that the real-space electron density plots and the dependence on the dielectric constant are in very good agreement with Figure 5(g) in Ref. [40]. Their numerical results appear to fall between our theoretical and attained relative amplitude for higher values of $\varepsilon_\mathrm{b}(x_2)$.

4.3.3 Plasmonic waveguide from amplitude effect

In the previous subsection, we demonstrated the possibility of localizing plasmon excitations in regions of lower dielectric screening, and confirmed that our calculations coincide with previous numerical results. Furthermore, Fig. 4.8(e) suggests that localization increases with greater contrast in the dielectric constant between regions. We now consider a waveguide setup analogous to those in Sec. 4.2, but with localization occurring in the region with a lower dielectric constant. Specifically, we consider a central channel with $\varepsilon_\mathrm{b}(x_1) = 1$ and outer regions with $\varepsilon_\mathrm{b}(x_2) = 9$. The dielectric constant profile is given by Eq. (4.2), where x_1 now refers to a point in the center and x_2 to a point in the outer regions. The central channel width is $\ell_\mathrm{w} = 100$ Å, and the total system width is 240 Å, with hard-wall boundary conditions at the outer edges. These boundary conditions imply that the induced potential goes to zero at these edges, leading to the quantization condition Eq. (4.4), which defines a one-to-one relation between the energy and momentum q_y for a given m. In this subsection, we vary different parameters, such as the plasmon energy and the quantum number m, to see how we can control the quasi-localization for waveguiding.

First, we analyze the effect of the plasmon energy. Figures 4.9(a)-4.9(c) show the real-space induced electron density $n_0(x, y)$ for the dielectric environment depicted in Fig. 4.9(d). For the first two plots, (a) and (b), the number of nodes in between the hard-wall boundaries is kept constant at $m = 4$, while the energy is increased from $E_\mathrm{pl} = 1.2$ eV in (a) to 1.4 eV in (b). This energy increase results in a higher momentum q_y, increasing the number of nodes along the waveguide direction.We observe slightly increased localization in the central region with increasing energy.

We also consider the effect of the number of nodes m. In Fig. 4.9(c), the plasmon is excited at $E_\mathrm{pl} = 1.2$ eV (as in (a)) but with a higher number of nodes in the x-direction, $m = 8$. This leads to a lower momentum q_y, evident in the decreased number of nodes along the y-direction. For this increase in m at constant energy, the localization in the central channel decreases.

While the localization varies with energy, momentum q_y, and node number m, the electron density n_0 remains localized in the central channel in all three cases shown in Fig. 4.9(a)-4.9(c). However, at lower energies, the opposite behavior can be observed. For example, at $E_\mathrm{pl} = 0.8$ eV and $m = 4$, the electron density localizes in the high-screening outer regions (Fig. 4.9(e) and 4.9(f)). Figure 4.9(e) shows that while the induced potential $V_\mathrm{pl}(x)$ is highly localized in the low-screening central channel, the induced electron density $n_0(x)$ is not. This suggests that this effect stems solely from the polarization $\Pi_0(x)$, which generally decreases with decreasing momentum $|\mathbf{q}|$ at constant energy. Because the momentum is lower in the central region (Fig. 4.1), the

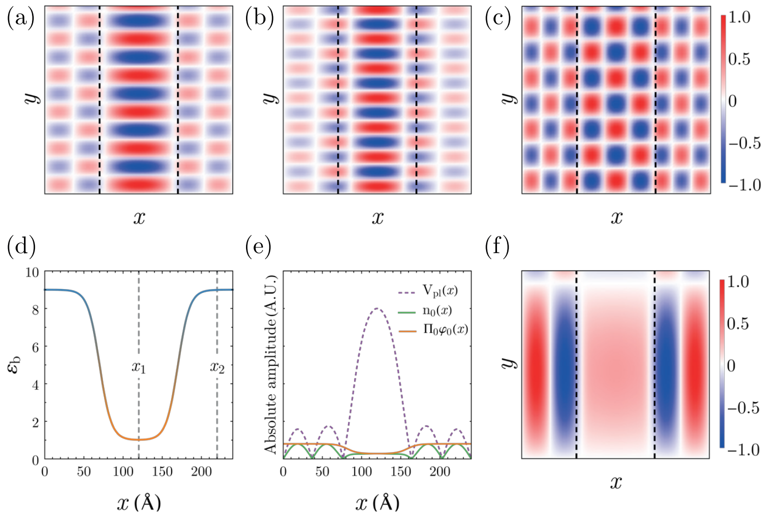


Figure 4.9: Localization of the plasmon in a waveguide setup. (a)-(c) Real-space induced electron density $n_0(x, y)$, plotted for different energies and number of nodes, namely (a) $E_{pl} = 1.2$ eV and $m = 4$, (b) $E_{pl} = 1.4$ eV and $m = 4$, (c) $E_{pl} = 1.2$ eV and $m = 8$. (d) Spatial variation of the dielectric constant as a function of x for the waveguide considered, where $\varepsilon_b(x_1) = 1$ and $\varepsilon_b(x_2) = 9$. (e) Absolute value of the induced plasmon potential V_{pl}, induced electron density n_0 and the amplitude of the electron density φ_0 as function of x, for $E_{pl} = 0.8$ eV and $m = 4$. (f) Real-space induced electron density $n_0(x, y)$, plotted for $E_{pl} = 0.8$ eV and $m = 4$.

polarization is also lower. In the higher energy cases (Fig. 4.9(a)-4.9(c)), the effect of the lower $\Pi_0(x)$ in the middle is less dominant; the localization is primarily driven by the increased amplitude $\varphi_0(x)$ in the low-screening region.

To get an overview of the quasi-localization and its dependence on E, q_y, and m, we plot the bound state spectrum in Fig. 4.10, similar to the plots in Sec. 4.2.2, but now with states above the energy $E_g(x_1, q_y)$. Each curve corresponds to a different value of m, starting with $m = 0$ at the bottom. The color gradient on each curve represents the relative amplitude difference between the central and outer regions, as given by

$$\nu = \frac{\Pi_0(x_1)\varphi_0(x_1) - \Pi_0(x_2)\varphi_0(x_2)}{\Pi_0(x_1)\varphi_0(x_1) + \Pi_0(x_2)\varphi_0(x_2)}. \tag{4.9}$$

Positive (green) values of ν indicate localization in the central, low-screening region, while negative (red) values indicate localization in the high-screening outer regions.

As discussed earlier, the Jacobian has a complex influence on the amplitude, depending on the derivative of the effective classical Hamiltonian with

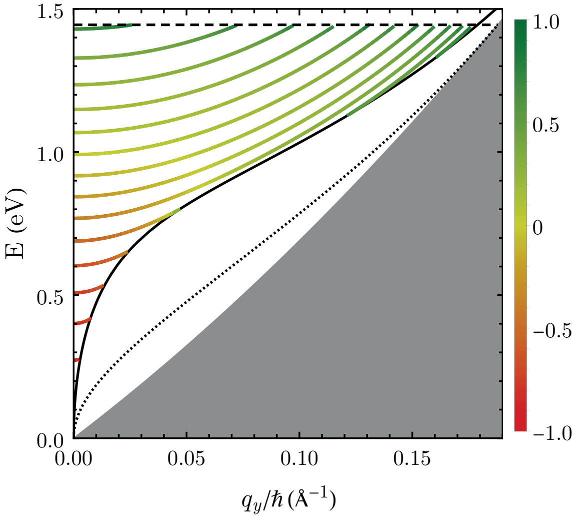


Figure 4.10: Spectrum of plasmonic bound states in a two-dimensional systems with a spatially varying dielectric constant as function of the perpendicular momentum q_y. The color gradient indicates the relative amplitude ν, where localization in the low-screened middle region is depicted with green and in the high-screened edge regions with red. The solid and dotted black lines indicate $E_\mathrm{g}(x_1, q_y)$ and $E_\mathrm{g}(x_2, q_y)$, respectively. Above the energy $E_\mathrm{g}(x_1, q_y)$, in the region where we are interested in, the plasmons are allowed in both spatial regions. In this case, there is no continuum of states, due to the hard-wall boundary conditions. Above the horizontal dashed black line on top, the Landau damped region is reached in the outer regions, which starts from the energy $E_{\mathrm{L,min}} = E_\mathrm{L}(x_2)$.

respect to q_x. Near the gap energy $E_\mathrm{g}(x_1, q_y)$ (solid black curve in Fig. 4.10), where $q_x \to 0$ in the central channel, the Jacobian is generally lower, enhancing localization in the middle. This is visible, for example, at the end of the $m = 4$ bound state curve. Furthermore, near the Landau energy $E_\mathrm{L}(x_2)$ (horizontal dashed black line), the Jacobian becomes large in the outer regions, reducing the plasmon excitation there and increasing central localization.

An analog to the bound state spectrum as function of the momentum q_y, given in Fig. 4.10, can be found in Ref. [40] figure 4(b). However, direct comparison is difficult, due to differences in system setup, with their system resembling the system discussed in the previous subsection. Besides, there is no clear distinction between the number of nodes considered. It is likely that the curve in their plot crosses multiple values of m along the dots.

We have thus seen that (quasi-)localization of plasmons in waveguide setups is possible (Fig. 4.9). The outcome, however, is not straightforward and depends on many different parameters, as is evident from Fig. 4.10. By varying parameters, such as the dielectric constant, plasmon energy, and mode number, we can control the degree of localization. This tunability allows us to achieve either strong or weak (quasi-)localization, depending on the desired application.

4.4 Conclusion

In this chapter, we applied the semiclassical theory to a specific, commonly used, model for the dielectric constant [34, 40, 70], representing a layered structure consisting of a thin film with effective height d, and dielectric constant ε_{M}, encapsulated by a dielectric substrate on both sides (see Sec. 2.2.2). An infinitely thin 2D free electron layer was placed in the middle of the thin film. Within our framework, the dielectric constants, effective height, and electron density can all be treated as spatially varying parameters. This model facilitated the analysis of two distinct types of localized states, based on total internal reflection or screening from the environment.

The first type of localization, discussed in Sec. 4.2.1, through the formation of bound states, arises from the effective classical Hamiltonian Eq. (2.26). This Hamiltonian generates classical plasmon trajectories, which can be periodic. When a classically allowed region is surrounded by classically forbidden regions, these periodic trajectories support bound states upon the application of the quantization condition Eq. (4.1). For 2D plasmons, the formation of these regions is strongly dependent on the momentum q_y. This localization mechanism is analogous to total internal reflection in optical waveguides. We note that these localized states are single states, meaning only one energy is excited per m for a certain momentum q_y. For applications in integrated photonics, it would be interesting to describe a plasmonic band structure, where one can excite multiple states per quantum number m for a certain momentum q_y. Such a band structure might exist in periodic systems or plasmonic crystals, analogous to electronic band structures, which will be discussed in Ch. 5.

In Sec. 4.2.2, we numerically implemented and extensively studied bound states supported by the effective classical Hamiltonian. We elucidated their origin and investigated how the spectrum depends on the spatial variation of different parameters by numerically solving the quantization condition Eq. (4.1). This yielded a dispersion relation connecting the energy, the momentum along the waveguide direction q_y, and a quantum number m. We demonstrated that by varying the dielectric substrate, electron density, or effective height of the middle layer, plasmonic bound states can be created. We systematically studied the conditions for the appearance of these bound states and computed their dispersion. Simultaneous variation of these parameters allowed us to control the lower and upper limits of the bound state spectrum in terms of energies and momenta q_y. These spectra, plotted as a function of the momentum along the waveguide, q_y, provide a direct link to experimental observables. The momentum q_y can be controlled experimentally by for example plasmonic antennas.

The second type of (quasi-)localization, studied in Sec. 4.3, arises from variations in dielectric screening, which modulate the amplitude of the plasmon and therefore the induced electron density. This (quasi-)localization is purely an amplitude effect, distinct from the phase-related localization of bound states. We have found that it is possible to localize the plasmon in regions with higher screening. We have also found that the localization of the plasmon depends on parameters like the substrate dielectric constant, the energy of the excited plasmon, and the momentum along the direction of the waveguide. The underlying mechanism is complex due to the interplay of several factors in the amplitude: the Jacobian, the effective dielectric function, and the total momentum.

In the numerical implementation of these quasi-localized states, we computed the real-space electron density using Eq. (1.60) for various values of the substrate dielectric constant, plasmon excitation energy, and momentum along the waveguide direction. We demonstrated that by varying these parameters, we can control the degree of localization. Notably, we observed that plasmons can be localized in regions with higher screening. Furthermore, the polarization function, Π_0, can lead to counter-intuitive localization behavior at low energies and momenta, where localization shifts to regions of higher screening. While our analysis primarily focused on variations in the dielectric constant, we anticipate similar effects from variations in other parameters, such as the electron density, as these also influence the dispersion and, consequently, the amplitude. The numerical results presented in this section are consistent with those reported in Ref. [40].

We remark that we only plotted the real-space induced potential for the quasi-localized states in Sec. 4.3, and not for the bound states in Sec. 4.2. The latter requires a different approach, since the induced potential Eq. (1.53) diverges at the classical turning points, due to the vanishing of the Jacobian. In other words, the asymptotic solution Eq. (1.53) no longer accurately describes the true solution. An accurate description in the vicinity of turning points could be obtained using an analogous construction in momentum space [48, 50, 55]. This would result in an expression for the induced potential that involves the Airy function. However, this construction has currently not yet been performed for plasmons in inhomogeneous systems and would require additional lengthy derivations [50]. We believe that this would be an interesting future direction, both from a practical and a fundamental point of view.

5

Plasmonic crystals and band structures

In this chapter, we derive an analytical condition for the allowed and forbidden plasmon bands in a quasi-one-dimensional periodic potential. We do this within the framework of semiclassical analysis for plasmons in two-dimensional systems, using the same model for the dielectric environment as in Ch. 4, with a finite height as discussed in Sec. 2.2.2. By applying Bloch's theorem, the problem of determining the allowed energy bands can be reduced to finding the eigenvalues of the transfer matrix for a single potential barrier. For the existence of plasmons, we utilize the Hamilton-Jacobi equation, which determines the plasmon's phase via the classical action. The induced potential can then exhibit an imaginary exponent for propagating waves or a real exponent for damped waves. We consider two distinct scenarios: tunneling (damped waves through the barriers) and above-barrier scattering (propagating waves over the barriers), depending on the momentum q_y. Finally, we apply this theory to two distinct periodic variations, a Gaussian array and the cosine function, illustrating the resulting band structures as functions of q_y.

This chapter details the full research and analysis from unpublished work in collaboration with K. J. A. Reijnders and M. I. Katsnelson.

5.1 Introduction periodic structures

Periodic modulation is a standard technique to tune the characteristics of electromagnetic waves, a principle well established in photonics with photonic crystals [93]. Plasmonics is a rapidly advancing field focused on controlling plasmons. Periodic structures hold significant promise for integrated photonics. The ability to manipulate plasmons at the nanoscale, overcoming the diffraction limit, is crucial for miniaturization. While Ref. [94] discusses one-dimensional periodic plasmonic systems, it does not observe forbidden bands. However, building upon our previous work, we recognize that periodic modulation, in principle, is a standard way to open gaps in the plasmon energy spectrum. Within our semiclassical framework, we consider both tunneling and above-barrier scattering to describe plasmons in a one-dimensional periodic structure. These two types of scattering lead to new quantization conditions, ultimately resulting in forbidden and allowed bands, akin to systems described by the Kronig-Penney model [2, 29]. Our general goal is to derive an analytical condition for the allowed and forbidden plasmon bands in a quasi-one-dimensional periodic potential. We achieve this by employing Bloch's theorem, which reduces the problem to finding the eigenvalues of the transfer matrix for a single potential barrier. We derive the transfer matrix with the complex phase integral method, considering two scenarios: tunneling and above-barrier scattering. The distinction between these scenarios greatly depends on the momentum perpendicular to the modulation, q_y.

In Sec. 5.2, we lay the theoretical foundation for understanding plasmonic band structures in periodic potentials. We extend the semiclassical analysis developed in previous chapters to derive an analytical condition for allowed and forbidden plasmon regions in a one-dimensional periodic potential. By utilizing Bloch's theorem to simplify the problem for a periodic potential, we show that this allows us to reduce the problem to finding the eigenvalues of a transfer matrix for passing a single potential barrier.

A prerequisite for analyzing plasmon scattering in periodic structures is to establish a principle of energy conservation. To this end, in Sec. 5.3, we derive an optical theorem tailored for plasmons. This theorem provides a fundamental statement of energy flux conservation when a plasmon scatters from a potential barrier. Physically, this conservation law imposes a crucial constraint on the system's reflection and transmission amplitudes. This principle is the direct analogue of probability conservation for particles described by the Schrödinger equation [63, 95]. Just as unitarity in quantum mechanics ensures that the probability currents of reflected and transmitted waves account for the entire incident probability flux, this optical theorem ensures a self-consistent, energy-preserving description of the plasmon scattering process.

Section 5.4 delves into the behavior of the plasmon around classical turning points, where the approximate solution for the induced potential is no longer an accurate representation. The standard way to resolve this issue is by employing the complex phase integral method. This general framework provides essential tools for deriving connection formulas and is broadly applied to second-order ordinary differential equations, such as the Schrödinger equation [63, 65, 95, 96]. From it, two powerful solution techniques have emerged to solve the following connection problem: the method of comparison equations and a method pioneered by Zwaan [66]. However, the plasmonic system under consideration is more complex than a standard Schrödinger equation, making a direct and rigorous application of these methods challenging. Nevertheless, throughout this section, we will apply this framework heuristically, justifying our steps with arguments of a physical nature.

More specifically, we will adopt the method pioneered by Zwaan to analytically continue the solution into the complex plane, navigating around the turning points before returning to the real axis. We tackle the connection problem, which connects the approximate solution of the wave before the turning point to the solution of the scattered wave. We derive the analytical relationships that define the conditions for plasmonic band structures. This involves considering two distinct scenarios: tunneling and above-barrier scattering, depending on the momentum q_y. We derive the transfer matrix which allows us to determine the relation for the allowed and forbidden bands. This is similar to the quantization condition for a single well, derived in the previous section, but now gives an upper and lower bound of the plasmonic bands.

Finally, in Sec. 5.5, we perform numerical calculations on our theory for two specific periodic variations: a Gaussian array and a cosine potential. We illustrate the resulting band structures as functions of q_y, showing the allowed and forbidden bands for each potential. We analyze the differences between the two potentials, particularly in terms of the width and shape of the bands, and discuss the implications of these differences for plasmon propagation and localization. This section provides concrete examples of how our theory can be applied to understand and design plasmonic crystals with desired band structures.

5.2 Periodic modulation for plasmons

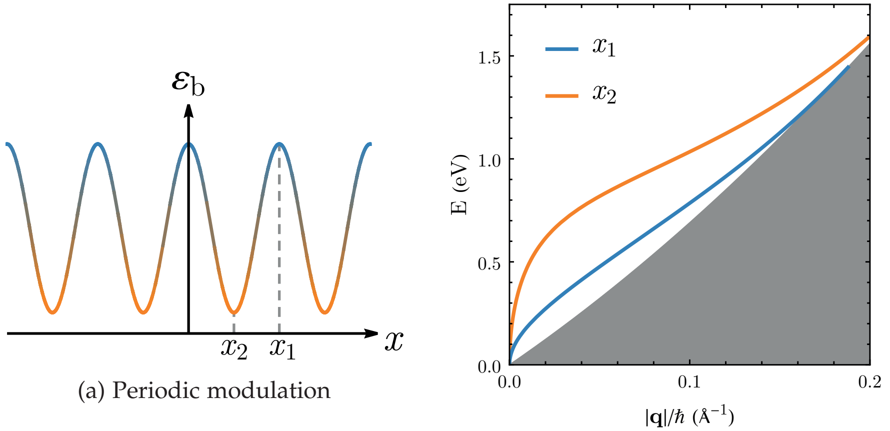
While models like the Kronig-Penney model provide a straightforward approach for calculating the band structure in periodic potential steps separated by homogeneous regions[2, 29], realistic potentials often exhibit smoother variations. These smoother potentials are effectively treated within the semiclassical approximation [63, 86, 97, 98], which is applicable when the length scale of the inhomogeneity is much longer than the electron wavelength. For the Schrödinger equation, these periodic potentials can then be solved employing complex phase integral methods to derive connection formulas and quantization conditions for the band structure. In previous chapters, we have demonstrated the successful application of the semiclassical approximation to plasmons in inhomogeneous environments [1, 50, 69]. Building on these insights, we aim to heuristically apply the complex phase integral methods to derive the band structure for plasmonic systems governed by the semiclassical approximation.

For particles governed by the Schrödinger equation, the distinction between classically allowed and forbidden regions is straightforward: the presence of a particle is classically allowed when its energy exceeds the potential energy of the barrier and forbidden when its energy is below the barrier. This results in either above-barrier scattering for energies higher than the potential or tunneling for energies below the potential.

Two-dimensional plasmons, however, have a gapless dispersion, complicating this picture and suggesting, at first glance, that they are classically allowed everywhere. Nevertheless, Ch. 4 demonstrated that classically forbidden regions can still exist when the variation occurs in only one direction and there is a finite momentum perpendicular to this variation. In this chapter, we consider variations in the x direction, and translational invariance in the y direction. Consequently, the momentum q_y, which is perpendicular to the variation, remains a good quantum number. For finite values of this perpendicular momentum, a gap opens in the dispersion, leading to classically forbidden regions as a function of q_y. We, therefore, expect that the turning points, and consequently the allowed and forbidden energy bands, also depend on this momentum q_y, analogous to the bound state spectrum of a single potential well as discussed in Ref. [69].

Such a periodic potential or periodically forbidden region can be realized through any mechanism that alters the dispersion relation, such as locally doping the material or patterning the dielectric substrate to vary the screening. A schematic example of a periodic modulation of the background dielectric constant, $\varepsilon_\mathrm{b}(x)$, is depicted in Fig. 5.1a. Two specific spatial points are highlighted, namely a maximum in the dielectric function at x_1, and a minimum at x_2. The dispersion relation, $E(x, |\mathbf{q}|)$, calculated from the effective classical Hamiltonian (2.26) at these two specific spatial points is shown in Fig. 5.1b.

In this chapter, we focus on variations in the dielectric environment; however, the theory is not limited to this, as demonstrated in Ch. 4. Let us investigate the classically allowed and forbidden regions. For the following discussion, it is useful to consider the gap energy, as defined in Ch. 4: $E_\mathrm{g}(x, q_y) = E(x, q_y; q_x = 0)$, where $E(x, q_y; q_x) = E(x, |\mathbf{q}|)$ is the local dispersion, with q_y as a parameter. The gap energy, when plotted as a function of q_y (since $q_x = 0$), replicates the dispersion relation shown in Fig. 5.1b, but with q_y as the independent variable instead of $|\mathbf{q}|$. Note that this gap energy can only be defined when q_y is a constant, which is the case when there is translational invariance in the y-direction. Thus, it provides a relationship between energy and the momentum q_y for which a classical turning point is locally present (i.e., $q_x \to 0$).



(a) Periodic modulation

(b) Dispersion relation

Figure 5.1: (a) Schematic representation of a periodically modulated background dielectric constant $\varepsilon_\mathrm{b}(x)$. At x_1, the dielectric constant reaches a maximum, and at x_2 it has a minimum. (b) Dispersion relation for a two-dimensional plasmon for two different local background dielectric constants ε_b. The blue dispersion is for a higher background dielectric constant, such as that at x_1 in (a). This higher dielectric constant increases the plasmon momentum, for a certain energy, pushing the dispersion relation toward the Landau damped region in gray. The orange dispersion is for a lower dielectric constant, e.g., at x_2 in (a).

As discussed in Ch. 4, a classically forbidden (and therefore tunneling) regime exists where the momentum q_x locally becomes imaginary. This occurs for energies between two gap energies: $E_\mathrm{g}(x_2, q_y) > E_\mathrm{p} > E_\mathrm{g}(x_1, q_y)$, where E_p is the energy of the excited plasmon. In Fig. 5.1b, this corresponds to energies (and corresponding q_y values) below the orange curve and above the blue curve. Conversely, for energies above the orange curve (i.e., for $E_\mathrm{p} > E_\mathrm{g}(x_2, q_y)$), the plasmon is classically allowed everywhere. However, it can still exhibit complex turning points due to so-called above-barrier scattering [63, 96]. These turning points were not considered in Ref. [69] because the potential barrier extends indefinitely, which significantly complicates the analysis for the purpose of that discussion.

To analyze the propagation of plasmons through such a periodic system, we first consider the interaction with a single unit cell of the periodic inhomogeneity, for instance, the region between two consecutive maxima of the dielectric constant. The effect of this single barrier on the plasmonic wave, in our case described by the induced potential, can be comprehensively described by a two-by-two transfer matrix, $\bar{T}$. The induced potential is given by Eq. (2.46), or more specifically, since we are solely interested in $z = 0$ and

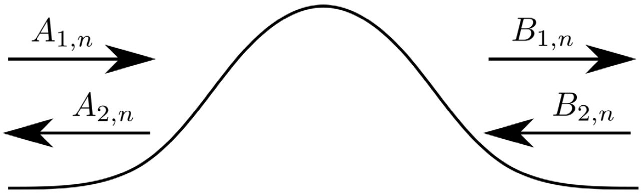


Figure 5.2: Schematic of the n^{th} potential barrier. On both sides of the barrier the plasmonic wave is a superposition of a forward-propagating component (amplitudes $A_{1,n}$ and $B_{1,n}$) and a backward-propagating component (amplitudes $A_{2,n}$ and $B_{2,n}$). The amplitudes on both side can be related to each other through the transfer matrix $\bar{T}$.

variations in the x-direction, given by

$$V_{\text{pl}}(x) = \frac{A(x)}{\sqrt{J(x)\varepsilon_{\text{eff}}(x)|\partial S/\partial \mathbf{x}|}} e^{iS(x)/\hbar} \tag{5.1}$$

The matrix, $\bar{T}$, relates the amplitudes of forward $A_{1,n}$ and backward $A_{2,n}$ propagating waves on the left side of the n^{th} barrier to the corresponding amplitudes $(B_{1,n}, B_{2,n})$ on the right side, as given by

$$\begin{pmatrix} A_{1,n} \\ A_{2,n} \end{pmatrix} = \bar{T} \begin{pmatrix} B_{1,n} \\ B_{2,n} \end{pmatrix}. \tag{5.2}$$

The amplitudes are depicted in Fig. 5.2, as a plasmon propagates through a single unit cell. The elements of the transfer matrix are constructed from the complex reflection r and transmission t coefficients of the single barrier. The elements in terms of these coefficients can be determined by recognizing that a single incoming plasmon from the left, with unitary amplitude results in a reflected and transmitted wave, which directly results in the first two elements [96]. The reverse process, an incoming plasmon from the right gives the other two elements equivalent to complex conjugation. The whole process of passing by a barrier can now be described by [96]

$$\begin{pmatrix} A_{1,n} \\ A_{2,n} \end{pmatrix} = \begin{pmatrix} 1/t & r^*/t^* \\ r/t & 1/t^* \end{pmatrix} \begin{pmatrix} B_{1,n} \\ B_{2,n} \end{pmatrix}, \tag{5.3}$$

where the star denotes complex conjugate. While the calculation of the coefficients is well-established for the Schrödinger equation using semiclassical methods [63, 86, 96–99], their specific form for plasmons requires careful consideration within our formalism, which will be done in the following sections.

Having defined the matrix for passing a single barrier, we now impose the condition of periodicity. For an infinite, perfectly periodic lattice, the system possesses discrete translational symmetry. This symmetry is captured by Bloch's theorem, which constrains the form of the solutions. In the one-dimensional case, this is known as Floquet's theorem (cf. Sec. 4.1.4 in Ref. [2]): $V_{\mathrm{pl}}(x_n) = \exp(-i\phi)V_{\mathrm{pl}}(x_{n+1})$, where ϕ is the Bloch phase [2, 29]. Applying this theorem to the amplitudes of our wave components, the vector of amplitudes at, for example, left of the $n+1^{th}$ barrier can be related to the amplitudes at the left of the n^{th} barrier by this Bloch phase factor, hence

$$\begin{pmatrix} A_{1,n} \\ A_{2,n} \end{pmatrix} = e^{-i\phi} \begin{pmatrix} A_{1,n+1} \\ A_{2,n+1} \end{pmatrix}. \tag{5.4}$$

Bloch's theorem requires the system to be identical on both sides of the barrier, and therefore the amplitudes should be taken in the same relative location within the adjacent unit cells. Note that, the transfer matrix shown in Eq. (5.3) only passes to the right side of the barrier. Before we can impose Bloch's theorem, we have to pass to the left side of the adjacent barrier, and therefore switch from amplitude $B_{i,n}$ to $A_{i,n+1}$. This process results in multiplication with an extra transfer matrix, which will be elaborated in Sec. 5.4.

By combining the physical propagation described by the transfer matrix with the symmetry constraint of Bloch's theorem, we arrive at the central condition for the existence of propagating wave solutions. Equating the two descriptions for the amplitudes yields an eigenvalue problem for the total transfer matrix. The allowed energy bands of the plasmonic crystal correspond to energies for which the Bloch phase ϕ is real, which requires the eigenvalues of $\bar{T}$ to be complex numbers with unit modulus, i.e., of the form $\exp(-i\phi)$. Conversely, energies that result in eigenvalues with a modulus different from one correspond to evanescent waves and lie within the band gaps. Before we can solve this eigenvalue problem to find the band structure, we must find the reflection and transmission coefficients. Before that we will establish a conservative relationship between the reflection and transmission coefficients, analogous to the optical theorem. This ensures the physical consistency of our transfer matrix, a topic we will address in the next section.

5.3 The optical theorem for plasmons

For the conventional Schrödinger equation, the unitarity condition for Hermitian Hamiltonians ensures conservation of probability density, and therefore particle or energy flow during a lossless scattering process. This unitarity condition leads to the optical theorem, which relates the amplitude of a scattered particle before and after the process according to $|r|^2 + |t|^2 = 1$ [63, 95]. This equation essentially states that for a lossless system, the total probability of reflection ($|r|^2$) and transmission ($|t|^2$) must equal the probability of the incoming wave. For plasmons, the optical theorem, which describes a similar relationship between scattering amplitudes and energy conservation, has been studied using different theories in Ref. [47].

Here, we aim to derive a condition analogous to the optical theorem within our semiclassical analysis of plasmons, governed by the Poisson equation (2.1). We focus on the lowest order in $\hbar$, which yields

$$\langle \nabla, \varepsilon(\mathbf{x}, z)\nabla \rangle \, \Gamma_0(\mathbf{x}, z) V_{\text{pl}}(\mathbf{x}) = -4\pi e^2 \Pi_0(\mathbf{x}) V_{\text{pl}}(\mathbf{x}). \tag{5.5}$$

In this formulation, Γ_0 is the principal symbol of the operator $\hat{\Gamma}$, which can be interpreted as a generalization of the instantaneous eigenfunction of the Poisson equation within the adiabatic approximation (see Ch. 2 for the explicit derivation).

We decompose the derivatives ∇ into in-plane and out-of-plane components. Multiplying both sides of Eq. (5.5) by V_{pl}^* and subtracting the complex conjugate of the resulting equation from the original equation yields an equation of the form: $fg^* - f^*g = 2i\text{Im}(f^*g)$. From this, we can readily see that the right-hand side cancels if we stay away from the Landau-damped region, i.e., $\text{Im}(\Pi_0) = 0$. Furthermore, for Hermitian electronic Hamiltonians, Γ_0 is purely real [1, 69], which results in $\text{Im}(\partial_z \varepsilon(\mathbf{x}, z) \partial_z \Gamma_0(\mathbf{x}, z)) = 0$, assuming a real and positive dielectric constant. This eliminates the out-of-plane components. Therefore, the remaining terms are given by

$$V_{\text{pl}}^*(\mathbf{x}) \, \langle \partial_{\mathbf{x}}, \varepsilon(\mathbf{x}, z)\partial_{\mathbf{x}} \rangle \, \Gamma_0(\mathbf{x}, z) V_{\text{pl}}(\mathbf{x}) - V_{\text{pl}}(\mathbf{x}) \, \langle \partial_{\mathbf{x}}, \varepsilon(\mathbf{x}, z)\partial_{\mathbf{x}} \rangle \, \Gamma_0^*(\mathbf{x}, z) V_{\text{pl}}^*(\mathbf{x}) = 0, \tag{5.6}$$

where the inner product $\langle a, b \rangle$ is now the standard inner product on $\mathbb{R}^2$, instead of $\mathbb{R}^3$.

As stated in the previous section, we are interested in periodic modulations in only one direction, the x direction. This greatly simplifies the dependence of the induced potential on the perpendicular y variable. The action can be decomposed into two distinct terms corresponding to the two directions, where the y component is simply given by $S_y = q_y y$ (this will be discussed in detail in the next section). Due to the fact that both $\varepsilon(x, z)$ and $\Gamma_0(x, z)$ do not depend on y in this case, it is easily shown that the derivatives with respect to y cancel out in the above equation, since $\text{Im}\left(V_{\text{pl}}^*(\mathbf{x})\varepsilon(x, z)\Gamma_0(x, z)\partial_y^2 V_{\text{pl}}(\mathbf{x})\right) = 0$. Therefore, we are left with only the derivatives with respect to x. The relation between the induced potential at $z = 0$ before and after interaction with a potential barrier is of primary concern. Therefore, considering only the active electron layer at $z = 0$ and assuming the existence of plasmons, which allows us to apply the Hamilton-Jacobi equation, for which $\Gamma_0(x, z = 0) = 1$ (see Sec. 2.1.2), results in

$$\partial_x \left(V_{\text{pl}}^*(\mathbf{x})\varepsilon(x, 0)\partial_x V_{\text{pl}}(\mathbf{x}) - V_{\text{pl}}(\mathbf{x})\varepsilon(x, 0)\partial_x V_{\text{pl}}^*(\mathbf{x}) \right) = 0, \tag{5.7}$$

which implies

$$\text{Im}\left(V_{\text{pl}}^*(\mathbf{x})\varepsilon(x, 0)\partial_x V_{\text{pl}}(\mathbf{x}) \right) = \text{const.} \tag{5.8}$$

This equation differs slightly from the standard optical theorem for the Schrödinger equation due to the different form of the induced potential compared to the wavefunction and the additional dielectric constant. Note that, the derivation of Eq. (5.8) neglects a term with the derivative of Γ_0 with respect to x, because it leads to higher-order terms in $\hbar$ (cf. appendix C.2).

In order to understand the physical importance of this result, let us consider an incoming plasmon from $x \to -\infty$ interacting with a potential barrier of finite width. We evaluate Eq. (5.8) at point a, after the barrier, and point b, before the barrier. Since Eq. (5.8) must be constant, we can equate its values at these two points. We substitute the induced potential Eq. (2.46) at $z = 0$ with amplitude factor $A_0^0 = 1$ for the incoming wave and consider the other amplitude terms given by $\varphi_0(x) = 1/\sqrt{J(x)\varepsilon_{\text{eff}}(x)|\partial S/\partial x|}$, which results in reflected and transmitted waves with amplitude factors r and t, respectively. Equating the expressions on both sides of the barrier gives

$$|t|^2 \frac{\varepsilon(x,0)}{J(x)\varepsilon_{\text{eff}}(x)}\bigg|_{x=x_a} + |r|^2 \frac{\varepsilon(x,0)}{J(x)\varepsilon_{\text{eff}}(x)}\bigg|_{x=x_b} = \frac{\varepsilon(x,0)}{J(x)\varepsilon_{\text{eff}}(x)}\bigg|_{x=x_b}. \tag{5.9}$$

Here, only the derivatives of the exponent in V_{pl} are retained, keeping only the leading order terms in $\hbar$, and therefore maintaining the accuracy of the theory [63].

Let us reflect on this result. For the Schrödinger equation, the unitarity condition ensures that probability is conserved. However, as discussed in Sec. 2.1.4, in the case of plasmons, the modulus of the induced potential does not represent a probability density; this role is fulfilled by the energy density. Although the energy density is conserved, the amplitude of the induced potential does not necessarily have to be conserved. As can be seen in Eq. (5.9), one has to take the local Jacobian and effective dielectric function into account, which might differ on either side of the barrier. Nevertheless, when the momentum and dielectric constant are equal on both sides of the barrier, this conservation condition reduces to the familiar Schrödinger case: $|t|^2 + |r|^2 = 1$. However, when this is not the case, one has to carefully consider the amplitude $\varphi_0(x)$ in the transmission and reflection coefficients.

5.4 The connection problem

The induced potential is an accurate representation of the plasmon wavefunction, in the context of the semiclassical approximation. However, it is not an accurate representation of the true solution near turning points, where $J \to 0$, which causes the amplitude φ_0 to diverge. In Ch. 4, this divergence was circumvented by avoiding these points and only analyzing the region between two turning points. A suitable Maslov index was used to describe the plasmon's behavior around the turning point when it reverses direction.

To determine the reflection and transmission coefficients, r and t, for a symmetric potential barrier, we must solve the connection problem across its two turning points. Following the approach pioneered by Zwaan [66], we circumvent the breakdown of the semiclassical approximation at the turning points by analytically continuing the solution into the complex position plane. This method involves integrating along a carefully chosen contour that bypasses the singularities associated with the turning points. While this technique and its underlying connection formulas are extensively studied and rigorously established for second-order ordinary differential equations [63, 86, 95–99], their application to our plasmonic system will be based on heuristic arguments. We will justify this extension by drawing strong analogies to the quantum mechanical problems where this method has been successfully applied.

We start by considering analytic continuation into the complex plane of the the induced potential Eq. (5.1). The action in Eq. (5.1) can be decomposed into two separate directions. For instance

$$S(x, y) = \int_{x_0}^{x} q_x(x')dx' + \int_{y_0}^{y} q_y(y')dy'. \tag{5.10}$$

As discussed before, throughout this chapter, we consider variations in only one direction, namely the x direction. Consequently, q_y is constant and serves as a good quantum number. The action can then be simplified to

$$S(x_0, x) = \int_{x_0}^{x} q_x(x')dx' + yq_y, \tag{5.11}$$

where the initial condition y_0 is arbitrary due to the translational invariance. For convenience, we set $y_0 = 0$. The coordinate y is also effectively arbitrary, as it only contributes a constant additional phase factor for constant y. In contrast, the initial condition for the action in x is not arbitrary, since we must consider the periodic variation in that direction. This variation may induce turning points, requiring careful treatment, as we discuss shortly.

We proceed by extending the induced potential formalism to the complex domain by considering complex values for x, denoted by ξ. We express the induced potential in a form convenient for the subsequent discussion. We have two solutions, representing right- and left-moving plasmons relative to a (complex) reference point ξ_0, which we write as

$$V_1(\xi_0, \xi) = \frac{A_1}{\sqrt{J(\xi)\varepsilon_{\text{eff}}(\xi)\sqrt{(\partial S/\partial \xi)^2 + q_y^2}}} e^{iS(\xi_0,\xi)/\hbar} \tag{5.12}$$

$$V_2(\xi_0, \xi) = \frac{A_2}{\sqrt{J(\xi)\varepsilon_{\text{eff}}(\xi)\sqrt{(\partial S/\partial \xi)^2 + q_y^2}}} e^{-iS(\xi_0,\xi)/\hbar}, \tag{5.13}$$

where ξ is the complex continuation of the coordinate x, and ξ_0 is a reference point (often chosen to be a turning point).

For a plasmon with momentum q_y and energy below the gap energy $E_\mathrm{p} < E_\mathrm{g}(x_2, q_y)$, real simple turning points exist, where q_x goes to zero. In the vicinity of such a simple turning point, for example, x_0 on the real axis, the momentum behaves like $q_x \propto \sqrt{x - x_0}$ (cf. App. D), vanishing at the turning point. The analytic continuation of the momentum q_x for the Schrödinger equation has been extensively discussed in references [63, 98]. For our wavefunction, it just follows the rules of a simple turning point for the conventional Schrödinger equation, as long as it has this square root form.

The amplitude, however, differs from the conventional $\propto J(x)$, containing additional terms dependent on x and $\mathbf{q}$. The analytic continuation of these terms is, nevertheless, rather simple, because these terms, excluding the Jacobian, depend only on the magnitude of the momentum $\mathbf{q}$, which is always real and positive, and therefore do not acquire additional phase factors during integration in the complex plane. The Jacobian also differs from the for the Schödinger equation conventional $J(x) \propto q_x$ (cf. Sec. 1.3.1), as it is given by $J(x) = dx/dt = \partial \mathcal{H}_0 / \partial q_x$. As discussed in Ch. 3 [1], it behaves like

$$J(x) \approx -q_x \frac{g_\mathrm{s} e^2 p_\mathrm{F}^2}{2m\varepsilon_\mathrm{eff} \hbar E^2 |\mathbf{q}|}, \tag{5.14}$$

in the limit of small q_x. Thus, near a turning point, where q_x becomes small, and non-zero q_y, we have $J(x) \propto q_x$, similar to the Schödinger equation, justifying the assumption that its analytic continuation around the turning point follows the rules for the Schödinger equation. With similar argumentation, we see that the rules for passing branch cut's in the complex plane are also the same as for a Schrödinger-like particle [63].

In order to visualize what happens in the complex plane, we introduce so-called Stokes diagrams. We consider two distinct scenarios: tunneling and above-barrier scattering, depending on the momentum q_y as discussed before. Subsequently, using Bloch's theorem, we find the eigenvalues of this transfer matrix to determine the allowed and forbidden energies.

5.4.1 Stokes diagrams

We recall that the goal is to determine the transfer matrix $\bar{T}$ for passing a barrier, which relates the amplitude of the induced potential before and after the barrier. Combining this with Bloch's theorem, we can then calculate the eigenvalues of this matrix, yielding the allowed and forbidden energy bands. As discussed before, close to a turning point, we may heuristically apply the rules for the Schrödinger equation for traversing turning points in the complex plane. In this section, we loosely follow and utilize the results from [63] and [96], while also recommending references like [86, 95, 98] for more detailed information. Our focus is on determining the transmission and reflection coefficients, t and r, which are related to each other through the optical theorem.

To construct the transfer matrix with the transmission and reflection coefficients, we must first define the turning points for plasmons and examine their dependence on the momentum q_y. A turning point, ξ_0, is defined as the point in the complex plane where $q_x \to 0$. From each simple turning point, three lines emanate, called anti-Stokes lines, where $\mathrm{Im}(S) = 0$. These lines solely allow for propagating waves, and no exponentially decaying components. On the real axis this corresponds to what was previously called the classically allowed region. On the other hand, three different lines emanate from each turning point, called Stokes lines, for which $\mathrm{Re}(S) = 0$. On these lines, we only have an exponentially decaying part of the solution, and no propagation part. It is informative to draw so-called Stokes diagrams, to clarify how the turning points and (anti-)Stokes lines, and therefore the barrier depend on the momentum q_y.

In addition to Stokes and anti-Stokes lines, the Stokes diagrams must also incorporate branch cuts. A branch cut is a curve that emanates from each turning point, which serves as a branch point for the momentum function $q_x(x)$. The necessity for these cuts arises because q_x is determined by solving the local dispersion relation, a process that typically involves a square root of a complex variable. The square root is inherently a multi-valued function; any non-zero complex number has two distinct square roots. If one follows a continuous path that encircles a branch point, the function does not return to its initial value but rather to the value on its other branch. To render $q_x(x)$ single-valued and well-defined, we introduce these cuts. While the specific placement of the branch cuts is an arbitrary choice made for mathematical convenience, this choice must be applied with strict consistency: any integration contour in the complex plane that crosses a branch cut must account for the resulting discontinuity by transitioning to the other branch of the function.

In the tunneling regime, for energies and momenta q_y in between the blue and orange line in Fig. 5.1b, all turning points are real. A simplified Stokes diagram for tunneling is shown in Fig. 5.3. We denote the turning points of the n^{th} barrier as a_n and b_n, representing the beginning and end of the tunneling region, respectively. For symmetric barriers, the turning points are symmetric about the point of maximum potential (or minimum dielectric constant), e.g. x_2 in Fig. 5.1a. Both turning points depend on q_y: they move closer to the maximum potential as q_y decreases, until they merge into a single second-order turning point at the edge of the tunneling regime, where the plasmon energy is equal to the gap energy $E_\mathrm{p} = E_\mathrm{g}(x_2, q_y)$. The next subsection discusses how to handle this higher-order turning point, which introduces an additional phase factor.

In the above-barrier scattering regime, for energies and momenta q_y above the orange line in Fig. 5.1b, the turning points are located in the complex plane. A simplified Stokes diagram is shown in Fig. 5.4. The turning points of the n^{th} barrier are again denoted by a_n and b_n, where a_n is in the lower half and b_n in the upper half of the complex plane. They are situated above and below the maximum potential at x_2 and are symmetric about the real axis. For small q_y, the turning points move further from the real axis, while for larger q_y, they approach each other, eventually merging into the same second-order turning point at $E_\mathrm{p} = E_\mathrm{g}(x_2, q_y)$ on the real axis, discussed in the previous paragraph.

Figures 5.3 and 5.4 depict simplified Stokes lines, which are not necessarily linear in reality. For bound potentials, as in our periodic modulation, the two anti-Stokes lines in the above-barrier regime do not approach the real axis, but asymptotically reach a finite value far from the turning point. Since we consider paths in the complex plane close to the turning points, the behavior of the (anti-)Stokes lines far from the turning point is not critical, provided the assumption that no other turning points are in the vicinity.

Next, we must choose a path in the complex plane around the turning point. In the tunneling regime, this avoids the singularity at the turning point while accurately describing the plasmon beyond it. In the above-barrier scattering regime, the turning points are initially in the complex plane and might seem irrelevant when moving over the real axis. However, this would incorrectly predict unitarity transmission. To incorporate an exponentially small reflection, we must choose a specific path around the turning point, returning to the real axis afterward.

5.4.2 Quantization condition for tunneling

We first consider the case of tunneling through a potential barrier. As previously discussed, a simplified schematic of the Stokes diagram for this regime is shown in Fig. 5.3. For plasmons, real turning points appear when the momentum perpendicular to the periodic modulation, q_y, is sufficiently large (see Sec. 5.2). The turning points of the n^{th} barrier are denoted by a_n and b_n, representing the start and end of the tunneling region, respectively.

For the Schrödinger equation, the reflection and transmission coefficients have been extensively studied [63, 86, 96–99]. In semiclassical analysis, careful attention must be paid to the chosen path in the complex plane to ensure the accuracy of the asymptotic expansion in terms of $\hbar$ [63]. In the case of plasmons, the analytic continuation of the wavefunction follows a similar procedure to that of the Schrödinger equation, provided that the periodic modulation does not introduce additional phases via branch cuts or singularities.

When the turning points are sufficiently separated, they can be treated as independent. We connect the transmitted wave to the incoming wave by analytically continuing the solution along a path in the complex position plane that bypasses the turning points. The choice of this contour is dictated by the fundamental requirement that the solution must be subdominant (exponentially decaying) along the large semicircular part of the path, ensuring its a unique solution [63, 96].

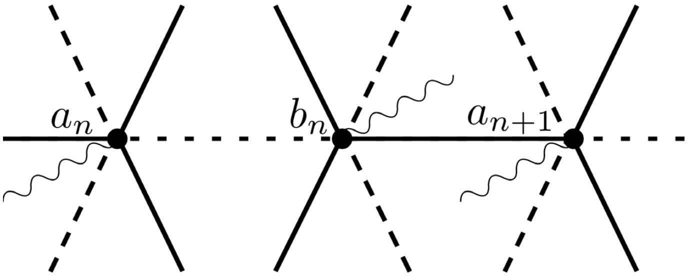


Figure 5.3: Stokes diagram for tunneling of plasmons through an overdense barrier. Dashed lines are Stokes lines, where $\mathrm{Re}(S) = 0$, and therefore only evanescent waves exist. Solid lines are anti-Stokes lines, where $\mathrm{Im}(S) = 0$, and therefore only propagating waves exist. The turning points are denoted with a_n and b_n, where a_n is considered the start of the tunneling regime, and b_n the end on the real x-axis. Branch cuts, given by wavy lines, emanate symmetrically from the turning points, meaning the branch cut emanating from turning point a_n is a point reflection of the branch cut emanating from the turning point b_n.

We are interested in the transmitted wave, which propagates forward in the positive x-direction. For the specific plasmon dynamics in our system, the Jacobian in Eq. (3.1) establishes that the group velocity is anti-parallel to the momentum ($v_x \propto -q_x$). Therefore, a forward-propagating plasmon is necessarily described by a negative momentum. According to the Riemann-Cauchy equations [63], a wave with negative momentum is mathematically subdominant on a contour in the lower half of the complex plane. Consequently, to correctly isolate the subdominant unique transmitted wave, we must choose our integration path in the lower half-plane.

This leads to the same contour choice as in Ref. [63]. In that work, the formulation also results in a right-moving wave having a negative momentum, which is therefore subdominant in the lower half-plane. The critical distinction lies in the origin of this property: in our case, the negative momentum for a forward wave is a direct physical consequence of the plasmon dispersion, rather than a feature of a specific mathematical convention of wave dynamics.

When the turning points are sufficiently separated, they can be treated as independent points. In this regime, a contour in the lower half of the complex plane can be chosen to connect the transmitted wave (amplitude factor t) to the incoming wave (amplitude factor 1), passing both turning points. Because the branch cuts are chosen to be symmetric, a contour that connects the transmitted wave to the incoming wave by passing through the lower half of the complex plane will necessarily cross one branch cut. We take a path which starts from the real axis after the second turning point, b_n, and ends on the real axis before the first turning point, a_n, thereby enclosing the tunneling region. This path, encircling the turning points in the lower half-plane, results in the transmitted wave being exponentially damped relative to the incoming wave. The transmission coefficient follows the semiclassical form [63, 96]

$$t = e^{-K/\hbar}, \tag{5.15}$$

where $K = \int_{a_n}^{b_n} |q_x| dx$ is real and positive, representing the classical action associated with tunneling through the barrier. As the separation between turning points increases, K grows, leading to an exponential suppression of transmission. This is consistent with standard WKB results for quantum mechanical barriers for the Schrödinger equation [63, 86, 96–99].

The reflection coefficient in this case is given by [63, 96]

$$r = ite^{K/\hbar} = i, \tag{5.16}$$

which appears to suggest unitarity reflection, as it implies that the wavefunction undergoes only a phase shift without loss of amplitude. However, this contradicts the presence of tunneling, as we cannot simultaneously have complete reflection and finite transmission. This apparent paradox arises because the above expression neglects additional corrections beyond the leading-order semiclassical approximation. In particular, using the optical theorem derived in Sec. 5.3, we obtain a more accurate description that ensures that the magnitudes of the reflection and transmission coefficients satisfy the necessary energy conservation constraints, resolving the apparent contradiction.

Consequently, by using the optical theorem, we can find the magnitude of the coefficients, valid independent of the chosen path and valid if the turning points are close together [63, 96]. The magnitude of the transmission coefficient is

$$|t| = e^{-K/\hbar} \left(1 + e^{-2K/\hbar}\right)^{-1/2}. \tag{5.17}$$

The reflection coefficient can then be written as

$$|r| = \left(1 + e^{-2K/\hbar}\right)^{-1/2}. \tag{5.18}$$

These expressions provide only the magnitudes of the coefficients. To fully specify the complex coefficients, we must determine their phases.

While a rigorous derivation is beyond the scope of this discussion, we can reconstruct the full coefficients by combining the magnitude information with known phase behaviors. For example, in the limit of well-separated turning points, the reflection coefficient acquires a phase of i. Moreover, as the turning points move closer, an additional phase shift, θ, develops [96–98], which evolves in the phase shift associated with a second-order turning point. Therefore, the full reflection and transmission coefficients can be expressed as

$$r = e^{i\pi/2+i\theta} \left(1 + e^{-2K/\hbar}\right)^{-1/2}, \tag{5.19}$$

$$t = e^{i\theta} e^{-K/\hbar} \left(1 + e^{-2K/\hbar}\right)^{-1/2}, \tag{5.20}$$

where θ is an additional phase that vanishes when the turning points are well separated, but becomes non-negligible when the turning points are close together.

In Ref [96, 98], this phase change can be captured by a single phase factor θ, which is given by

$$\theta = \arg\left[\Gamma\left(\frac{1}{2} + \frac{iK}{\pi\hbar}\right)\right] + \frac{K}{\pi\hbar} - \frac{K}{\pi\hbar}\ln\left(\frac{|K|}{\pi\hbar}\right). \tag{5.21}$$

This formula is derived using the method of comparison equations for the case where two turning points are sufficiently close applied to a second-order ordinary differential equation. In this case, the equation for the Weber function can be applied, whereas in the case of isolated turning points, the simpler Airy equation is sufficient [63, 96–98]. While this derivation is formally for a second-order ordinary differential equation, the reasoning is applicable to our system. This is because the derivation uses the Weber equation as the comparison equation, which is still applicable for our system. However, we have not applied the method of comparison equations explicitly to our system of equations. Nevertheless, we assume the result of Eq. (5.21) holds for our system without re-deriving it.

Now we have derived the formulas for the transmission and reflection coefficients, we can turn back to the transfer matrix given by Eq. (5.3). As stated before, we have to multiply with a second transfer matrix that transfers the amplitudes to the next turning points a_{n+1}. This matrix is given by

$$\begin{pmatrix} B_{1,n} \\ B_{2,n} \end{pmatrix} = \begin{pmatrix} e^{L/\hbar} & 0 \\ 0 & e^{-L/\hbar} \end{pmatrix} \begin{pmatrix} A_{1,n+1} \\ A_{2,n+1} \end{pmatrix}, \tag{5.22}$$

with $L = i\int_{b_n}^{a_{n+1}} q_x dx$, where q_x is real now. The integral L is thus imaginary, and describes the action of traversing between two barriers. Performing a matrix multiplication to get the full transfer matrix over one periodic period, yielding

$$\begin{pmatrix} A_{1,n} \\ A_{2,n} \end{pmatrix} = \begin{pmatrix} \frac{1}{t}e^{L/\hbar} & \frac{r^*}{t^*}e^{-L/\hbar} \\ \frac{r}{t}e^{L/\hbar} & \frac{1}{t^*}e^{-L/\hbar} \end{pmatrix} \begin{pmatrix} A_{1,n+1} \\ A_{2,n+1} \end{pmatrix}. \tag{5.23}$$

As discussed in Sec. 5.2, because we are considering a periodic system, we may apply Bloch's theorem Eq. (5.4). This implies that $\exp(-i\phi)$ has to be an eigenvalue of the full transfer matrix $\bar{T}$.

By solving for the eigenvalues of this matrix, we find a relation between the allowed energies, the quantum number m, and the momentum q_y. Taking the determinant of the matrix in Eq. (5.23), we find the eigenvalue secular equation, given by

$$\lambda^2 - 2\lambda|t|^{-1}\cos\left(\frac{L}{\hbar} + \theta\right) + 1 = 0, \tag{5.24}$$

where $\lambda = \exp(\pm i\phi)$. It can be readily shown that this results in the following quantization condition [63, 97, 99]

$$\cos(\phi) = |t|^{-1} \cos\left(\frac{L}{\hbar} + \theta\right). \tag{5.25}$$

where, conventionally, real values of ϕ correspond to propagating Bloch waves (allowed energy bands), while complex values of ϕ would indicate evanescent waves (forbidden energy bands or band gaps). In Ref. [63], it is shown that the Bloch phase can be expressed in terms of the allowed quantum state ν and the total number of barriers N. For a finite number of barriers, it is shown that $\phi = \nu\pi/(N+1)$, with N the number of barriers, and ν the quantum state, within each band m. In our case of infinite barriers, ϕ gives a continuous number of states within the boundaries $\cos(\phi) = \pm 1$, corresponding to the first and final state, respectively. From these considerations, we expect that the states in the allowed continuum are distributed according to $\cos(\phi) \propto \cos(\nu)$. The density of states would therefore reach a maximum at the edges and decreases inwards, with minima in the middle of each band.

Following Refs. [63, 97], we can use the identity that if

$$\cos\left(\frac{L}{\hbar} + \theta\right) = \left(1 + e^{2K/\hbar}\right)^{-1/2}, \tag{5.26}$$

then

$$\tan\left(\frac{L}{\hbar} + \theta\right) = e^{K/\hbar}, \tag{5.27}$$

to write the condition for the edges of the allowed bands as

$$\frac{L}{\hbar} + \theta = \left(m - \frac{1}{2}\right)\pi \pm \tan^{-1} e^{-K/\hbar}, \tag{5.28}$$

where we also used the identity $\tan^{-1}(e^{K/\hbar}) = \frac{1}{2}\pi - \tan^{-1}(e^{-K/\hbar})$. This equation is the condition for the m^{th} band in the tunneling regime, where the ± describes the top and bottom edge of the allowed energy bands, respectively, and m can be any positive integer. The first part of Eq. (5.28) (apart from θ) is exactly the same as the quantization condition for a single well discussed in Ch. 4. The quantum number m is therefore the band equivalent of the quantum state in a single well, each of which splits up due to the last factor and becomes a band of continuous states in the case of infinite periodic potentials.

In summary, the tunneling regime is characterized by plasmons that penetrate through potential barriers. This interaction between the plasmon states in adjacent classically allowed regions lifts their degeneracy and leads to the formation of continuous energy bands. The condition given by Eq. (5.28) can be interpreted as the quantization condition that defines these bands, analogous to the formation of molecular orbitals from atomic orbitals, resulting in the formation of bonding and anti-bonding superpositions. Within this framework, the $\pm$ signs in Eq. (5.28) determine the energies of these bonding and anti-bonding states, thus defining the band edges. The phase factor θ. derived from the method of comparison equations, becomes crucial when the turning points approach each other and can no longer be treated independently. For plasmons, this occurs when the energy is close to the top of the potential barrier, $E_g(x_2, q_y)$, for a given q_y.

5.4.3 Quantization condition for above-barrier scattering

For energies above $E_{\mathrm{g}}(x_2)$, corresponding to smaller values of q_y, real turning points no longer exist. Instead, for the n^{th} barrier, we find two symmetric complex conjugate turning points, denoted by a_n and b_n, as depicted by the Stokes diagram in Fig. 5.4. This situation gives rise to the non-classical phenomenon of above-barrier reflection, which is fundamentally different from the behavior of classical particles. A classical particle with energy exceeding the potential maximum would simply pass over the barrier with unit probability and no reflection. Plasmons, however, behave as waves, and their partial reflection is physically expected because the potential barrier acts as a region of smoothly varying impedance. In wave physics, any such change in a medium's properties, even one that does not create a classically forbidden region, can act as a source of scattering. The complex conjugate turning points are the mathematical signature of this effect [63, 65, 96]; their presence, though off the real axis, signifies that the potential is structured enough to disturb the wave's propagation and generate a reflected component, which can open up a gap in the spectrum.

Applying similar steps as in the previous subsection, using the optical theorem for above-barrier scattering, the transmission and reflection coefficients are given by [63, 96]

$$t = e^{i\theta}\left(1 + e^{-2K/\hbar}\right)^{-1/2}, \tag{5.29}$$

and

$$r = e^{i\pi/2 + i\theta} e^{-K/\hbar}\left(1 + e^{-2K/\hbar}\right)^{-1/2}, \tag{5.30}$$

where $K = \int_{a_n}^{b_n} |q_x| dx$ is again real and positive, representing the action associated with traveling around the complex turning points in the above-barrier scattering regime. The integral K still grows as the separation between the turning point increases, but, unlike the tunneling regime, for the above-barrier regime, this leads to exponential suppression of the reflectivity. Here, a_n and b_n are the complex conjugate turning points, satisfying $\mathrm{Im}(a_n) < 0$ and $\mathrm{Im}(b_n) > 0$.

The behavior of these complex turning points, and therefore the scattering coefficients, strongly depends on the momentum q_y. A smaller value of q_y corresponds to a plasmon energy E_{p} that is higher above the potential barrier

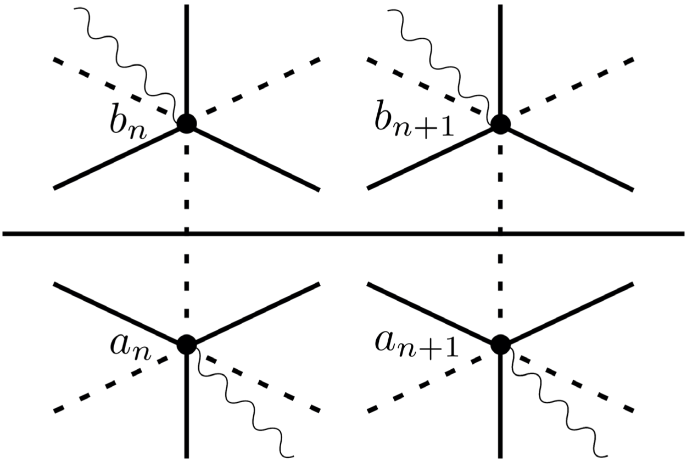


Figure 5.4: Stokes diagram for above-barrier scattering of plasmons over an underdense barrier. Dashed lines are Stokes lines, where $\mathrm{Re}(S) = 0$, and therefore only evanescent waves exist. Solid lines are anti-Stokes lines, where $\mathrm{Im}(S) = 0$, and therefore only propagating waves exist. The turning points are denoted by a_n and b_n, where $\mathrm{Im}(a_n) < 0$ and $\mathrm{Im}(b_n) > 0$. The turning points are exactly above and below the real space projection of the minima in $\varepsilon_\mathrm{b}(x)$. Branch cuts, given by wavy lines, emanate symmetrically from the turning points, meaning the branch cut emanating from turning point a_n is a point reflection of the branch cut emanating from the turning point b_n.

maximum $E_\mathrm{g}(x_2, q_y)$. Intuitively, one would expect that this leads to reduced above-barrier reflection. This intuition is confirmed by the movement of the turning points in the complex plane. When $E_\mathrm{p} = E_\mathrm{g}(x_2, q_y)$, the turning points lie on the real axis at the potential maximum. As q_y decreases and the energy increases relative to the barrier, the imaginary parts of the complex conjugate pair a_n and b_n increase, moving them further away from the real axis. This increases the integral K and, hence, reduces the reflectivity. However, a crucial aspect of this wave phenomenon is that the reflection does not approach zero but rather a finite value in the limit of $q_y \to 0$. This is because the imaginary parts of the turning points themselves approach a finite limiting value instead of moving to infinity. This value can be calculated analytically for specific forms of the periodic modulation, as we will briefly consider in the next section when analyzing two particular modulation profiles.

The form of the reflection and transmission coefficients is similar to the tunneling case. However, the key distinction is that the turning points a_n and b_n are now complex. While K remains real in the above-barrier regime, the presence of complex turning points fundamentally alters the physics. Whereas in the tunneling regime, the transmission coefficient is small, in the above-barrier regime, the reflection coefficient is small. To connect amplitudes across one Bloch period, we must again introduce an additional transfer matrix. This matrix accounts for the change of reference frame, from the n^{th} barrier to the $(n+1)^{th}$ barrier, completing one full Bloch period.

In the above-barrier regime, the concept of single-well bound states is not well defined [97]. This is because the complex turning points, in the analogous single-well problem, would be undefined at infinity. However, we can define forbidden regions (band gaps) where plasmon propagation is suppressed due to above-barrier scattering. The quantization condition for the edges of these forbidden gaps is given by

$$L + \theta = m\pi \pm \tan^{-1} e^{-K/\hbar}, \tag{5.31}$$

where $L = i \int_{\mathrm{Re}(b_n)}^{\mathrm{Re}(a_{n+1})} q_x dx$ is imaginary and is the added action to change reference frames between two barriers. The integration limits are now the real-axis projections of the complex turning points. This equation determines the upper and lower boundaries of the forbidden gaps in the energy spectrum.

In summary, the two regimes, tunneling and above-barrier scattering, provide a comprehensive description of plasmonic crystals. In the tunneling regime, Eq. (5.28) defines the upper and lower bounds for the allowed bands, dictating the energies at which plasmons can propagate through the periodic structure by tunneling. Conversely, in the above-barrier regime, Eq. (5.31) defines the upper and lower bounds for the forbidden bands, indicating the energy ranges where plasmon propagation is prohibited due to destructive interference. In both regimes, the phase factor θ plays a crucial role when the energy approaches the top of the potential barrier, specifically near $E_{\mathrm{g}}(x_2, q_y)$ for a given q_y. This factor accounts for the phase accumulated by the plasmon as it interacts with the turning points, which become closely spaced near the barrier top. The introduction of θ is essential for accurately describing the transition between allowed and forbidden bands, ensuring the continuity of the band structure near $E_{\mathrm{g}}(x_2, q_y)$.

5.5 Numerical computation of explicit periodic modulations

In this section, we apply the theoretical framework developed in the preceding sections to numerically analyze two specific periodic systems. Our goal is to calculate the energy-momentum dispersion relations for these systems, revealing the allowed and forbidden energy bands. We achieve this by numerically solving the quantization conditions derived for the tunneling (Eq. (5.28)) and above-barrier scattering (Eq. (5.31)) regimes. These equations relate the energy and momentum q_y for a given quantum number m.

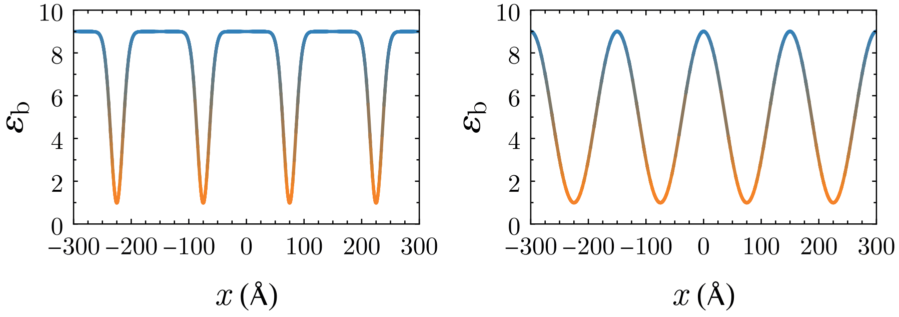
We consider a layered dielectric environment similar to the one described in Ch. 4 (Ref. [69]), consisting of a two-dimensional conducting layer (containing free electrons) embedded within a layered dielectric structure. The effective dielectric constant for such a system is given by

$$\varepsilon_{\text{eff}}(x) = \varepsilon_{\text{M}} \frac{1 - \tilde{\varepsilon}_{\text{b}}^{2}(x) e^{-2|\mathbf{q}|d/\hbar}}{1 + 2\tilde{\varepsilon}_{\text{b}}(x) e^{-|\mathbf{q}|d/\hbar} + \tilde{\varepsilon}_{\text{b}}^{2}(x) e^{-2|\mathbf{q}|d/\hbar}}, \tag{5.32}$$

which is characteristic of layered materials in the literature [34, 40, 69, 70, 100]. The active layer is encapsulated by a dielectric with constant $\varepsilon_{\text{m}} = 10$ and thickness $d = 5.76$ Å. Two semi-infinite layers with background dielectric constant ε_{b} surround this structure. The influence of the background dielectric is captured by $\tilde{\varepsilon}_{\text{b}} = (\varepsilon_{\text{M}} - \varepsilon_{\text{b}})/(\varepsilon_{\text{M}} + \varepsilon_{\text{b}})$. While, in principle, all parameters in $\varepsilon_{\text{eff}}(x)$ can be spatially dependent, provided the semiclassical approximation (small h) remains valid, we restrict our analysis to spatial variations in the background dielectric constant $\varepsilon_{\text{b}}(x)$.

The choice of the specific functional form for $\varepsilon_{\text{b}}(x)$ is subject to certain constraints. Beyond the semiclassical condition, for which the inhomogeneities must be smooth and only simple turning points are considered, there are additional requirements that are not tied to the functional form of $\varepsilon_{\text{b}}(x)$ itself but instead arise from the derivation of the phase shift θ for second-order ODEs. In particular, the method of comparison equations requires that the Weber function equation be applicable [63, 96–98]. While we do not formally derive or prove these conditions, we nonetheless adopt them as working assumptions. In the above-barrier scattering regime, further numerical constraints appear due to the complex nature of the turning points. When the momentum q_y is small (corresponding to smaller relative changes in q_x) or for wider barriers, the turning points are located deep in the complex plane. This can introduce complications because certain functions, such as $\tanh(x)$, possess branch cuts in the complex plane (e.g., at $i\pi/2$), which may interfere with the numerical determination of the turning points.

The choice of the specific functional form for $\varepsilon_{\text{b}}(x)$ is subject to certain constraints. Beyond the semiclassical condition, for which the inhomogeneities, i.e. the dielectric constant, must be smooth and only result in simple turning point. The method of comparison equations, which was applied to second order differential equations to derive the phase shift θ (Eq. (5.21)) requires that the equation for the Weber function must be applicable [63, 96–98]. Once again, while we do not formally derive θ or apply the method of comparison equations, we nonetheless heuristically adopt it. For the above-barrier scattering regime, an additional numerical constraint arises from the complex nature of the turning points. When the momentum q_y is small (corresponding to smaller relative changes in q_x) or for wider barriers, the turning points lie deep in the complex plane. Certain functions, such as $\tanh(x)$, have branch cuts in the complex plane (e.g., at $i\pi/2$). These branch cuts can complicate the numerical determination of the turning points. Ultimately, the applicability of the Weber function description, together with the well-behaved behavior of q_x around the turning points, constitutes the central set of theoretical constraints in this framework.



(a) Gaussian modulation

(b) Cosine modulation

Figure 5.5: (a) Periodic modulation of the background dielectric constant $\varepsilon_b(x)$, using a Gaussian array modulation. The maximum background dielectric constant is given by $\varepsilon_b(x_1) = 9$, whereas the minimum is given by $\varepsilon_b(x_2) = 1$. (b) Periodic modulation of the background dielectric constant $\varepsilon_b(x)$ using a cosine function. The maximum and the minimum background dielectric constant are the same as in (a). The period in both modulations is 150 Å.

As previously explained in Chs. 3 and 4, plasmons perceive a decrease in the dielectric constant as a potential barrier, as this decrease reduces the local plasmon momentum $|\mathbf{q}|$. Therefore, we consider a system where the background dielectric constant $\varepsilon_b(x)$ alternates between $\varepsilon_b(x_1) = 9$ (at x_1, a potential minimum) and $\varepsilon_b(x_2) = 1$ (at x_2, a potential maximum). This can be seen as an active material layer, encapsulated by a dielectric with $\varepsilon_b = 9$, with etched away parts, leaving a vacuum with $\varepsilon_b = 1$. We investigate two specific functional forms for this periodic modulation: a Gaussian array and the cosine function, depicted in Figs. 5.5a and 5.5b. In both cases, the barriers have a characteristic decay length $\ell = 15$ Å and a period $\ell_p = 150$ Å.

The chosen decay length ℓ allows us to verify the validity of the semiclassical approximation through the calculation of the semiclassical parameter. For a typical metal-like system with a carrier density $n^{(0)} = 1.8 \times 10^{14}$ cm^{-2} and an effective mass $m_{\text{eff}} = 0.42 m_e$, the electron wavelength can be estimated using the Thomas-Fermi approximation Eq. (1.49). This yields a semiclassical parameter $h = \lambda_{\text{el}}/\ell = 0.2$, which is within the range of applicability of the semiclassical approximation.

5.5.1 Gaussian Modulation

The Gaussian function is of interest due to its narrow width, controlled by the decay length ℓ. This narrow width leads to closely spaced turning points for a wide range of q_y values, resulting in small values of the integral K and therefore a large transmission coefficient. Consequently, we expect broad allowed bands in the tunneling regime and broad forbidden bands in the above-barrier scattering regime. Such a modulation is relevant for systems with small, localized periodic variations, such as arrays of adatoms or narrow strips etched from the background substrate.

The periodic Gaussian array is given by

$$\varepsilon_\mathrm{b}(x) = \sum_n \left(\varepsilon_\mathrm{b}(x_2) - \varepsilon_\mathrm{b}(x_1)\right) e^{-\left(\frac{x}{\ell} + n\frac{\ell_\mathrm{p}}{\ell}\right)^2} + \varepsilon_\mathrm{b}(x_1), \tag{5.33}$$

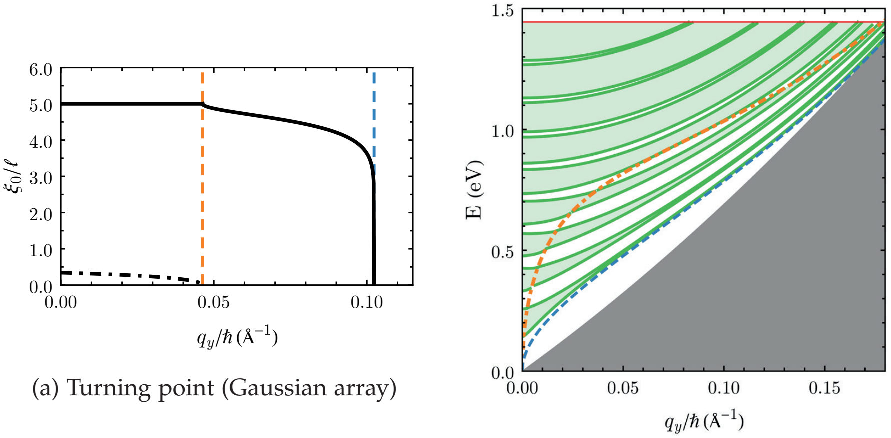
where the sum extends over all integers n, the maximum dielectric constant is $\varepsilon_\mathrm{b}(x_1) = 9$ and the minimum $\varepsilon_\mathrm{b}(x_2) = 1$. Although this expression involves an infinite sum, we only need to consider the transfer matrix for a single barrier to calculate the band structure. Because the period ℓ_p is much larger than the decay length ℓ, the influence of adjacent Gaussian functions on the single barrier is negligible. Even more, this is justified by the fact that we require that there are no extra turning points nearby. The resulting periodic modulation of the background dielectric constant is shown in Fig. 5.5a.

As discussed in the previous chapters and in Refs. [1, 69], the turning points depend on the momentum q_y perpendicular to the barriers. Consequently, the integration limits in L and K, and therefore the quantization conditions, are also functions of q_y. The turning points for the Gaussian barrier are plotted in Fig. 5.6a, for plasmon energy $E_\mathrm{p} = 0.8$ eV. The tunneling and above-barrier scattering regimes are separated by the momentum corresponding to the equality $E_\mathrm{p} = E_\mathrm{g}(x_2, q_y)$; this corresponding q_y value is indicated by the orange dashed line. In the above-barrier regime, the real part of the turning point (solid black line) remains constant at the minimum of the dielectric constant, x_2. In the tunneling regime, the real part of the turning point decreases slowly with increasing q_y and then drops sharply to zero as q_y approaches the momentum corresponding to $E_\mathrm{p} = E_\mathrm{g}(x_1, q_y)$, depicted by the blue dashed line.

The imaginary part of the turning point (black dot-dashed line), applicable to the above-barrier regime, approaches a constant value as $q_y \to 0$. This constant depends on the barrier's shape, specifically the minimum and maximum dielectric constants and the functional form of the modulation. It can be derived analytically by taking the Hamilton-Jacobi equation $\mathcal{H}_0 = 0$ and taking the limit to $q_y \to 0$. For the Gaussian barrier, this constant value is given by $\mathrm{Im}(\xi_0)_{q_y \to 0} = \sqrt{\ln\left(\varepsilon_\mathrm{b}(x_1)/(\varepsilon_\mathrm{b}(x_1) - \varepsilon_\mathrm{b}(x_2))\right)}$.

Applying the theory from the previous subsections, we can calculate the allowed and forbidden bands using Eqs. (5.28) and (5.31), respectively. These bands are shown in Fig. 5.6b, where the thick green lines represent the top and bottom of the allowed bands. As expected, the bands from the two regimes join at the transition energy $E_\mathrm{g}(x_2, q_y)$, where the Stokes diagrams of the two regimes converge, as discussed in Sec. 5.4.1.

As the behavior of the turning point in the tunneling regime in Fig. 5.6a suggests, the allowed bands are broad when the momentum q_y is close to the orange dashed line ($E_\mathrm{g}(x_2, q_y)$). This is because the turning points are near the maximum of the inhomogeneity and thus close together, indicating



(a) Turning point (Gaussian array)

(b) Band structure (Gaussian array)

Figure 5.6: (a) Real (black solid line) and imaginary (black dashed line) parts of the turning points in terms of decay length ℓ as functions of the perpendicular momentum q_y for the Gaussian array modulation. Real turning points exist in the tunneling regime, for large q_y up to the dashed blue line ($E_\text{g}(x_1, q_y)$). For smaller q_y, from the orange dashed line ($E_\text{g}(x_2, q_y)$), the turning points become complex, corresponding to the above-barrier scattering regime. (b) Band structure for the Gaussian periodic modulation. The thick green lines represent the top and bottom of the allowed bands, and the lighter green regions are a continuum of allowed states. The white regions are forbidden bands. The gray region is the Landau damped region, and the gap energies at x_1 and x_2 are the dashed curves in blue and orange, respectively.

a short tunneling length. As the momentum approaches the blue dashed line ($E_\text{g}(x_1, q_y)$), the tunneling length increases rapidly, and the allowed band become narrower. Beside the effect of the momentum q_y, it can be seen in the figure that the allowed bands become narrower with increasing energy.

The allowed bands are generally broader in the above-barrier regime, than in the tunneling regime, reflecting the high transmission probability, and low reflectivity characteristic of this regime. As expected from Fig. 5.6a, the complex turning points move further from the real axis for smaller q_y, which further decreases the width of the forbidden bands. Moreover, similar to the tunneling regime, the forbidden bands become progressively narrower with increasing energy.

Note that the lowest allowed band in the above-barrier regime extends down to zero energy, indicating gapless propagation in the long-wavelength limit. This contrasts sharply with the tunneling regime at higher q_y, where low-energy propagation is forbidden.

In conclusion, the Gaussian periodic barriers, being relatively narrow, facilitate both tunneling and above-barrier scattering. This characteristic is directly reflected in the plasmonic band spectrum depicted in Fig. 5.6b. We observe broad allowed bands in the tunneling regime for small energies and momenta q_y, resulting from a high probability of plasmons tunneling. Similarly, the broad forbidden bands in the above-barrier regime for large q_y, suggesting that plasmons can still encounter significant resistance when propagating over the barriers. These observations underscore the influence of both the energy and momentum q_y on the plasmonic band structure, highlighting their interplay with tunneling and above-barrier scattering phenomena.

5.5.2 Cosine Modulation

In contrast to the sharp peaks of the Gaussian array, the cosine modulation provides a smooth, continuous variation of the dielectric constant, which may be more representative of realistic systems where the substrate material is gradually varied. However, due to the broader nature of the cosine modulation, the turning points move away from each other more rapidly as a function of q_y compared to the Gaussian case. This leads to smaller tunneling coefficients and, hence, narrower bands. Similarly, in the above-barrier regime, the wider potential results in the turning points lying deeper in the complex plane, leading to smaller reflection coefficients and narrower forbidden bands.

We consider a modulation in the background dielectric constant given by

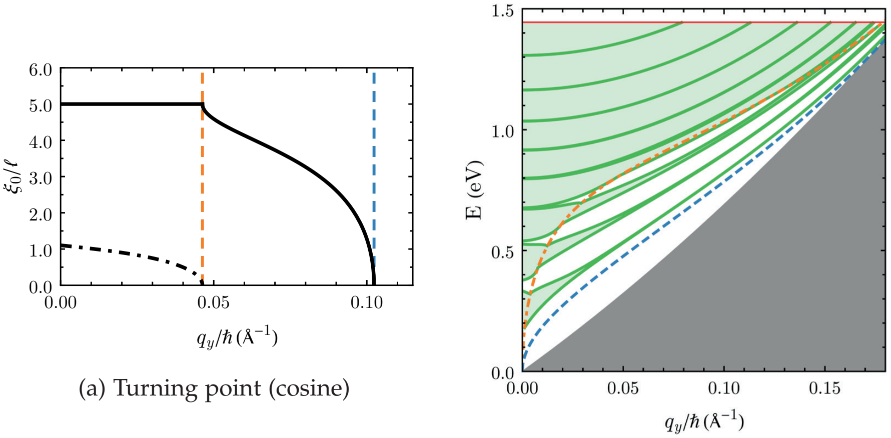
$$\varepsilon_{\mathrm{b}}(x) = \frac{\varepsilon_{\mathrm{b}}(x_1) + \varepsilon_{\mathrm{b}}(x_2)}{2} + \frac{\varepsilon_{\mathrm{b}}(x_1) - \varepsilon_{\mathrm{b}}(x_2)}{2} \cos \frac{2\pi x}{\ell_{\mathrm{p}}}, \tag{5.34}$$

where, again, the maximum dielectric constant is $\varepsilon_{\mathrm{b}}(x_1) = 9$ and the minimum $\varepsilon_{\mathrm{b}}(x_2) = 1$. The resulting periodic modulation of the background dielectric constant is shown in Fig. 5.5b.

As before, the turning points depend on q_y, which influences the integration limits in L and K, and thus the quantization conditions. The turning points for the cosine modulation are plotted in Fig. 5.7a, for a representative plasmon energy $E_{\mathrm{p}} = 0.8$ eV. The transition between the tunneling and above-barrier regimes occurs again at q_y, for which the equality $E_{\mathrm{p}} = E_{\mathrm{g}}(x_2, q_y)$ is met, indicated by the orange dashed line. The real part of the turning point (solid black line) remains constant at x_2 in the above-barrier regime. In the tunneling regime, the real part decreases more sharply with increasing q_y than in the Gaussian case, reaching zero for q_y, for which $E_{\mathrm{p}} = E_{\mathrm{g}}(x_1, q_y)$, marked by the blue dashed line.

As was the case for the Gaussian array in the previous subsection, the imaginary part of the turning point (black dot-dashed line) in the above-barrier regime approaches a constant value as $q_y \to 0$. For the cosine modulation, this limit is given by $\mathrm{Im}(\xi_0)_{q_y \to 0} = \cos^{-1}\left(\varepsilon_{\mathrm{b}}(x_1)/(\varepsilon_{\mathrm{b}}(x_1) - \varepsilon_{\mathrm{b}}(x_2))\right)$. This value is larger for the cosine function than for the Gaussian, suggesting larger values of K and, therefore, weaker reflection.

Applying the same theoretical framework, we calculate the allowed and



(a) Turning point (cosine)

(b) Band structure (cosine)

Figure 5.7: (a) Real (black solid line) and imaginary (black dashed line) parts of the turning points in terms of decay length ℓ as functions of the perpendicular momentum q_y for the cosine modulation. Real turning points exist in the tunneling regime, for large q_y up to the dashed blue line ($E_g(x_1, q_y)$). For smaller q_y, from the orange dashed line ($E_g(x_2, q_y)$), the turning points become complex, corresponding to the above-barrier scattering regime. (b) Band structure for the cosine periodic modulation. The thick green lines represent the top and bottom of the allowed bands, and the lighter green regions are a continuum of allowed states. The white regions are forbidden bands. The gray region is the Landau damped region, and the gap energies at x_1 and x_2 are the dashed curves in blue and orange, respectively.

forbidden bands for the cosine modulation, as shown in Fig. 5.7b. The thick green lines depict the edges of the allowed bands, which merge at the gap energy $E_g(x_2, q_y)$. In the tunneling regime, the allowed bands are narrower than in the Gaussian case, a consequence of the larger values of K and correspondingly smaller transmission coefficients. Similarly, in the above-barrier regime, larger values of K lead to minimal reflection, resulting in very narrow forbidden bands. Due to spectral broadening at finite temperatures, these narrow forbidden bands are unlikely to be observable in experiments.

In conclusion, the cosine periodic barriers, which may be more representative of realistic systems, are generally broader than their Gaussian counterparts. This increased barrier width leads to reduced tunneling and above-barrier scattering probabilities for nearly all energies and momenta q_y. We observe narrow allowed bands in the tunneling regime, resulting from a lower probability of plasmons tunneling through the barriers compared to the Gaussian case. Similarly, the very narrow forbidden bands in the above-barrier regime suggest that plasmons encounter less resistance when propagating over the barriers, yielding a nearly continuous spectrum. These observations highlight the sensitivity of the plasmonic band structure to the shape and width of the periodic barriers, emphasizing the differences between Gaussian and cosine modulations in controlling plasmon propagation.

In summary, this numerical computation has demonstrated the efficacy of our semiclassical theory in computing the band structure for plasmons in periodic potentials. The strong dependence of the band structure on the momentum q_y suggests that experimental observation of forbidden gaps would be more readily achievable in systems with a well-defined q_y. Conversely, excitation via commonly used SNOM techniques [7, 8], which generates radially symmetric plasmons propagating outward from the tip, would likely excite plasmons across a broad range of q_y, potentially obscuring the observation of distinct forbidden gaps [94]. Therefore, for experimental verification of our theoretical predictions, it is crucial to employ excitation methods that allow for more precise control over the plasmon momentum.

5.6 Conclusion

In this chapter, we derived an analytical condition for the allowed and forbidden plasmon bands in a one-dimensional periodic potential within the framework of the semiclassical approximation for plasmons in two-dimensional systems. By heuristically applying complex phase integral methods, we obtained a quantization condition for the plasmon band structure analogous to the semiclassical treatment of systems described by the Schrödinger equation, but depending on the momentum q_y perpendicular to the modulation.

Within this formalism, the plasmon behavior and dynamics are governed by the effective classical Hamiltonian, given by Eq. (2.26). Compared to the wavefunction of the Schrödinger equation, the plasmon wavefunction is represented by the induced potential, as given in Eq. (5.12). Despite the gapless nature of the plasmon dispersion, classically forbidden regions emerge under spatially varying conditions, and the presence of a finite momentum q_y, perpendicular to these spatial variations. This momentum effectively introduces an energy gap in the system, leading to phenomena such as tunneling and above-barrier scattering, similar to Schrödinger-like systems with periodic potentials. The role of this perpendicular momentum is now crucial in determining the turning points, i.e. the start of a classically forbidden region, and the overall plasmon band structure.

The formation of plasmonic bands follows from Bloch's theorem and the eigenvalues of the transfer matrix associated with traversing a single potential barrier. Unlike conventional bound-state problems for plasmons [69], where discrete energy levels arise as a function of q_y, the periodicity of the system results in a continuous band structure. For a system with infinite periodicity, this results in a condition, for the edges of the allowed bands in the tunneling regime given by Eq. (5.28), similar to the quantization condition in Ch. 4, but with an extra term that splits the single well spectra into the edges of continuous bands. For the above-barrier scattering regime, the edges of the forbidden bands are given by Eq. (5.31). The band structure is then a function of the energy, the perpendicular momentum q_y, and quantum number m, labeling the different bands of the plasmon spectrum. An important result is that the band structure heavily depends on the momentum q_y, changing from a forbidden gap to an allowed band for different values of this momentum. Therefore, it might be difficult to see the plasmonic band gap experimentally. Typically these experiments are done by bringing a physical tip close to a sample and subsequent irradiation of light (SNOM) [8, 94, 101], which excites a broad range of q_y values, due to the symmetry.

From a numerical perspective, we analyzed different potential profiles and their effects on the band formation. For a Gaussian array and a cosine potential, the band structure exhibits significant differences based on the width of the modulation. In the tunneling regime, a narrow potential leads to well-separated, broad bands, whereas a wider potential results in narrow bands due to reduced tunneling between adjacent wells. In the above-barrier scattering regime, the broader potential reduces the reflection, which results in very narrow forbidden bands, which are probably not visible due to spectral broadening in experiments. This highlights the strong dependence of the band width on the periodic potential's shape and spatial extent.

For wider barriers, such as a $\tanh(x)$-shaped potential, we observed that the turning points move deeper into the imaginary plane as the barrier width increases. This leads to a decrease in the tunneling coefficient t, effectively making the bands thinner, which would result in something similar to a single-well case. This behavior suggests that wider potentials suppress tunneling effects, reducing both below-barrier and above-barrier scattering contributions to the band formation, conform general results for wide barriers in the WKB approximation. However, in the case of the $\tanh(x)$, the analytic continuation of the potential, particularly the presence of a branch cut at $i\pi/2$, might play a difficult role in shaping these effects.

All information regarding the dielectric environment is captured in the effective dielectric function $\varepsilon_{\text{eff}}(\mathbf{x})$. Although we focused on variations in the dielectric background, our theoretical framework is not limited to these variations, and similar results might be obtained when the framework is applied to other modulation mechanisms, such as local doping of the electron density. It is important to note that a varying electron density also varies the boundary of the electron-hole continuum (Landau damping). In Ref. [50], it is argued that a Landau turning point indicates a hard-wall boundary condition, for which no tunneling is applicable, since the induced potential at such a turning point goes to zero. In other words, the wave just stops.

Finally, we note that experimental realizations of the predicted plasmonic band gaps may be affected by spectral broadening and Landau damping at finite temperatures. These effects could reduce the visibility of forbidden bands in practical settings. However, our approach provides a foundation for further studies on controlling plasmon propagation via spatial periodic modulations, which may help understand plasmonic crystals and plasmonic metamaterials.

6

Conclusion and Outlook

The primary objective of this thesis has been to develop an intuitive, (semi-)analytical theory for quantum plasmons in two-dimensional inhomogeneous media, that can complement numerical research and aid in the interpretation of experiments. A key challenge in plasmonics has been the lack of a general framework capable of accurately describing plasmons in spatially varying environments, where existing approaches often rely on computationally expensive numerical methods or lack analytical insight. To address this, we constructed a semiclassical approach within the random phase approximation (RPA), providing a powerful tool to describe quantum plasmons with intuitive classical techniques. This approach, outlined in Ch. 2, naturally led to an effective classical Hamiltonian, which governs plasmonic behavior in inhomogeneous systems, capturing both plasmon trajectories and the phase of the induced potential that accounts for the wave-like character. Our results are particularly relevant for 2D plasmonic systems, which, unlike their 3D counterparts, exhibit a distinct gapless square-root dispersion relation, enhanced tunability (invasively and noninvasively), and strong interactions with the dielectric environment. By providing a tractable and physically intuitive theoretical framework, this work advances the understanding of plasmon dynamics in realistic, complex environments, paving the way for further analytical and experimental developments in the field.

While the semiclassical theory developed in this thesis provides an analytically tractable and intuitive description of plasmons in 2D inhomogeneous media, it is subject to specific limitations that must be considered when applying it to experimental and numerical settings. The primary condition for its validity is the existence of a small dimensionless parameter h, which determines the accuracy of the expansion. As discussed in Sec. 2.1.5, this parameter depends on the ratio between the electron wavelength and the characteristic length scale of the inhomogeneity. The semiclassical approximation is only valid when the inhomogeneity varies smoothly compared to the electron wavelength, ensuring that local homogeneity can be assumed. Another important limitation arises from screening effects, which introduce an additional dimensionless parameter, κ, that quantifies the ratio between the screening length and the inhomogeneity length scale. If the screening length is comparable to or larger than the characteristic scale of the system, the semiclassical approach formally breaks down.

However, in many practical cases, the semiclassical approximation remains applicable beyond its strict formal limit, often providing accurate qualitative and even quantitative predictions outside its nominal regime of validity. These constraints set a fundamental lower bound on the spatial variations that can be accurately described within this framework, making it less applicable to systems with sharp interfaces, nanoscale defects, or strongly confined plasmonic waveguides.

Additionally, while our approach correctly captures plasmon dispersion and wave-like behavior, it is constrained by the limitations of the RPA [2, 28], which breaks down at low electron densities. This limitation somewhat coincides with the semiclassical constraints, since at low densities, the electron wavelength increases within the Thomas-Fermi approximation, therefore increasing h. Furthermore, the RPA does not account for local interactions and strong correlation effects in strongly interacting electron systems. A theory for plasmons in strongly correlated systems is much more complicated (see, for example, Ref. [102]).

Despite these limitations, the semiclassical approximation applied to the RPA remains a valuable theoretical tool, particularly for bridging analytical models with computational techniques and providing physical intuition in systems where fully numerical methods may lack transparency. For typical experimental systems, we believe that the validity conditions are satisfied. In typical metallic systems, the semiclassical approach remains accurate for spatial variations on the order of 4 nm, while in semiconductors, where electronic wavelengths are larger, the approach extends to length scales around 125 nm. This suggests that for many realistic plasmonic systems, the semiclassical framework provides a reliable and insightful description of plasmon dynamics in inhomogeneous environments.

We applied the general semiclassical framework, developed in this thesis, to physically relevant and experimentally verifiable problems. In Ch. 3, we considered the problem of plasmon scattering on radially symmetric inhomogeneities. Using a simple model for the dielectric environment, introduced in Sec. 2.2.1, we analyzed how local variations in electron density affect the propagation of plasmons. A decrease in electron density was interpreted as an attractive potential, while an increase acted as a repelling potential. More generally, any modification of the system that lowers the plasmon dispersion relation—leading to higher momenta for a given energy—can be understood as an attractive potential.

The scattering problem was formulated in terms of the effective classical Hamiltonian, which provided a clear trajectory-based picture of plasmon dynamics. The wave-like nature of plasmons was incorporated through the induced potential, leading to a phase shift in the scattered wave. By relating this phase shift to the total and differential scattering cross sections, we established a direct connection between the semiclassical theory and experimentally measurable quantities. The semiclassical approach successfully captures key scattering phenomena, including forward scattering, interference effects, and backscattering, which emerge due to variations in the characteristic length scale and peak electron density of the inhomogeneity.

By performing explicit calculations for specific systems, we explored how different inhomogeneity profiles influence the scattering process. We found that interference effects become increasingly pronounced as the system enters the deep semiclassical regime (i.e. smooth spatial variations), leading to oscillatory behavior in the total scattering cross section. These oscillations were linked to overlapping classical trajectories, and we believe them to be a direct visualization of constructive and destructive interference patterns. The numerical results demonstrated that the semiclassical approach provides a physically intuitive and computationally efficient method for understanding plasmon scattering in inhomogeneous 2D systems.

Overall, Ch. 3 illustrated how the developed theory can be applied to experimentally relevant scattering problems. The semiclassical framework not only provides a direct link between classical trajectories and phase shifts but also highlights the essential role of inhomogeneities in shaping plasmon propagation. The insights gained here set the stage for further applications of this theory to more complex plasmonic systems, including plasmon waveguiding with bound states and plasmonic crystals.

Besides studying plasmon scattering, we also applied the developed semiclassical framework to plasmonic waveguides, as discussed in Ch. 4. In this context, we investigated how structured dielectric environments can be used to confine and guide plasmonic excitations. By extending the concept of a repelling potential, we demonstrated that spatial variations in the dielectric constant and electron density can lead to plasmon localization. This allowed us to identify two distinct waveguiding mechanisms.

The first relies on total internal reflection, where a plasmonic bound state is formed due to the presence of a forbidden region, analogous to photonic waveguides. However, such localization is only possible when the plasmon carries a sufficiently high momentum along the waveguide direction. In contrast, for plasmons traveling purely parallel to the variations, no forbidden region emerges, and the momentum simply adjusts to the local dielectric properties.

The second type of localization arises, among other reasons, due to these adjustments of the momentum to the local screening, which modifies the plasmon amplitude without introducing classically forbidden regions. This quasi-localization effect depends on the interplay between the dielectric environment and the plasmon dispersion, providing an alternative mechanism for plasmonic waveguiding.

Through numerical computation, we studied the conditions under which these localized states formed, demonstrating that increasing the dielectric contrast enhanced the degree of localization. By computing the induced electron density, we observed that there are various complex mechanisms in this type of quasi-localization, due to the interplay of different effects from polarization, screening, and momentum changes.

Overall, Ch. 4 illustrated how the semiclassical framework can be applied to describe plasmonic waveguiding in 2D materials. The results established a direct connection between dielectric structuring and plasmon localization, highlighting how variations in the dielectric constant and electron density can be used to engineer effective plasmonic waveguides. These insights provided a foundation for exploring more complex plasmonic structures, such as periodic systems and plasmonic crystals, which were discussed in the subsequent chapter.

In Ch. 5, we derived a theoretical framework to describe plasmonic band structures in periodically modulated systems, extending the semiclassical approach used in previous chapters. We derived an analytical condition for the allowed and forbidden plasmon bands in a quasi-one-dimensional periodic potential. By employing Bloch's theorem, we reduced the problem to determining the eigenvalues of the transfer matrix for traversing a single potential barrier. We obtained this transfer matrix by studying the analytical continuation of our semiclassical expression in the complex plane, which yields transmission and reflection coefficients.

We considered two distinct scenarios: tunneling and above-barrier scattering, depending on the momentum perpendicular to the modulation, q_y. The tunneling occurs through a finite classically forbidden region, as discussed in Sec. 4.2, whereas above-barrier scattering occurs for higher energies. For both scenarios, we derived analytical expressions that define the allowed and forbidden bands. These expressions can be interpreted as quantization conditions, similar to the quantization condition for bound states in Ch. 4, but now with a splitting into continuous bands.

We numerically computed the band structure for two specific periodic modulations: a Gaussian array and the cosine function. For the Gaussian modulation, the plasmonic band spectrum revealed broad allowed bands in the tunneling regime and relatively broad forbidden bands in the above-barrier regime, reflecting the narrowness of the Gaussian barriers, which facilitates both tunneling and above-barrier scattering. In contrast, the cosine modulation, characterized by broader barriers, exhibited narrower allowed bands in the tunneling regime and very narrow forbidden bands in the above-barrier regime, indicating reduced tunneling and above-barrier scattering probabilities.

The results of Ch. 5 demonstrate that our semiclassical theory provides a powerful tool for analyzing plasmonic band structures in periodic potentials, offering valuable insights into the interplay between barrier shape, momentum, and plasmon propagation. To further this analysis, we can consider a 2D crystal with periodic modulation in both in-plane directions, such as done in experiments in for example Ref. [103]. This 2D array could, in principle, be studied based on our theory of plasmon scattering developed in Ch. 3. Such an extension would be similar to the Korringa–Kohn–Rostoker (KKR) theory of band structure for the Schrödinger equation [2]. This goes beyond the scope of this thesis, although it would be an interesting subject for future studies.

Finally, let us briefly review some discussion points mentioned throughout this thesis. First, we did not consider Landau damping or the Landau turning points considered in Ref. [50], which occur when the plasmon energy is higher than the critical Landau energy E_{L} in one region, but plasmons can still exist in another region. This phenomenon was not considered in this thesis, since it seems to be less applicable for 2D plasmons than for 3D plasmons, where the region supporting plasmon propagation becomes increasingly limited as the dispersion slowly approaches and ultimately merges with the Landau-damped region at higher energies (cf. Ref. [28]). Second, finite temperature effects were not explicitly included in our calculations. However, as discussed in Ch. 1, they can be added heuristically or phenomenologically through a broadening factor for the plasmonic states. Lastly, while we considered only isotropic Hamiltonians of the form $\hat{\mathbf{p}}^2$ in this article, which are Hermitian and do not contribute a Berry phase for our semiclassical Ansatz, it would be interesting to investigate different (e.g., matrix) Hamiltonians, as discussed for Dirac systems in Ref. [79], in which this phase could play a significant role, particularly in influencing the quantization conditions through an additional phase term.

6.1 Outlook

The semiclassical framework developed in this thesis provides a comprehensive and versatile description of plasmons in two-dimensional inhomogeneous media. By capturing both the classical trajectories and the wave-like nature of these collective excitations, this approach offers a powerful tool for analyzing plasmon dynamics in a wide range of experimentally relevant systems. The theory is not only applicable to well-established plasmonic phenomena, such as scattering, waveguiding, and periodic structures, but also extends naturally to new experimental platforms where inhomogeneities play a fundamental role. Given its analytical transparency and direct connection to classical physics, this framework serves as a valuable complement to fully numerical methods, providing deeper insight into the underlying mechanisms that govern plasmon behavior.

While the theory is remarkably complete within its domain of validity, there remains room for further development. As mentioned throughout the thesis, our semiclassical approach could be extended to incorporate additional quantum effects beyond the random phase approximation, such as strong correlations, local interactions, and different electronic Hamiltonians with possible Berry phases. Below, we outline several different promising directions that fit naturally within the semiclassical analysis and could be readily pursued.

An exciting avenue for further research involves incorporating classical light, described by Maxwell's equations, into our semiclassical framework. To couple light with plasmons, spatial variations at the boundary are crucial. This is because momentum conservation prevents photons from directly coupling to the electronic system due to the momentum mismatch; light typically has a smaller momentum than plasmons at the same energy. One promising approach to bridge this momentum gap is to introduce a periodic inhomogeneity with a period on the order of the wavelength of the plasmon that one wishes to excite. This inhomogeneity can induce Umklapp scattering [2, 29], providing the necessary momentum boost to the photons, allowing them to couple with plasmons. This can be achieved for both 2D and surface plasmons. To develop this theory, one can consider light in the classical limit, which is reasonable because light generally has a much longer wavelength than plasmons for the same energy [2, 77]. This classical limit can be described by the full Maxwell equations, as given in Sec. 1.2.1. A semiclassical Ansatz would then be needed not only for the electric field, via the induced potential, but also for the magnetic field.

Another interesting extension involves the study of bilayer systems, such as those found in van der Waals heterostructures. In the absence of electronic hybridization between layers, each layer sustains its own plasmonic excitations, leading to two distinct polarizations and separate amplitudes of the induced plasmon potential. Our semiclassical approach can be applied to derive an effective classical Hamiltonian governing such bilayer systems. The analysis predicts a splitting of the plasmon dispersion into a higher-energy branch exhibiting the usual square-root dispersion and a lower-energy branch with a linear dispersion [104]. Investigating the implications of this double dispersion relation on plasmonic bound states and transport properties would be a valuable direction for future research. However, it is important to note that while the effective Hamiltonian can be obtained straightforwardly, solving the full transport equation in this scenario presents additional challenges that require careful consideration. Nevertheless, we would still like to briefly consider the effective classical Hamiltonian. We can write the left-hand side of the Poisson equation (2.1), in the form

$$-4\pi e^2 n(\mathbf{x}, z) = -4\pi e^2 \begin{pmatrix} e^2\Pi_0^{(0)}\delta(z) & 0 \\ 0 & \Pi_0^{(d)}\delta(z-d) \end{pmatrix} \begin{pmatrix} \varphi_0^{(0)} \\ \varphi_0^{(d)} \end{pmatrix} V_{\text{pl}}. \tag{6.1}$$

Following the method of variation of parameters, when then have

$$f(z') = -4\pi e \begin{pmatrix} \Pi_0^{(0)}\delta(z') & 0 \\ 0 & \Pi_0^{(d)}\delta(z'-d) \end{pmatrix}, \tag{6.2}$$

Substitution in Eq. (2.21), where the function $g(z,z')$ Eq. (2.22), now has two contributions, one at $z'=0$, from the bottom polarization, and the other at $z'=d$ from the top polarization. Substituting this in the principal symbol of Γ_0 yields

$$\Gamma_0(z)=\begin{cases}\left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}(0,z)|\mathbf{q}|}\Pi_0^{(0)}\frac{w_1(z)}{w_1(0)} \quad \frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}(d,z)|\mathbf{q}|}\Pi_0^{(d)}\frac{w_1(z)}{w_1(d)}\right), & z>d\\ \left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}(0,z)|\mathbf{q}|}\Pi_0^{(0)}\frac{w_1(z)}{w_1(0)} \quad \frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}(d,z)|\mathbf{q}|}\Pi_0^{(d)}\frac{w_2(z)}{w_2(d)}\right), & d>z>0\,,\\ \left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}(0,z)|\mathbf{q}|}\Pi_0^{(0)}\frac{w_1(z)}{w_2(0)} \quad \frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}(d,z)|\mathbf{q}|}\Pi_0^{(d)}\frac{w_2(z)}{w_2(d)}\right), & 0>z\end{cases} \tag{6.3}$$

where we naturally find different effective dielectric functions, depending on the layer, namely

$$\varepsilon_{\text{eff}}(z_i,z_j)=\frac{\hbar}{2|\mathbf{q}|}\frac{\det W(z_i)}{w_1(z_i)w_2(z_j)}. \tag{6.4}$$

In this bilayer formulation Γ_0 is now a matrix. We use the self-consistency equation, where the amplitude is also a matrix. This yields a coupled set of equations between both amplitudes, which can be solved by diagonalizing the matrix. After eliminating $\varphi_0^{(d)}$, we obtain the effective classical Hamiltonian given by

$$\begin{aligned}\mathcal{H}_0(\mathbf{x},\mathbf{q})=1-\frac{2\pi e^2\hbar}{|\mathbf{q}|}\Pi_0^{(0)}(\mathbf{x},\mathbf{q})& \qquad (6.5)\\ \times\left(\frac{1}{\varepsilon_{\text{eff}}(0,0)}+\frac{2\pi e^2\hbar}{|\mathbf{q}|\varepsilon_{\text{eff}}(d,0)\varepsilon_{\text{eff}}(0,d)}\frac{\Pi_0^{(d)}(\mathbf{x},\mathbf{q})}{1-\frac{2\pi e^2\hbar}{|\mathbf{q}|\varepsilon_{\text{eff}}(d,d)}\Pi_0^{(d)}(\mathbf{x},\mathbf{q})}\right)&.\end{aligned}$$

As stated before, this result has to be treated with caution, since we did not look at the behavior of the amplitude φ_0, which has to be solved from the transport equation.

A related extension concerns surface plasmons at metal-dielectric interfaces, where screening effects from mobile charges play a crucial role. In certain limits, such as the Debye-Hückel regime, analytical solutions for the effective dielectric function can be obtained [70]. Applying our semiclassical theory to such systems could yield insights into plasmon transport in weakly screened environments by layers with roaming charges, drawing connections to previous studies on electrostatic screening at interfaces. The effective dielectric function for such a system would be given by [70]

$$\varepsilon_{\text{eff}}(\mathbf{x})=\varepsilon_{\text{M}}(\mathbf{x})\frac{1-\tilde{\varepsilon}_{\text{A}}(\mathbf{x})\tilde{\varepsilon}_{\text{B}}(\mathbf{x},\mathbf{q})e^{-2|\mathbf{q}|d/\hbar}}{1+\left(\tilde{\varepsilon}_{\text{A}}(\mathbf{x})+\tilde{\varepsilon}_{\text{B}}(\mathbf{x},\mathbf{q})\right)e^{-|\mathbf{q}|d/\hbar}+\tilde{\varepsilon}_{\text{A}}(\mathbf{x})\tilde{\varepsilon}_{\text{B}}(\mathbf{x},\mathbf{q})e^{-2|\mathbf{q}|d/\hbar}}, \tag{6.6}$$

where $\tilde{\varepsilon}_{\text{B}}(\mathbf{x},\mathbf{q})=(\varepsilon_{\text{M}}(\mathbf{x})-\varepsilon_{\text{B}}(\mathbf{x})\sqrt{1+\kappa_{\text{D}}^2/|\mathbf{q}|^2})/(\varepsilon_{\text{M}}(\mathbf{x})+\varepsilon_{\text{B}}(\mathbf{x})\sqrt{1+\kappa_{\text{D}}^2/|\mathbf{q}|^2})$, with κ_{D} the inverse Debye length for the bottom layer, which characterizes the distance over which the electrostatic potential is significantly screened.

The Debye-Hückel theory is valid for moderate charges, which assumes very low-density metals, and is therefore formally not applicable to a metallic interface. A more realistic approach would involve transitioning from a purely two-dimensional model to a three-dimensional picture incorporating full dynamical screening effects from bulk electrons, potentially unifying different plasmonic systems within a single theoretical framework. This might be possible with theory developed in Ref. [50] for plasmons in 3D inhomogeneous systems. In combination with the transition from 3D to 2D, as discussed in Ref. [105, 106], which quantizes and therefore localizes the motion in the third out-of-plane dimension. This would be an interesting avenue to explore, potentially connecting theories for plasmons in various dimensions, all of them from a quantum perspective.

Beyond bilayer and surface plasmon systems, it would be valuable to explore how our semiclassical theory applies to electronic systems with forbidden regions, such as charge-density wave materials or semiconductors and insulators with small band gaps [107–110]. In such systems, the interplay between electronic tunneling and plasmon dynamics could lead to novel transport phenomena, requiring modifications to the standard semiclassical analysis. Investigating these effects would provide deeper insight into the role of quantum coherence and nonlocal interactions in plasmonic systems.

Additionally, another promising direction for further study involves refining the model for the dielectric function in the out-of-plane direction. While the present work considered specific layered forms for the dielectric profile, more sophisticated models, such as exponential, hyperbolic cosine, and hyperbolic tangent functions, could be explored, which can be solved analytically with hypergeometric or Heun functions, depending on the number of poles of the underlying differential equations. These models offer the potential to describe nonlinear screening effects more accurately and may lead to analytically tractable solutions with richer physical interpretations, similar to the analytical solvable model in Ref. [96].

Finally, time-dependent extensions of the semiclassical Hamiltonian can be explored, which would be particularly relevant for experiments involving time-dependent external fields. In the limit of slow oscillations, a semiclassical treatment of time evolution could be employed [56], whereas for high-frequency driving fields, time-averaging techniques could be applied to engineer effective plasmonic band structures [111, 112]. A full treatment combining Maxwell's equations with time-dependent semiclassical plasmon dynamics would be an exciting step towards a more comprehensive description of driven plasmonic systems.

In summary, while the semiclassical framework developed in this thesis already provides a robust and widely applicable tool for understanding plasmons in inhomogeneous systems, there remain many opportunities for further theoretical development. By extending the approach to include light-matter interactions, bilayer systems, surface plasmons, electron tunneling effects, refined dielectric models, and time-dependent fields, the semiclassical theory could be made even more powerful, offering deeper insights into the fundamental physics of plasmonic excitations and their interactions with structured environments

Acknowledgments

The work was supported by the European Union’s Horizon 2020 research and innovation programme under European Research Council synergy grant 854843 “FASTCORR”.

Research data management

This work has been carried out in accordance with the research data management policy of the Institute for Molecules and Materials of Radboud University, the Netherlands. Numerical calculations were performed in Mathematica [90]. All notebooks have been stored on the Radboud Data Repository (RDR), and are publicaly available via the following digital object identifier (DOI): 10.34973/yyc1-6w49.

Appendix

A Alternative derivation of the induced potential $V(\mathbf{x}, z)$

In this appendix, we present an alternative derivation of the effective classical Hamiltonian Eq. (2.26) and the transport equation (2.31), directly applied to the simple model of the dielectric discussed in Sec. 2.2.1. Instead of using the operator separation technique discussed in Sec. 2.1.1, we use the Ansatz (2.4) for $V(\mathbf{x}, z)$ to solve the Poisson equation (2.1) order by order in $\hbar$. On the one hand, we consider this method less elegant, since it mixes the separation of the in-plane and out-of-plane degrees of freedom with the application of the semiclassical Ansatz. On the other hand, it may be easier to understand for readers less familiar with operator techniques.

On the left-hand side of the Poisson equation (2.1), we have

$$\begin{aligned}
&\langle \nabla, \varepsilon(\mathbf{x}, z) \nabla \rangle V(\mathbf{x}, z) \\
&\quad = \Bigg(- \frac{\varepsilon(\mathbf{x}, z)}{\hbar^2} V_0(\mathbf{x}, z) \left(\frac{\partial S}{\partial \mathbf{x}} \right)^2 + \frac{\partial}{\partial z} \varepsilon(\mathbf{x}, z) \frac{\partial V_0}{\partial z} \\
&\qquad - \frac{\varepsilon(\mathbf{x}, z)}{\hbar} V_1(\mathbf{x}, z) \left(\frac{\partial S}{\partial \mathbf{x}} \right)^2 + \hbar \frac{\partial}{\partial z} \varepsilon(\mathbf{x}, z) \frac{\partial V_1}{\partial z} \\
&\qquad + i \frac{2\varepsilon(\mathbf{x}, z)}{\hbar} \left\langle \frac{\partial V_0}{\partial \mathbf{x}}, \frac{\partial S}{\partial \mathbf{x}} \right\rangle + i \frac{\varepsilon(\mathbf{x}, z)}{\hbar} V_0(\mathbf{x}, z) \frac{\partial^2 S}{\partial \mathbf{x}^2} \\
&\qquad + \frac{i}{\hbar} V_0(\mathbf{x}, z) \left\langle \frac{\partial \varepsilon(\mathbf{x}, z)}{\partial \mathbf{x}}, \frac{\partial S}{\partial \mathbf{x}} \right\rangle \Bigg) e^{iS(\mathbf{x})/\hbar},
\end{aligned} \tag{A.1}$$

where we left out all higher-order terms. As discussed in Sec. 2.1.1 and Sec. 2.1.5, the second derivative of V_0 with respect to the out-of-plane direction z belongs to the leading-order term, since the combination $\hbar/z$ is of order one when proper dimensionless parameters are introduced.

The induced electron density $n(\mathbf{x}, z)$ on the right-hand side of the Poisson equation (2.1) is given by Eqs. (2.2) and (1.60). Using our previous Ansatz (1.53) for the induced potential $V_{\text{pl}}(\mathbf{x})$ in the plane, as well as a relation similar to Eq. (2.46), cf. Refs. [48, 55], we obtain the first two terms in the asymptotic expansion of $n(\mathbf{x}, z)$. In the following two subsections, we analyze the terms of leading and subleading order in the Poisson equation, respectively.

A.1 Leading-order term

Collecting the leading-order terms in Eq. (A.1), we have

$$\left(-\frac{1}{\hbar^2}\varepsilon_i(\mathbf{x})\left(\frac{\partial S}{\partial \mathbf{x}}\right)^2+\varepsilon_i(\mathbf{x})\frac{\partial^2}{\partial z^2}\right)V_0(\mathbf{x},z)e^{iS(\mathbf{x})/\hbar}= \\ -4\pi e^2\delta(z)\Pi_0\left(\mathbf{x},\frac{\partial S}{\partial \mathbf{x}}\right)\varphi_0(\mathbf{x})e^{iS(\mathbf{x})/\hbar}, \quad \text{(A.2)}$$

where we made use of the fact that $\varepsilon_i(\mathbf{x})$ is piecewise constant, see Eq. (2.54). The right-hand side contains the leading-order term of the induced electron density $n(\mathbf{x},z)$.

We proceed by noting that the action exponents on both sides cancel out, given that they do not vanish. We can then solve the remaining differential equation in the way discussed in Sec. 2.1.2. We find that

$$V_0(\mathbf{x},z)=c_0\left(\mathbf{x},\frac{\partial S}{\partial \mathbf{x}}\right)e^{-\frac{|z|}{\hbar}\left|\frac{\partial S}{\partial \mathbf{x}}\right|}, \quad \text{(A.3)}$$

with

$$c_0\left(\mathbf{x},\frac{\partial S}{\partial \mathbf{x}}\right)=\frac{2\pi e^2\hbar}{\varepsilon_{\mathrm{avg}}(\mathbf{x})|\partial S/\partial\mathbf{x}|}\Pi_0\left(\mathbf{x},\frac{\partial S}{\partial \mathbf{x}}\right)\varphi_0(\mathbf{x}), \quad \text{(A.4)}$$

where we used the definition $2\varepsilon_{\mathrm{avg}}(\mathbf{x})\equiv\varepsilon_{\mathrm{A}}(\mathbf{x})+\varepsilon_{\mathrm{B}}(\mathbf{x})$. Expression (A.3) corresponds to the leading-order term of $\left(\hat{\Gamma}V_{\mathrm{pl}}\right)(\mathbf{x},z)$ in the main text, see Eqs. (2.58) and (2.46).

At this point, we can directly apply the self-consistency condition (2.3) at $z=0$ to the leading-order terms. We have

$$V_0(\mathbf{x},z=0)e^{iS(\mathbf{x})/\hbar}=\varphi_0(\mathbf{x})e^{iS(\mathbf{x})/\hbar}, \quad \text{(A.5)}$$

from which we obtain the Hamilton-Jacobi equation

$$\mathcal{H}_0\left(\mathbf{x},\frac{\partial S}{\partial \mathbf{x}}\right)=1-\frac{2\pi e^2\hbar}{\varepsilon_{\mathrm{avg}}(\mathbf{x})|\partial S/\partial\mathbf{x}|}\Pi_0\left(\mathbf{x},\frac{\partial S}{\partial \mathbf{x}}\right)=0. \quad \text{(A.6)}$$

This is the same result as the Hamilton Jacobi equation following from the effective classical Hamiltonian Eq. (2.26) in the main text, taking into account Eq. (2.58). From a technical point of view, one may say that this derivation of the leading-order term is not that different from our previous derivation in Sec. 2.1. The main difference is that the derivatives with respect to **x** directly act on the semiclassical Ansatz, leading to the replacement of **q** by $\partial S/\partial\mathbf{x}$.

A.2 Subleading-order term

The derivation of the subleading-order term $V_1(x,z)$ is however quite different from our derivation in the main text. We first calculate $n_1(\mathbf{x})$, the subleading part of $(\hat{\Pi}V_{\mathrm{pl}})(\mathbf{x})$, using a relation similar to Eq. (2.46), cf. Refs. [48, 55]. In this way, we find

$$\begin{aligned} n_1(\mathbf{x}) =& n_{1,\mathrm{a}}(\mathbf{x})e^{iS(\mathbf{x})/\hbar} \\ n_{1,\mathrm{a}}(\mathbf{x}) =& \Pi_0\left(\mathbf{x},\frac{\partial S}{\partial \mathbf{x}}\right)\varphi_1(\mathbf{x}) + \Pi_1\left(\mathbf{x},\frac{\partial S}{\partial \mathbf{x}}\right)\varphi_0(\mathbf{x}) \\ & -\frac{i}{2}\sum_{jk}\frac{\partial^2\Pi_0}{\partial q_j \partial q_k}\left(\mathbf{x},\frac{\partial S}{\partial \mathbf{x}}\right)\frac{\partial^2 S}{\partial x_j \partial x_k}\varphi_0(\mathbf{x}) \\ & - i\left\langle \frac{\partial \Pi_0}{\partial \mathbf{q}}\left(\mathbf{x},\frac{\partial S}{\partial \mathbf{x}}\right), \frac{\partial \varphi_0}{\partial \mathbf{x}}\right\rangle, \end{aligned} \tag{A.7}$$

where $\Pi_1(\mathbf{x},\partial S/\partial \mathbf{x})$ satisfies Eq. (1.57).

Next, we consider the terms of subleading order in Eq. (A.1). Inserting our solution (A.3) for $V_0(\mathbf{x},z)$, computing the various derivatives, and using that $\varepsilon(\mathbf{x},z)$ is piecewise constant, we find

$$\begin{aligned} &\left(-\frac{1}{\hbar^2}\varepsilon_i(\mathbf{x})\left(\frac{\partial S}{\partial \mathbf{x}}\right)^2 + \varepsilon_i(\mathbf{x})\frac{\partial^2}{\partial z^2}\right)V_1(\mathbf{x},z) \\ &\qquad = -\frac{i}{\hbar^2}\left(f_{1,i}(\mathbf{x}) + f_{2,i}(\mathbf{x})\frac{|z|}{\hbar}\right)e^{-\frac{|z|}{\hbar}\left|\frac{\partial S}{\partial \mathbf{x}}\right|} - 4\pi e^2\delta(z)n_{1,\mathrm{a}}(\mathbf{x}), \end{aligned} \tag{A.8}$$

where we canceled the exponent $\exp(iS(\mathbf{x})/\hbar)$ on both sides. Moreover, we defined $f_{1,i}(\mathbf{x})$ and $f_{2,i}(\mathbf{x})$ by

$$f_{1,i} = \left\langle \frac{\partial \varepsilon_i}{\partial \mathbf{x}}, \frac{\partial S}{\partial \mathbf{x}}\right\rangle c_0 + 2\varepsilon_i\left\langle \frac{\partial c_0}{\partial \mathbf{x}}, \frac{\partial S}{\partial \mathbf{x}}\right\rangle + \varepsilon_i c_0 \frac{\partial^2 S}{\partial \mathbf{x}^2}, \tag{A.9}$$

$$f_{2,i} = -\frac{2\varepsilon_i c_0}{|\partial S/\partial \mathbf{x}|}\sum_{j,k}\frac{\partial S}{\partial x_j}\frac{\partial S}{\partial x_k}\frac{\partial^2 S}{\partial x_j \partial x_k}, \tag{A.10}$$

where c_0 was defined in Eq. (A.4) and the subscript $i = A$ $(i = B)$ denotes the region above (below) the plane $z = 0$, just as in Sec. 2.1.1.

We can solve the differential equation for V_1 using the methods discussed in Sec. 2.1.1. We first determine the solution of the homogeneous equation, given by

$$V_{1,i,\mathrm{H}}(\mathbf{x},z) = c_{1,i}^{-}e^{-\frac{|z|}{\hbar}\left|\frac{\partial S}{\partial \mathbf{x}}\right|} + c_{1,i}^{+}e^{\frac{|z|}{\hbar}\left|\frac{\partial S}{\partial \mathbf{x}}\right|}. \tag{A.11}$$

Second, we solve the inhomogeneous equation using the method of undetermined coefficients. To this end, we employ the Ansatz

$$V_{1,i,\mathrm{P}}(\mathbf{x}, z) = \left(\alpha_i \frac{z}{\hbar} + \beta_i \frac{z^2}{\hbar^2}\right) e^{-\frac{|z|}{\hbar}\left|\frac{\partial S}{\partial \mathbf{x}}\right|}. \tag{A.12}$$

Inserting this Ansatz into the differential Eq. (A.8), we obtain the constants α_i and β_i. After some calculus, we find that the constants β_i are the same above and below the plane, and given by

$$\beta = -\frac{i}{2} \frac{2\pi e^2 \hbar}{\varepsilon_{\mathrm{avg}}(\mathbf{x})} \frac{\Pi_0\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}\right)}{|\partial S/\partial \mathbf{x}|^3} \varphi_0(\mathbf{x}) \sum_{j,k} \frac{\partial S}{\partial x_j} \frac{\partial^2 S}{\partial x_j \partial x_k} \frac{\partial S}{\partial x_k}, \tag{A.13}$$

where we dropped the subscript. Although the constant β vanished for the operator separation, it is non-zero in this derivation. We come back to this shortly. Using our result for β, we find α_i as

$$\begin{aligned} \alpha_i = \frac{i}{2} \frac{s_i}{|\partial S/\partial \mathbf{x}|} \Bigg(& -\frac{c_0}{|\partial S/\partial \mathbf{x}|^2} \sum_{j,k} \frac{\partial S}{\partial x_j} \frac{\partial S}{\partial x_k} \frac{\partial^2 S}{\partial x_j \partial x_k} \\ & + \frac{c_0}{\varepsilon_i} \left\langle \frac{\partial \varepsilon_i}{\partial x}, \frac{\partial S}{\partial x} \right\rangle + 2 \left\langle \frac{\partial c_0}{\partial x}, \frac{\partial S}{\partial x} \right\rangle + c_0 \frac{\partial^2 S}{\partial \mathbf{x}^2} \Bigg), \end{aligned} \tag{A.14}$$

where $s_i = 1$ for $i = \mathrm{A}$ and $s_i = -1$ for $i = \mathrm{B}$, as in the main text. The full solution is given by the sum of the homogeneous (A.11), and particular (A.12) solution.

As before, the constants $c_{1,i}^-$, and $c_{1,i}^+$ are determined via the boundary conditions. First, the induced potential tends to zero as $|z| \to \infty$, which yields $c_{1,i}^+ = 0$. Second, the potential has to be continuous at the interface $z = 0$, whence $c_{1,\mathrm{A}}^- = c_{1,\mathrm{B}}^- = c_1$. The final boundary condition concerns the derivative of V at the interface. It can either be derived from the differential equation (A.8) or from the requirement that the $\mathcal{D}$-field is discontinuous at the interface, with the discontinuity given by the surface charge determined by the induced electron density. We find that the constant c_1 is given by

$$c_1 = \frac{2\pi e^2 \hbar}{\varepsilon_{\mathrm{avg}}(\mathbf{x})} \frac{n_1(\mathbf{x})}{|\partial S/\partial \mathbf{x}|} + \frac{1}{|\partial S/\partial \mathbf{x}|} \frac{\varepsilon_{\mathrm{A}}(\mathbf{x})\alpha_{\mathrm{A}} - \varepsilon_{\mathrm{B}}(\mathbf{x})\alpha_{\mathrm{B}}}{2\varepsilon_{\mathrm{avg}}(\mathbf{x})}. \tag{A.15}$$

The total solution for $V_{1,i}$ is then given by

$$V_{1,i}(\mathbf{x}, z) = \left(c_1 + \alpha_i \frac{z}{\hbar} + \beta \frac{z^2}{\hbar^2}\right) e^{-\frac{|z|}{\hbar}\left|\frac{\partial S}{\partial \mathbf{x}}\right|}. \tag{A.16}$$

We may briefly compare this result with the expression for $\Gamma_{1,i}$ found in Ref. [1] (Eq. (49)). The most striking difference between the two is the presence of a term with z^2 in Eq. (A.16). In the main text, this term only arises when one computes the action of the pseudodifferential operator $\hat{\Gamma}$ on the semiclassical

Ansatz for V_{pl}, as considered in Eq. (2.30). Specifically, it arises when both derivatives in $\partial^2\Gamma_0/\partial q_j \partial q_k$ are applied to the exponent. In both derivations, one obtains the same result for $V_{1,i}$.

A.3 Transport equation

Our next step is to apply the self-consistency condition (2.3) at $z = 0$ to the subleading-order terms. In this way, we should obtain the transport equation. The condition reads

$$V_{1,i}(\mathbf{x}, z = 0)e^{iS(\mathbf{x})/\hbar} = \varphi_1(\mathbf{x})e^{iS(\mathbf{x})/\hbar}, \tag{A.17}$$

which directly yields $\varphi_1 - c_1 = 0$. At this point, one has to insert the expressions (A.14) for α_i into Eq. (A.15) and compute all derivatives. One then computes the derivatives $\partial\mathcal{H}_0/\partial q_j$ and $\partial^2\mathcal{H}_0/\partial q_j \partial q_k$, and uses them to rewrite the terms in c_0. After somewhat lengthy calculations, the equation $\varphi_1 - c_1 = 0$ becomes

$$\begin{aligned}\left(1 - \frac{2\pi e^2\hbar}{\varepsilon_{\text{avg}}\,|\partial S/\partial \mathbf{x}|}\Pi_0\right)\varphi_1 + \Bigg(&\frac{-2\pi i e^2\hbar}{\varepsilon_{\text{avg}}|\partial S/\partial \mathbf{x}|^3}\left\langle \frac{\partial S}{\partial \mathbf{x}}, \frac{\partial \Pi_0}{\partial \mathbf{x}}\right\rangle \\ &+ \frac{i}{2}\frac{2\pi e^2\hbar}{\varepsilon_{\text{avg}}^2|\partial S/\partial \mathbf{x}|^3}\Pi_0\left\langle \frac{\partial S}{\partial \mathbf{x}}, \frac{\partial \varepsilon_{\text{avg}}}{\partial \mathbf{x}}\right\rangle \\ &- \frac{2\pi e^2\hbar}{\varepsilon_{\text{avg}}|\partial S/\partial \mathbf{x}|}\Pi_1\Bigg)\varphi_0 - i\left\langle \frac{\partial \mathcal{H}_0}{\partial \mathbf{q}}, \frac{\partial \varphi_0}{\partial \mathbf{x}}\right\rangle \\ &- \frac{i}{2}\sum_{j,k}\frac{\partial^2\mathcal{H}_0}{\partial q_j \partial q_k}\frac{\partial^2 S}{\partial x_k \partial x_j}\varphi_0 = 0.\end{aligned} \tag{A.18}$$

First, the terms in front of φ_1 cancel because of the Hamilton-Jacobi equation (A.6). Using the definition of $\mathcal{H}_1(\mathbf{x}, \mathbf{q}) \equiv -c_1$, we may write the remaining terms as

$$\begin{aligned}\mathcal{H}_1\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}\right)\varphi_0 - i\left\langle \frac{\partial \mathcal{H}_0}{\partial \mathbf{q}}\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}\right), \frac{\partial \varphi_0}{\partial \mathbf{x}}\right\rangle& \\ - \frac{i}{2}\sum_{j,k}\frac{\partial^2\mathcal{H}_0}{\partial q_j \partial q_k}\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}\right)&\frac{\partial^2 S}{\partial x_k \partial x_j}\varphi_0 = 0,\end{aligned} \tag{A.19}$$

which exactly coincides with our previous transport equation (2.31).

We have thus shown that the Ansatz (2.4) for $V(\mathbf{x}, z)$ leads to the same results as the operator separation discussed in Sec. 2.1.1, namely the effective classical Hamiltonian (2.26) and the Hamilton-Jacobi equation and the transport equation (2.31). However, the calculations are more tedious since this method mixes the separation of the in-plane and out-of-plane degrees of freedom with the application of the semiclassical Ansatz.

B Energy density in the plane $z = 0$

In Sec. 2.1.4, we computed the integrated energy density $\mathcal{U}_{\mathrm{I}}(\mathbf{x})$ for our two-dimensional problem. In the derivation, we set the contribution $\mathcal{U}_{\mathrm{pl}}(\mathbf{x})$ of the plane $z = 0$ to zero. In this appendix, we derive a general formula for the leading-order term of the semiclassical energy density in a medium with plasmons. Moreover, we show that $\mathcal{U}_{\mathrm{pl}}(\mathbf{x}) = 0$ in our example, that is, when the layer is infinitely thin.

As in the main text, the relation between the electric field and the induced potential is given by $\boldsymbol{\mathcal{E}} = e^{-1}\nabla V$. However, the relation between the displacement field and the electric field is more complicated than in the main text. We start from the general relation between the electric field and the displacement field, namely

$$\boldsymbol{\mathcal{D}}(\mathbf{x}, t) = \int dt' \int d\mathbf{x}' \varepsilon(\mathbf{x}, \mathbf{x}', t - t') \boldsymbol{\mathcal{E}}(\mathbf{x}', t'). \tag{B.1}$$

We perform a Fourier transform with respect to time, i.e.,

$$\boldsymbol{\mathcal{D}}(\mathbf{x}, \omega) = \int d\mathbf{x}' \varepsilon(\mathbf{x}, \mathbf{x}', \omega) \boldsymbol{\mathcal{E}}(\mathbf{x}', \omega). \tag{B.2}$$

Throughout this appendix, we use ω instead of E, not only to make a more explicit connection with the conventions in the literature [25, 77], but also to create a clear distinction in notation with the electric field. We can rewrite Eq. (B.2) as

$$\boldsymbol{\mathcal{D}}(\mathbf{x}, \omega) = (\hat{\varepsilon}\boldsymbol{\mathcal{E}})(\mathbf{x}, \omega), \tag{B.3}$$

where $\hat{\varepsilon}$ is an operator that corresponds to the dielectric function. We now need to extract the proper definition of the operator $\hat{\varepsilon}$ from our semiclassical analysis.

Since plasmons are self-sustained oscillations, they are defined by the vanishing of the displacement field. For homogeneous systems [2, 28], this requirement translates to the secular equation $\varepsilon(\mathbf{q}, \omega)V(\mathbf{q}) = 0$, generally speaking. When we consider inhomogeneous systems and apply the semiclassical approximation [50], the secular equation becomes the operator equation $\hat{\mathcal{H}}V = 0$, where $\hat{\mathcal{H}}$ is a pseudodifferential operator. Indeed, we can write the result of our derivation in Sec. 2.1 as $\hat{\mathcal{H}}V_{\mathrm{pl}} = 0$, where $V_{\mathrm{pl}}(\mathbf{x}, \omega) = \varphi_0(\mathbf{x}) \exp(iS(\mathbf{x})/\hbar)$ and the principal symbol $\mathcal{H}_0(\mathbf{x}, \mathbf{q}, \omega)$ of $\hat{\mathcal{H}}$ is the effective classical Hamiltonian given by Eq. (2.26). Comparing this result to Eq. (B.3), we may say that $\hat{\varepsilon}$ is a pseudodifferential operator. It is, however, not equal to $\hat{\mathcal{H}}$, since its principal symbol should equal $\varepsilon_{\mathrm{avg}}$ in the absence of a polarization. We therefore argue that the principal symbol $\varepsilon_0(\mathbf{x}, \mathbf{q}, \omega)$ of $\hat{\varepsilon}$ equals

$$\varepsilon_0(\mathbf{x}, \mathbf{q}, \omega) = \varepsilon_{\mathrm{avg}}(\mathbf{x}) - \frac{2\pi e^2 \hbar}{|\mathbf{q}|} \Pi_0(\mathbf{x}, \mathbf{q}, \omega), \tag{B.4}$$

which gives the correct result in the homogeneous case.

Equations (B.3) and (B.4) allow us to derive the leading-order term of the displacement field $\boldsymbol{\mathcal{D}}(\mathbf{x},\omega)$. Since $V(\mathbf{x},\omega)$ has the form of a semiclassical Ansatz, so does $\boldsymbol{\mathcal{E}}(\mathbf{x},\omega)$. We can therefore use the general formula for the commutation of a pseudo-differential operator with a rapidly oscillating exponent [48, 55], and write

$$\boldsymbol{\mathcal{D}}(\mathbf{x},\omega) = (\hat{\varepsilon}\boldsymbol{\mathcal{E}})(\mathbf{x},\omega) = \varepsilon_0\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}, \omega\right)\boldsymbol{\mathcal{E}}(\mathbf{x},\omega)(1+\mathcal{O}(\hbar)), \tag{B.5}$$

cf. Eq. (2.30). From here on, we only consider the leading-order term, and therefore omit the $\mathcal{O}(\hbar)$.

We can now obtain an expression for the energy density by repeating the derivation in Ref. [77], almost verbatim. Since $\varepsilon_0\left(\mathbf{x}, \partial S/\partial \mathbf{x}, \omega\right)$ depends on ω, the medium is dispersive and one cannot consider a purely monochromatic field. Instead, we let $\boldsymbol{\mathcal{E}}(\mathbf{x},t) = \boldsymbol{\mathcal{E}}_c(\mathbf{x},t)\exp(-i\omega_c t)$, where $\boldsymbol{\mathcal{E}}_c(\mathbf{x},t)$ varies only slowly with time. Hence, when we write down the Fourier expansion of $\boldsymbol{\mathcal{E}}(\mathbf{x},t)$, namely,

$$\boldsymbol{\mathcal{E}}(\mathbf{x},t) = \int \frac{d\omega}{2\pi}\boldsymbol{\mathcal{E}}_c(\mathbf{x},\omega)e^{-i(\omega_c+\omega)t}, \tag{B.6}$$

only the components $\boldsymbol{\mathcal{E}}_c(\mathbf{x},\omega)$ with $\omega \ll \omega_c$ are significant. Expanding the displacement field $\boldsymbol{\mathcal{D}}(\mathbf{x},t)$ in Fourier components up to leading order, and using Eq. (B.5), we find that

$$\frac{\partial \boldsymbol{\mathcal{D}}}{\partial t} = \int \frac{d\omega}{2\pi} f(\mathbf{x},\omega_c+\omega)\boldsymbol{\mathcal{E}}_c(\mathbf{x},\omega)e^{-i(\omega_c+\omega)t}, \tag{B.7}$$

where $f(\mathbf{x},\omega) = -i\omega\varepsilon_0\left(\mathbf{x}, \partial S/\partial \mathbf{x}, \omega\right)$, cf. Ref. [77]. Since the Fourier components $\boldsymbol{\mathcal{E}}_c(\mathbf{x},\omega)$ are very small for $\omega \gg \omega_c$, we can expand $f(\mathbf{x},\omega)$ to first order in ω around ω_c, that is,

$$\frac{\partial \boldsymbol{\mathcal{D}}}{\partial t} = \int \frac{d\omega}{2\pi}\left(f(\mathbf{x},\omega_c) + \omega\frac{\partial f}{\partial \omega}(\mathbf{x},\omega_c)\right)\boldsymbol{\mathcal{E}}_c(\mathbf{x},\omega)e^{-i(\omega_c+\omega)t}. \tag{B.8}$$

Using the definition of the Fourier transform, the right-hand side can be rewritten as

$$\begin{aligned}\frac{\partial \boldsymbol{\mathcal{D}}}{\partial t} = -i\omega\varepsilon_0\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}, \omega\right)&\boldsymbol{\mathcal{E}}_c(\mathbf{x},t)e^{-i\omega t} \\ &+ \frac{\partial}{\partial \omega}\left(\omega\varepsilon_0\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}, \omega\right)\right)\frac{\partial \boldsymbol{\mathcal{E}}_c}{\partial t}(\mathbf{x},t)e^{-i\omega t},\end{aligned} \tag{B.9}$$

where we now omitted the subscript c on ω.

We now substitute the result (B.9) in the Poynting theorem (2.47). Assuming that ε_0 is a real function, we find the leading-order term of the semiclassical energy density by taking out the time derivative. We obtain

$$\mathcal{U}^{\text{SC}}(\mathbf{x}) = \frac{1}{16\pi}\frac{\partial}{\partial\omega}\left(\omega\varepsilon_0\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}, \omega\right)\right) \boldsymbol{\mathcal{E}}(\mathbf{x}, t) \cdot \boldsymbol{\mathcal{E}}^*(\mathbf{x}, t). \tag{B.10}$$

Note that this expression reduces to the much simpler result (2.48) when we consider a function $\varepsilon(\mathbf{x})$ that does not depend on ω. When we consider plasmons, we can simplify this expression by noting that $\varepsilon_0\left(\mathbf{x}, \partial S/\partial\mathbf{x}, \omega\right)$ vanishes by virtue of the Hamilton-Jacobi equation. We obtain

$$\mathcal{U}^{\text{SC}}(\mathbf{x}) = \frac{\omega}{16\pi e^2}\frac{\partial\varepsilon_0}{\partial\omega}\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}, \omega\right)|\nabla V(\mathbf{x}, t)|^2. \tag{B.11}$$

where we also used the relation between the electric field and the induced potential.

In deriving the result (B.11), we did not specify the number of dimensions, which means that it is equally valid in three and two dimensions. When we consider our two-dimensional problem, both $\boldsymbol{\mathcal{D}}$ and $\boldsymbol{\mathcal{E}}$ gain an additional coordinate z, and $\mathbf{x}$ becomes two-dimensional. Throughout this article, we considered an infinitely thin charge layer at $z = 0$. However, one may intuitively argue that we can use expression (B.11) between $z = -\epsilon$ and $z = \epsilon$, where ϵ is a small number that is determined by the requirement that the induced potential $V(\mathbf{x}, z, t)$ has not yet decayed significantly. Looking at Eq. (2.46), we observe that this is equivalent to $\epsilon\left|\partial S/\partial\mathbf{x}\right|/\hbar \ll 1$, meaning that ϵ should be much smaller than the plasmon wavelength, in accordance with the derivation in Sec. 2.1.1.

We obtain an expression for $\mathcal{U}_{\text{P}}(\mathbf{x})$, the energy density that comes from the two-dimensional plane at $z = 0$, by integrating Eq. (B.11) from $z = -\epsilon$ to $z = \epsilon$. Using Eq. (2.46) for $V(\mathbf{x}, z, t)$, we have

$$\mathcal{U}_{\text{pl}}(\mathbf{x}) = \frac{1}{8\pi e^2\hbar}\frac{|A_0^0|^2}{|J|}\frac{\omega}{\varepsilon_{\text{avg}}}\frac{\partial\varepsilon_0}{\partial\omega}\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}, \omega\right)\left(1 - e^{-2\frac{\epsilon}{\hbar}\left|\frac{\partial S}{\partial \mathbf{x}}\right|}\right). \tag{B.12}$$

Since we previously required that $\epsilon\left|\partial S/\partial\mathbf{x}\right|/\hbar \ll 1$, we can Taylor expand the exponent up to first order, which yields

$$\mathcal{U}_{\text{pl}}(\mathbf{x}) = \frac{1}{8\pi e^2\hbar}\frac{|A_0^0|^2}{|J|}\frac{\omega}{\varepsilon_{\text{avg}}}\frac{\partial\varepsilon_0}{\partial\omega}\left(\mathbf{x}, \frac{\partial S}{\partial \mathbf{x}}, \omega\right) \times 2\frac{\epsilon}{\hbar}\left|\frac{\partial S}{\partial \mathbf{x}}\right|. \tag{B.13}$$

Comparing this to the result (2.50), we observe that it is of higher order. Because we set out to derive the leading-order term of the integrated energy density $\mathcal{U}_{\text{I}}(\mathbf{x})$, we conclude that the contribution of the two-dimensional plane is effectively zero, i.e. $\mathcal{U}_{\text{pl}}(\mathbf{x}) = 0$. The derivation shows that the same result likely holds for layers with a finite thickness, cf. the discussion in the introduction and the conclusion, provided that $\epsilon\left|\partial S/\partial\mathbf{x}\right|/\hbar \ll 1$.

C Additional derivations for the general description

C.1 Derivation of subprincipal symbol Γ_1 at $z = 0$

In this appendix, we compute the integral (2.41) explicitly at the point $z = 0$, using integration by parts. We show that the final result can be cast in the form (2.42).

Starting from Eqs. (2.41) and (2.39), we have

$$\Gamma_1(z=0) = -4\pi e^2 \Pi_1 g(0,0) + \int_{-\infty}^{\infty} g(0,z') f_{1s}(z') dz', \tag{C.1}$$

where $g(z,z')$ is given by Eq. (2.22) and f_{1s} by Eq. (2.40). The latter quantity is rather involved and contains F_0, F_1 and Γ_0. We first note that F_1 does not contain any differential operators in z, whereas F_0 does, see Eqs. (2.10) and (2.11). To shorten the notation, we write

$$F_0\left(\mathbf{x}, \mathbf{q}, z, \frac{\partial}{\partial z}\right) = \hat{F}_0, \tag{C.2}$$

and omit the arguments $(\mathbf{x}, \mathbf{q})$ throughout most of this appendix.

Given the structure of $g(0,z')$, see Eq. (2.22), we can split the integration in Eq. (C.1) into two parts, over the intervals $(-\infty, 0]$ and $[0, \infty)$. We start with the computation of the integral over the interval $[0,\infty)$. From Eqs. (C.1), (2.22) and (2.40), and applying the product rule to the second term in f_{1s}, we have

$$\begin{aligned}\int_0^{\infty} g(0,z') f_{1s}(z') dz' = \frac{w_2(0)}{\det(W)} \frac{2\pi e^2 \hbar}{\varepsilon_{\mathrm{eff}}|\mathbf{q}|} \frac{\Pi_0}{w_1(0)} &\left(\int_0^{\infty} w_1(z') F_1(z') w_1(z') dz' \right. \\ &\left. - i \int_0^{\infty} w_1(z') \sum_j \frac{\partial \hat{F}_0}{\partial q_j} \frac{\partial w_1}{\partial x_j}(z') dz' \right) \\ &- i \frac{w_2(0)}{\det(W)} \sum_j \frac{\partial}{\partial x_j} \left(\frac{2\pi e^2 \hbar}{\varepsilon_{\mathrm{eff}}|\mathbf{q}|} \frac{\Pi_0}{w_1(0)} \right) \\ &\times \int_0^{\infty} w_1(z') \frac{\partial \hat{F}_0}{\partial q_j} w_1(z') dz'. \end{aligned} \tag{C.3}$$

We proceed by removing F_1 from the above expression. Since w_i, where $i = \{1,2\}$, is a solution of the homogeneous differential equation (2.16), we have $\hat{F}_0 w_i = 0$. Taking a mixed partial derivative of this relation with respect to x_j and q_j, we have

$$\frac{i}{2}\left(\sum_j \hat{F}_0 \frac{\partial^2 w_i}{\partial x_j \partial q_j} + \sum_j \frac{\partial \hat{F}_0}{\partial x_j} \frac{\partial w_i}{\partial q_j} + \sum_j \frac{\partial \hat{F}_0}{\partial q_j} \frac{\partial w_i}{\partial x_j} \right) = -\frac{i}{2} \sum_j \frac{\partial^2 \hat{F}_0}{\partial x_j \partial q_j} w_i = F_1 w_i, \tag{C.4}$$

where the last equality holds by virtue of the last equality in Eq. (2.11). With this last relation, we can remove F_1 from Eq. (C.3), and obtain

$$\begin{aligned}\int_0^\infty g(0,z')f_{1s}(z')dz' = \frac{i}{2}\frac{w_2(0)}{\det(W)}\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\frac{\Pi_0}{w_1(0)}\sum_j\int_0^\infty w_1(z')\Bigg(\hat{F}_0\frac{\partial^2 w_1}{\partial x_j \partial q_j}(z')\\ + \frac{\partial \hat{F}_0}{\partial x_j}\frac{\partial w_1}{\partial q_j}(z') - \frac{\partial \hat{F}_0}{\partial q_j}\frac{\partial w_1}{\partial x_j}(z')\Bigg)dz'\\ - i\frac{w_2(0)}{\det(W)}\sum_j\frac{\partial}{\partial x_j}\left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\frac{\Pi_0}{w_1(0)}\right)\\ \int_0^\infty w_1(z')\frac{\partial \hat{F}_0}{\partial q_j}w_1(z')dz'. \end{aligned} \tag{C.5}$$

Although this expression looks more complicated than our initial expression (C.3), it no longer contains F_1.

In what follows, we show how we can evaluate each of the terms in Eq. (C.5) using integration by parts. When we look at the first term in Eq. (C.5), we observe that it contains the product $w_1\hat{F}_0(\partial^2 w_1/\partial x_j\partial q_j)$. The idea of our procedure is to transfer the differential operator $\hat{F}_0$ directly to w_1, at the cost of a few boundary terms, and then to use that $\hat{F}_0 w_1 = 0$, which holds because w_1 is a solution of the homogeneous differential equation. Explicitly, we have

$$\begin{aligned}\int_0^\infty w_1\hat{F}_0\frac{\partial^2 w_1}{\partial x_j\partial q_j}dz' &= \int_0^\infty w_1\frac{\partial}{\partial z}\left(\varepsilon\frac{\partial}{\partial z}\right)\frac{\partial^2 w_1}{\partial x_j\partial q_j}dz' - \int_0^\infty w_1\frac{|\mathbf{q}|^2}{\hbar^2}\varepsilon\frac{\partial^2 w_1}{\partial x_j\partial q_j}dz'\\ &= \left[w_1\varepsilon\frac{\partial}{\partial z}\left(\frac{\partial^2 w_1}{\partial x_j\partial q_j}\right)\right]_0^\infty - \int_0^\infty\frac{\partial w_1}{\partial z}\varepsilon\frac{\partial}{\partial z}\left(\frac{\partial^2 w_1}{\partial x_j\partial q_j}\right)dz'\\ &\quad - \int_0^\infty\frac{|\mathbf{q}|^2}{\hbar^2}\varepsilon w_1(z')\frac{\partial^2 w_1}{\partial x_j\partial q_j}dz'\\ &= \left[w_1\varepsilon\frac{\partial}{\partial z}\left(\frac{\partial^2 w_1}{\partial x_j\partial q_j}\right)\right]_0^\infty - \left[\frac{\partial w_1}{\partial z}\varepsilon\frac{\partial^2 w_1}{\partial x_j\partial q_j}\right]_0^\infty\\ &\quad + \int_0^\infty\left(\frac{\partial}{\partial z}\left(\varepsilon\frac{\partial w_1}{\partial z}\right) - \frac{|\mathbf{q}|^2}{\hbar^2}\varepsilon w_1\right)\frac{\partial^2 w_1}{\partial x_j\partial q_j}dz', \end{aligned} \tag{C.6}$$

where we omitted the argument z' throughout. The last term now contains $\hat{F}_0 w_1$, see Eq. (2.10), which vanishes because w_1 solves the homogeneous equation (2.16). This means that the integral is given by the two boundary terms. We now recall that we assumed that ε becomes constant for $z \to \infty$, and constructed our homogeneous solutions in such a way that w_1 decays exponentially as $z \to \infty$. This implies that all of its derivatives also vanish in this limit, and hence the boundary terms above vanish at infinity.

We therefore have

$$\sum_j \int_0^\infty w_1 \hat{F}_0 \frac{\partial^2 w_1}{\partial x_j \partial q_j} dz' = -\sum_j w_1 \varepsilon \frac{\partial}{\partial z}\left(\frac{\partial^2 w_1}{\partial x_j \partial q_j}\right) + \sum_j \frac{\partial w_1}{\partial z} \varepsilon \left(\frac{\partial^2 w_1}{\partial x_j \partial q_j}\right), \quad \text{(C.7)}$$

where all functions are to be evaluated at $z = 0$.

Let us now consider the last term in Eq. (C.5). Since $\hat{F}_0 w_1 = 0$, we have $(\partial \hat{F}_0 / \partial q_j) w_1 + \hat{F}_0 (\partial w_1 / \partial q_j) = 0$, and

$$\int_0^\infty w_1 \frac{\partial \hat{F}_0}{\partial q_j} w_1 dz' = -\int_0^\infty w_1 \hat{F}_0 \frac{\partial w_1}{\partial q_j} dz' = -\left[w_1 \varepsilon \frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial q_j}\right)\right]_0^\infty + \left[\frac{\partial w_1}{\partial z} \varepsilon \frac{\partial w_1}{\partial q_j}\right]_0^\infty, \quad \text{(C.8)}$$

where the last equality follows from repeated integration by parts, just as in Eq. (C.6).

We now show that the two remaining terms in Eq. (C.5) are equal to

$$\int_0^\infty w_1 \frac{\partial \hat{F}_0}{\partial x_j} \frac{\partial w_1}{\partial q_j} dz' - \int_0^\infty w_1 \frac{\partial \hat{F}_0}{\partial q_j} \frac{\partial w_1}{\partial x_j} dz' = \left[\frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial q_j}\right) \varepsilon \frac{\partial w_1}{\partial x_j}\right]_0^\infty$$

$$+ \left[w_1 \frac{\partial \varepsilon}{\partial x_j} \frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial q_j}\right)\right]_0^\infty - \left[\frac{\partial w_1}{\partial z} \frac{\partial \varepsilon}{\partial x_j} \frac{\partial w_1}{\partial q_j}\right]_0^\infty - \left[\frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial x_j}\right) \varepsilon \frac{\partial w_1}{\partial q_j}\right]_0^\infty . \quad \text{(C.9)}$$

First, note that $\partial \hat{F}_0 / \partial q_j = 2\varepsilon q_j / \hbar^2$ is not a differential operator, which implies that

$$\int_0^\infty w_1 \frac{\partial \hat{F}_0}{\partial q_j} \frac{\partial w_1}{\partial x_j} dz' = \int_0^\infty \left(\frac{\partial \hat{F}_0}{\partial q_j} w_1\right) \frac{\partial w_1}{\partial x_j} dz' = -\int_0^\infty \left(\hat{F}_0 \frac{\partial w_1}{\partial q_j}\right) \frac{\partial w_1}{\partial x_j} dz'$$

$$= -\left[\varepsilon \frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial q_j}\right) \frac{\partial w_1}{\partial x_j}\right]_0^\infty + \int_0^\infty \varepsilon \frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial q_j}\right) \frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial x_j}\right) dz'$$

$$+ \int_0^\infty \frac{\partial w_1}{\partial q_j} \varepsilon \frac{q^2}{\hbar^2} \frac{\partial w_1}{\partial x_j} dz', \quad \text{(C.10)}$$

where the second equality follows from the text above Eq. (C.8), and the last equality follows from integration by parts.

Computing $\partial\hat{F}_0/\partial x_j$ explicitly, repeatedly integrating by parts, and using that $(\partial\hat{F}_0/\partial x_j)w_1+\hat{F}_0(\partial w_1/\partial x_j)=0$, we also find that

$$\begin{aligned}\int_0^\infty w_1\frac{\partial\hat{F}_0}{\partial x_j}\frac{\partial w_1}{\partial q_j}dz' &= \left[w_1\frac{\partial\varepsilon}{\partial x_j}\frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial q_j}\right)\right]_0^\infty-\left[\frac{\partial w_1}{\partial z}\frac{\partial\varepsilon}{\partial x_j}\frac{\partial w_1}{\partial q_j}\right]_0^\infty\\ &\quad+\int_0^\infty\left(\frac{\partial\hat{F}_0}{\partial x_j}w_1\right)\frac{\partial w_1}{\partial q_j}dz'\\ &= \left[w_1\frac{\partial\varepsilon}{\partial x_j}\frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial q_j}\right)\right]_0^\infty-\left[\frac{\partial w_1}{\partial z}\frac{\partial\varepsilon}{\partial x_j}\frac{\partial w_1}{\partial q_j}\right]_0^\infty\\ &\quad-\int_0^\infty\left(\hat{F}_0\frac{\partial w_1}{\partial x_j}\right)\frac{\partial w_1}{\partial q_j}dz'\\ &= \left[w_1\frac{\partial\varepsilon}{\partial x_j}\frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial q_j}\right)\right]_0^\infty-\left[\frac{\partial w_1}{\partial z}\frac{\partial\varepsilon}{\partial x_j}\frac{\partial w_1}{\partial q_j}\right]_0^\infty \qquad \text{(C.11)}\\ &\quad-\left[\varepsilon\frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial x_j}\right)\frac{\partial w_1}{\partial q_j}\right]_0^\infty+\int_0^\infty\varepsilon\frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial x_j}\right)\frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial q_j}\right)dz'\\ &\quad+\int_0^\infty\frac{\partial w_1}{\partial x_j}\varepsilon\frac{q^2}{\hbar^2}\frac{\partial w_1}{\partial q_j}dz'.\end{aligned}$$

Subtracting Eq. (C.10) from Eq. (C.11), we see that the remaining integrals cancel, and we obtain Eq. (C.9).

Inserting the results (C.7), (C.8) and (C.9) in Eq. (C.5) and evaluating all boundary terms, we obtain

$$\begin{aligned}\int_0^\infty g(0,z')f_{1s}(z')dz' &= -\frac{i}{2}\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\frac{\Pi_0}{w_1}\frac{w_2}{\det(W)}\\ &\quad\times\sum_j\frac{\partial}{\partial x_j}\left(w_1\varepsilon\frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial q_j}\right)-\frac{\partial w_1}{\partial q_j}\varepsilon\frac{\partial w_1}{\partial z}\right)\\ &\quad+i\sum_j\frac{\partial}{\partial x_j}\left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\frac{\Pi_0}{w_1}\right)\frac{w_2}{\det(W)}\\ &\quad\times\left(\frac{\partial w_1}{\partial z}\varepsilon\frac{\partial w_1}{\partial q_j}-w_1\varepsilon\frac{\partial}{\partial z}\left(\frac{\partial w_1}{\partial q_j}\right)\right), \qquad \text{(C.12)}\end{aligned}$$

where all functions of z are to be evaluated at $z=0$ from here on. We can then use the relation

$$w_i^2\frac{\partial}{\partial q_j}\left(\frac{1}{w_i}\varepsilon\frac{\partial w_i}{\partial z}\right)=w_i\varepsilon\frac{\partial}{\partial z}\left(\frac{\partial w_i}{\partial q_j}\right)-\frac{\partial w_i}{\partial q_j}\varepsilon\frac{\partial w_i}{\partial z}, \qquad \text{(C.13)}$$

to find

$$\begin{aligned}\int_0^\infty g(0,z')f_{1s}(z')dz' = &-\frac{i}{2}\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\frac{\Pi_0}{w_1}\frac{w_2}{\det(W)}\sum_j \frac{\partial}{\partial x_j}\left(w_1^2\frac{\partial}{\partial q_j}\left(\frac{1}{w_1}\varepsilon\frac{\partial w_1}{\partial z}\right)\right)\\ &- i\sum_j \frac{\partial}{\partial x_j}\left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\frac{\Pi_0}{w_1}\right)\frac{w_2}{\det(W)}w_1^2\frac{\partial}{\partial q_j}\left(\frac{1}{w_1}\varepsilon\frac{\partial w_1}{\partial z}\right).\end{aligned} \tag{C.14}$$

Performing the derivatives of w_1 with respect to x_j, this can also be written as

$$\begin{aligned}\int_0^\infty g(0,z')f_{1s}(z')dz' = &-\frac{i}{2}\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\Pi_0\frac{w_1 w_2}{\det(W)}\sum_j \frac{\partial^2}{\partial x_j \partial q_j}\left(\frac{1}{w_1}\varepsilon\frac{\partial w_1}{\partial z}\right)\\ &- i\sum_j \frac{\partial}{\partial x_j}\left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\Pi_0\right)\frac{w_1 w_2}{\det(W)}\frac{\partial}{\partial q_j}\left(\frac{1}{w_1}\varepsilon\frac{\partial w_1}{\partial z}\right).\end{aligned} \tag{C.15}$$

The integral over the interval $(-\infty, 0]$ can be performed in exactly the same way as in Eqs. (C.6), (C.8) and (C.9). When one interchanges w_1 and w_2 in those outcomes, and changes the integration limits from $[0,\infty)$ to $(-\infty, 0]$, one obtains the results for the interval $(-\infty, 0]$. One can then perform the same steps, to arrive at a result similar to Eq. (C.15). Note, however, that the change of integration limits leads to a relative minus sign between the upper and lower half, since for the lower half all boundary terms vanish at the lower limit of integration.

Combining both results, we find that the integral in Eq. (C.1) equals

$$\begin{aligned}\int_{-\infty}^\infty g(0,z')f_{1s}(z')dz' = &\frac{i}{2}\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\Pi_0\frac{w_1 w_2}{\det(W)}\sum_j \frac{\partial^2}{\partial x_j \partial q_j}\left(\frac{1}{w_2}\varepsilon\frac{\partial w_2}{\partial z} - \frac{1}{w_1}\varepsilon\frac{\partial w_1}{\partial z}\right)\\ &+ i\sum_j \frac{\partial}{\partial x_j}\left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\Pi_0\right)\frac{w_1 w_2}{\det(W)}\\ &\times\frac{\partial}{\partial q_j}\left(\frac{1}{w_2}\varepsilon\frac{\partial w_2}{\partial z} - \frac{1}{w_1}\varepsilon\frac{\partial w_1}{\partial z}\right).\end{aligned} \tag{C.16}$$

We can now use Eqs.(2.24) and (2.25), which show that $w_1 w_2/\det(W) = \hbar/(2|\mathbf{q}|\varepsilon_{\text{eff}})$. This implies that

$$\begin{aligned}\int_{-\infty}^\infty g(0,z')f_{1s}(z')dz' = &\frac{i}{2}\frac{2\pi e^2\hbar}{(\varepsilon_{\text{eff}}|\mathbf{q}|)^2}\Pi_0\sum_j \frac{\partial^2(\varepsilon_{\text{eff}}|\mathbf{q}|)}{\partial x_j \partial q_j}\\ &+ \frac{i}{\varepsilon_{\text{eff}}|\mathbf{q}|}\sum_j \frac{\partial}{\partial x_j}\left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\Pi_0\right)\frac{\partial(\varepsilon_{\text{eff}}|\mathbf{q}|)}{\partial q_j}.\end{aligned} \tag{C.17}$$

This leads us to our final result for Γ_1 at $z = 0$, Eq. (C.1). Using Eq. (2.22) to compute $g(0,0)$, we find

$$\Gamma_1(z=0) = \frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\Pi_1 + \frac{i}{2}\frac{2\pi e^2\hbar}{(\varepsilon_{\text{eff}}|\mathbf{q}|)^2}\Pi_0\sum_j \frac{\partial^2(\varepsilon_{\text{eff}}|\mathbf{q}|)}{\partial x_j \partial q_j} + i\sum_j \frac{\partial}{\partial x_j}\left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\Pi_0\right)\frac{1}{\varepsilon_{\text{eff}}|\mathbf{q}|}\frac{\partial(\varepsilon_{\text{eff}}|\mathbf{q}|)}{\partial q_j}. \tag{C.18}$$

As discussed in Sec. 2.1.3, it is the quantity

$$\mathcal{H}_1 + \frac{i}{2}\frac{\partial^2\mathcal{H}_0}{\partial x_j \partial q_j}, \tag{C.19}$$

that enters in the amplitude of the potential, where $\mathcal{H}_1 = -\Gamma_1(z=0)$. In the final part of this appendix, we compute this quantity, and show that it can be cast in the form given in Eq. (2.42). Since $\mathcal{H}_0$ is given by Eq. (2.26), we have

$$\frac{\partial^2\mathcal{H}_0}{\partial x_j \partial q_j} = \frac{2\pi e^2\hbar}{(\varepsilon_{\text{eff}}|\mathbf{q}|)^2}\Pi_0\frac{\partial^2(\varepsilon_{\text{eff}}|\mathbf{q}|)}{\partial x_j \partial q_j} + \frac{\partial}{\partial x_j}\left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\Pi_0\right)\frac{1}{\varepsilon_{\text{eff}}|\mathbf{q}|}\frac{\partial(\varepsilon_{\text{eff}}|\mathbf{q}|)}{\partial q_j} + \frac{\partial}{\partial q_j}\left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\Pi_0\right)\frac{1}{\varepsilon_{\text{eff}}|\mathbf{q}|}\frac{\partial(\varepsilon_{\text{eff}}|\mathbf{q}|)}{\partial x_j} - \frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\frac{\partial^2\Pi_0}{\partial x_j \partial q_j}. \tag{C.20}$$

Combining Eqs.(C.18) and (C.20), we find

$$\mathcal{H}_1 + \frac{i}{2}\sum_j \frac{\partial^2\mathcal{H}_0}{\partial x_j \partial q_j} = -\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\left(\Pi_1 + \frac{i}{2}\sum_j \frac{\partial^2\Pi_0}{\partial x_j \partial q_j}\right) + \frac{i}{2}\sum_j \frac{\partial}{\partial q_j}\left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\Pi_0\right)\frac{\partial \ln(\varepsilon_{\text{eff}}|\mathbf{q}|)}{\partial x_j} - \frac{i}{2}\sum_j \frac{\partial}{\partial x_j}\left(\frac{2\pi e^2\hbar}{\varepsilon_{\text{eff}}|\mathbf{q}|}\Pi_0\right)\frac{\partial \ln(\varepsilon_{\text{eff}}|\mathbf{q}|)}{\partial q_j}. \tag{C.21}$$

Using the definition of the Poisson bracket

$$\{f,g\} = \sum_j \left(\frac{\partial f}{\partial x_j}\frac{\partial g}{\partial q_j} - \frac{\partial f}{\partial q_j}\frac{\partial g}{\partial x_j}\right), \tag{C.22}$$

and the definition of $\mathcal{H}_0$, see Eq. (2.26), we finally arrive at Eq. (2.42).

C.2 Derivation of an expression for the energy density

In this appendix, we compute the integrated energy density for the potential $V(\mathbf{x},z)$ given by Eq. (2.46). We integrate expression (2.48), which was discussed in Ref. [1], and show that Eq. (2.50) holds.

We first note that ∇ in Eq. (2.48) is the three-dimensional gradient, that is, $\nabla = (\partial/\partial\mathbf{x}, \partial/\partial z)$. Without much loss of generality, we may assume that Γ_0 is real. This roughly corresponds to an Hermitian Hamiltonian, cf. Eq. (2.15). In the terminology introduced in Sec. 4.2, see also Refs. [1, 50], we may say that this situation corresponds to plasmons in a classically allowed region. Note that having a real-valued Γ_0 corresponds to having real-valued functions w_i, see Eq. (2.23). Substituting the potential (2.46) in Eq. (2.48) and only taking the leading-order terms in $\hbar$ into account, we find

$$|\nabla V|^2 = \left|\frac{\partial V}{\partial \mathbf{x}}\right|^2 + \left|\frac{\partial V}{\partial z}\right|^2 = \frac{|A_0^0|^2}{|J(\mathbf{x})|}\frac{1}{\varepsilon_{\rm eff}|\partial S/\partial\mathbf{x}|}\left(\Gamma_0(z)\frac{1}{\hbar^2}\left|\frac{\partial S}{\partial\mathbf{x}}\right|^2\Gamma_0(z) + \frac{\partial\Gamma_0}{\partial z}\frac{\partial\Gamma_0}{\partial z}\right). \tag{C.23}$$

We note that taking the derivative of either Γ_0 or the amplitude with respect to $\mathbf{x}$ leads to higher-order terms in $\hbar$, which we therefore neglect. Moreover, we remark that this result holds regardless of the Berry phase in Eq. (2.46), since it cancels upon taking the absolute value.

We consider the integrated energy density $\mathcal{U}_{\rm I}(\mathbf{x})$, defined as the integral of the energy density $U(\mathbf{x}, z)$ over z, that is

$$\begin{aligned}\mathcal{U}_{\rm I}(\mathbf{x}) &= \int_{-\infty}^{\infty}\mathcal{U}(\mathbf{x},z)dz \\ &= \frac{1}{16\pi e^2}\frac{|A_0^0|^2}{|J(\mathbf{x})|}\frac{1}{\varepsilon_{\rm eff}|\partial S/\partial\mathbf{x}|} \\ &\quad \int_{-\infty}^{\infty}\left(\varepsilon(\mathbf{x},z)\Gamma_0(z)\frac{1}{\hbar^2}\left|\frac{\partial S}{\partial\mathbf{x}}\right|^2\Gamma_0(z) + \varepsilon(\mathbf{x},z)\frac{\partial\Gamma_0}{\partial z}\frac{\partial\Gamma_0}{\partial z}\right)dz. \end{aligned} \tag{C.24}$$

As in the previous appendix, we separate the integral into two parts, corresponding to the upper $[0, \infty)$, and lower $(-\infty, 0]$ halves of the system. Since Γ_0 is symmetric in $z = 0$ upon interchanging w_1 and w_2, see Eq. (2.23), one can infer the outcome for the lower half from the outcome for the upper half. Considering the upper half, and inserting our expression (2.23) for Γ_0, we obtain

$$\begin{aligned}\int_0^{\infty}\mathcal{U}dz &= \frac{1}{16\pi e^2}\frac{|A_0^0|^2}{|J(\mathbf{x})|}\frac{1}{\varepsilon_{\rm eff}|\partial S/\partial\mathbf{x}|} \\ &\quad \int_0^{\infty}\frac{1}{w_1^2(0)}\left(w_1(z)\frac{\varepsilon(\mathbf{x},z)}{\hbar^2}\left|\frac{\partial S}{\partial\mathbf{x}}\right|^2 w_1(z) + \frac{\partial w_1}{\partial z}\varepsilon(\mathbf{x},z)\frac{\partial w_1}{\partial z}\right)dz, \end{aligned} \tag{C.25}$$

where we applied the Hamilton-Jacobi equation, $\mathcal{H}_0(\mathbf{x}, \partial S/\partial\mathbf{x}) = 0$, to impose the condition that $2\pi e^2\hbar\Pi_0/(\varepsilon_{\rm eff}|\partial S/\partial\mathbf{x}|)$ must equal 1.

Integrating the second term by parts, we find

$$\begin{aligned}\int_0^\infty \mathcal{U} dz = {} & \frac{1}{16\pi e^2}\frac{|A_0^0|^2}{|J(\mathbf{x})|}\frac{1}{\varepsilon_{\text{eff}}|\partial S/\partial \mathbf{x}|}\frac{1}{w_1^2(0)} \\ & \times\left(\int_0^\infty w_1(z)\left(\frac{\varepsilon(\mathbf{x},z)}{\hbar^2}\left|\frac{\partial S}{\partial \mathbf{x}}\right|^2 w_1(z) - \frac{\partial}{\partial z}\left(\varepsilon(\mathbf{x},z)\frac{\partial w_1}{\partial z}\right)\right)dz\right. \\ & \left. + \left[w_1(z)\varepsilon(\mathbf{x},z)\frac{\partial w_1}{\partial z}\right]_0^\infty\right). \end{aligned} \tag{C.26}$$

The remaining integral on the right-hand side vanishes because $w_1(z)$ satisfies the homogeneous differential equation $\hat{F}_0 w_1(z) = 0$, cf. the discussion above Eq. (C.4) in the previous appendix.

Since both w_1 and its derivatives go to zero as $z \to \infty$, part of the boundary term also vanishes, and we are left with the contribution at $z = 0$. Adding the contribution from the lower half, which comes with a relative minus sign because the boundaries are $-\infty$ and 0, we obtain

$$\begin{aligned}\mathcal{U}_{\text{I}}(\mathbf{x}) &= \int_{-\infty}^\infty \mathcal{U}(\mathbf{x},z)dz \qquad \text{(C.27)} \\ &= \frac{1}{16\pi e^2}\frac{|A_0^0|^2}{|J(\mathbf{x})|}\frac{1}{\varepsilon_{\text{eff}}|\partial S/\partial \mathbf{x}|}\left(-\frac{\varepsilon(\mathbf{x},0)}{w_1(0)}\frac{\partial w_1(0)}{\partial z} + \frac{\varepsilon(\mathbf{x},0)}{w_2(0)}\frac{\partial w_2(0)}{\partial z}\right).\end{aligned}$$

Finally, using the definition of ε_{eff} from Eq. (2.25), we find

$$\mathcal{U}_{\text{I}}(\mathbf{x}) = \frac{1}{16\pi e^2}\frac{|A_0^0|^2}{|J(\mathbf{x})|}\frac{2\varepsilon_{\text{eff}}|\partial S/\partial \mathbf{x}|}{\hbar\varepsilon_{\text{eff}}|\partial S/\partial \mathbf{x}|} = \frac{1}{8\pi e^2\hbar}\frac{|A_0^0|^2}{|J(\mathbf{x})|}. \tag{C.28}$$

This result exactly coincides with Eq. (81) in Ref. [1]. However, this time we started from an arbitrary model for $\varepsilon(\mathbf{x},z)$, instead of a simplified model.

D Simple turning point

The behavior of q_x around a turning point, where it vanishes, is fundamental to understanding the nature of plasmonic bound states. In this appendix, we demonstrate that these turning points are simple turning points [48, 75, 76, 86], independent of the system parameters.

To analyze this behavior, we must consider the role of the effective dielectric function $\varepsilon_{\text{eff}}(\mathbf{x}, \mathbf{q})$. While $\varepsilon_{\text{eff}}(\mathbf{x}, \mathbf{q})$ is momentum dependent, we argue that this variation does not qualitatively alter the scaling of q_x near the turning point. To justify this statement, let us examine two limiting cases: as seen in Fig. 4.2(b), in the large $|\mathbf{q}|$ limit, screening becomes constant at ε_{M}. Conversely, in the small $|\mathbf{q}|$ limit, the system behaves similarly as previously analyzed in Ref. [1]. At both these limits, there are no discontinuities or divergences in $\varepsilon_{\text{eff}}(\mathbf{x}, \mathbf{q})$. Moreover, it behaves smoothly as a function of $\mathbf{q}$, meaning that we can perform a Taylor expansion to first order in $\mathbf{q}$. We therefore argue that the variation of $\varepsilon_{\text{eff}}(\mathbf{x}, \mathbf{q})$ does not modify the behavior of q_x at the turning points.

When q_y is small, we can analyze the behavior near the turning points analytically. We start our analysis from the effective classical Hamiltonian, Eq. (2.26), which in the limit of small momenta $|\mathbf{q}| \ll 1$ takes the approximate form

$$\mathcal{H}_0 \approx 1 - \frac{g_{\text{s}} e^2 p_{\text{F}}^2(\mathbf{x})|\mathbf{q}|}{2m\varepsilon_{\text{avg}}(\mathbf{x})\hbar E^2} + \mathcal{O}(|\mathbf{q}^2|), \tag{D.1}$$

where the average dielectric function is defined as

$$2\varepsilon_{\text{avg}} = \varepsilon_{\text{A}} + \varepsilon_{\text{B}}, \tag{D.2}$$

which follows naturally from the effective dielectric function $\varepsilon_{\text{eff}}(|\mathbf{q}| \to 0)$. Since plasmons are defined by the equation $\mathcal{H}_0 = 0$, we directly have

$$|\mathbf{q}| = \frac{2m\varepsilon_{\text{avg}}(x)\hbar E^2}{g_{\text{s}} e^2 p_{\text{F}}^2(x)}. \tag{D.3}$$

Expressing the total momentum in terms of its components, $|\mathbf{q}|^2 = q_x^2 + q_y^2$, we obtain

$$q_x^2 = \frac{2m\varepsilon_{\text{avg}}(x)\hbar E^2}{g_{\text{s}} e^2 p_{\text{F}}^2(x)} - q_y^2. \tag{D.4}$$

By definition, $q_x = 0$ at the turning point, which gives us a relation between x_{c}, q_y and E.

$$q_y^2 = \left(\frac{2m\varepsilon_{\text{avg}}(x_{\text{c}})\hbar E^2}{g_{\text{s}} e^2 p_{\text{F}}^2(x_{\text{c}})} \right)^2. \tag{D.5}$$

To determine the nature of the turning point, we Taylor-expand q_x^2 around x_{c}, which yields

$$q_x^2 = \left(\frac{2m\varepsilon_{\mathrm{avg}}(x_{\mathrm{c}})\hbar E^2}{g_{\mathrm{s}}e^2 p_{\mathrm{F}}^2(x_{\mathrm{c}})}\right)^2 - q_y^2 + (x - x_{\mathrm{c}})\left(\frac{2m\varepsilon_{\mathrm{avg}}(x)\hbar E^2}{g_{\mathrm{s}}e^2 p_{\mathrm{F}}^2(x)}\right)^2 \left(2\frac{\varepsilon'_{\mathrm{avg}}(x)}{\varepsilon_{\mathrm{avg}}(x)} - 4\frac{p'_{\mathrm{F}}(x)}{p_{\mathrm{F}}(x)}\right)\bigg|_{x_{\mathrm{c}}}. \tag{D.6}$$

With Eq. (D.5), we can substitute q_y as a function of x_{c} and E. We find

$$q_x^2 = (x - x_{\mathrm{c}})\left(\frac{2m\varepsilon_{\mathrm{avg}}(x)\hbar E^2}{g_{\mathrm{s}}e^2 p_{\mathrm{F}}^2(x)}\right)^2 \left(2\frac{\varepsilon'_{\mathrm{avg}}(x)}{\varepsilon_{\mathrm{avg}}(x)} - 4\frac{p'_{\mathrm{F}}(x)}{p_{\mathrm{F}}(x)}\right)\bigg|_{x_{\mathrm{c}}}. \tag{D.7}$$

By definition, this confirms that the turning point is simple, as q_x^2 depends linearly on $(x - x_{\mathrm{c}})$. To eliminate the explicit dependence on x_{c}, we may express x_{c} as a function of (q_y, E), i.e., $x_{\mathrm{c}} = x_{\mathrm{c}}(q_y, E)$.

For small q_y, we have thus explicitly shown that the turning points are simple. For larger values of q_y, we numerically confirmed that the proportionality $q_x^2 \propto x - x_c$ continues to hold, meaning that the turning points remain simple.

Bibliography

[1] T. M. Koskamp, M. I. Katsnelson, and K. J. A. Reijnders, Phys. Rev. B **108**, 085414 (2023).

[2] S. V. Vonsovsky and M. I. Katsnelson, *Quantum solid-state physics* (Springer-Verlag, Berlin Heidelberg, 1989).

[3] M. S. Tame, K. R. McEnery, Ş. K. Özdemir, J. Lee, S. A. Maier, and M. S. Kim, Nat. Phys. **9**, 329 (2013).

[4] W. L. Barnes, A. Dereux, and T. W. Ebbesen, Nature **424**, 824 (2003).

[5] A. N. Grigorenko, M. Polini, and K. S. Novoselov, Nat. Photonics **6**, 749 (2012).

[6] E. Ozbay, Science **311**, 189 (2006).

[7] Z. Fei, G. O. Andreev, W. Bao, L. M. Zhang, A. S. McLeod, C. Wang, M. K. Stewart, Z. Zhao, G. Dominguez, M. Thiemens, M. M. Fogler, M. J. Tauber, A. H. Castro-Neto, C. N. Lau, F. Keilmann, and D. N. Basov, Nano Letters **11**, 4701 (2011), pMID: 21972938.

[8] J. Chen, M. Badioli, P. Alonso-González, S. Thongrattanasiri, F. Huth, J. Osmond, M. Spasenović, A. Centeno, A. Pesquera, P. Godignon, A. Zurutuza Elorza, N. Camara, F. J. G. de Abajo, R. Hillenbrand, and F. H. L. Koppens, Nature **487**, 77 (2012).

[9] G. Di Martino, Y. Sonnefraud, S. Kéna-Cohen, M. Tame, Ş. K. Özdemir, M. S. Kim, and S. A. Maier, Nano Letters **12**, 2504 (2012).

[10] J. A. Schuller, E. S. Barnard, W. Cai, Y. C. Jun, J. S. White, and M. L. Brongersma, Nature materials **9**, 193 (2010).

[11] L. Landau, J. Exp. Theor. Phys. **16**, 574 (1946).

[12] X. Li, D. Xiao, and Z. Zhang, New Journal of Physics **15** (2013).

[13] S. Ghosh, G. Menichetti, M. I. Katsnelson, and M. Polini, Phys. Rev. B **107**, 195302 (2023).

[14] R. J. Koch, T. Seyller, and J. A. Schaefer, Phys. Rev. B **82**, 201413 (2010).

[15] C. R. Yonzon, D. A. Stuart, X. Zhang, A. D. McFarland, C. L. Haynes, and R. P. Van Duyne, Talanta **67**, 438 (2005).

[16] W. Srituravanich, N. Fang, C. Sun, Q. Luo, and X. Zhang, Nano Letters **4**, 1085 (2004).

[17] J. Pendry, A. Holden, D. Robbins, and W. Stewart, IEEE Transactions on Microwave Theory and Techniques **47**, 2075 (1999).

[18] D. R. Smith, J. B. Pendry, and M. C. K. Wiltshire, Science **305**, 788 (2004).

[19] S. I. Bozhevolnyi, V. S. Volkov, E. Devaux, J.-Y. Laluet, and T. W. Ebbesen, Nature **440**, 508 (2006).

[20] A. A. Rukhadze and V. P. Silin, Soviet Physics Uspekhi **4** (1961).

[21] J. M. Fitzgerald, P. Narang, R. V. Craster, S. A. Maier, and V. Giannini, Proc. IEEE **104**, 2307 (2016).

[22] J. A. Scholl, A. L. Koh, and J. A. Dionne, Nature **483**, 421 (2012).

[23] A. H. Castro Neto, F. Guinea, N. M. Peres, K. S. Novoselov, and A. K. Geim, Reviews of Modern Physics **81**, 109 (2009).

[24] A. K. Geim and I. V. Grigorieva, Nature **499**, 419 (2013).

[25] J. D. Jackson, *Classical electrodynamics* (Wiley, New York, 1999).

[26] L. Landau, Collected Papers of L.D. Landau **10** (1945).

[27] J. D. Jackson, Journal of Nuclear Energy. Part C, Plasma Physics, Accelerators, Thermonuclear Research **1**, 171 (1960).

[28] G. F. Giuliani and G. Vignale, *Quantum theory of the electron liquid* (Cambridge University Press, Cambridge, 2005).

[29] C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 2005).

[30] A. N. Grigorenko, M. Polini, and K. S. Novoselov, Nature Photonics **6**, 749 (2012).

[31] M. I. Katsnelson, *The Physics of Graphene* (Cambridge University Press, 2020).

[32] A. Czachor, A. Holas, S. r. Sharma, and K. S. Singwi, Physical Review B **25** (1982).

[33] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).

[34] T. O. Wehling, E. Şaşıoğlu, C. Friedrich, A. I. Lichtenstein, M. I. Katsnelson, and S. Blügel, Phys. Rev. Lett. **106**, 236805 (2011).

[35] C. Steinke, T. O. Wehling, and M. Rösner, Phys. Rev. B **102**, 115111 (2020).

[36] M. Rösner, E. Şaşıoğlu, C. Friedrich, S. Blügel, and T. Wehling, Phys. Rev. B **92**, 085102 (2015).

[37] F. Stern and W. E. Howard, Physical Review **163**, 816 (1967).

[38] T. Westerhout, E. van Veen, M. I. Katsnelson, and S. Yuan, Phys. Rev. B **97**, 205434 (2018).

[39] M. Rösner, C. Steinke, M. Lorke, C. Gies, F. Jahnke, and T. O. Wehling, Nano Lett. **16**, 2322 (2016).

[40] Z. Jiang, S. Haas, and M. Rösner, 2D Mater. **8**, 035037 (2021).

[41] S. Yuan, R. Roldán, A.-P. Jauho, and M. I. Katsnelson, Phys. Rev. B **87**, 085430 (2013).

[42] C. L. Haynes, C. R. Yonzon, X. Zhang, and R. P. Van Duyne, Journal of Raman Spectroscopy **36**, 471 (2005).

[43] A. C. R. Pipino, G. C. Schatz, and R. P. V. Duyne, Physical review B **49**, 8320 (1994).

[44] S. Yuan, R. Roldán, and M. I. Katsnelson, Phys. Rev. B **84**, 035439 (2011).

[45] F. Jin, R. Roldán, M. I. Katsnelson, and S. Yuan, Phys. Rev. B **92**, 115440 (2015).

[46] Y. Li, Z. Zhan, X. Kuang, Y. Li, and S. Yuan, Comput. Phys. Commun. **285**, 108632 (2023).

[47] I. Torre, M. I. Katsnelson, A. Diaspro, V. Pellegrini, and M. Polini, Phys. Rev. B **96**, 035433 (2017).

[48] V. P. Maslov and M. V. Fedoryuk, *Semi-classical approximation in quantum mechanics* (Reidel, Dordrecht, 1981).

[49] A. Martinez, *An introduction to semiclassical and microlocal analysis* (Springer-Verlag, New York, 2002).

[50] K. J. A. Reijnders, T. Tudorovskiy, and M. I. Katsnelson, Ann. Phys. **446**, 169116 (2022).

[51] B. Kh. Ishmukhametov, Phys. Status Solidi (B) **45**, 669 (1971).

[52] B. Kh. Ishmukhametov and M. I. Katsnelson, Fiz. Metallov i Metallovedenie **40**, 736 (1975).

[53] B. Kh. Ishmukhametov, M. I. Katsnelson, V. N. Larionov, and A. M. Ustyuzhanin, phys. stat. sol. (b) **104**, K75 (1981).

[54] B. Kh. Ishmukhametov, M. I. Katsnelson, V. N. Larionov, and A. M. Ustjuzhanin, Phys. Lett. A **82**, 387 (1981).

[55] V. Guillemin and S. Sternberg, *Geometric asymptotics* (American Mathematical Society, Providence, Rhode Island, 1977).

[56] V. V. Belov, S. Yu. Dobrokhotov, and T. Ya. Tudorovskiy, J. Eng. Math. **55**, 183 (2006).

[57] R. Goldston and P. Rutherford, *Introduction to Plasma Physics* (IOP Publishing Ltd., 1995).

[58] E. H. Lieb, Rev. Mod. Phys. **53**, 603 (1981).

[59] S. Raza, S. I. Bozhevolnyi, M. Wubs, and N. Asger Mortensen, Journal of Physics: Condensed Matter **27**, 183204 (2015).

[60] P. M. Platzman and P. A. Wolff, *Waves and Interactions in Solid State Plasmas* (Academic Press, New York, 1973).

[61] D. J. Griffiths, *Introduction to quantum mechanics*, 2nd ed. (Pearson Prentice Hall, Upper Saddle River, 2005).

[62] B. Kh. Ishmukhametov and M. I. Katsnelson, Fiz. Met. Metalloved. **40**, 736 (1975).

[63] J. Heading, *An Introduction to Phase-Integral Methods* (Methuen, London, 1962).

[64] S. Flügge, *Practical Quantum Mechanics* (Springer-Verlag, Berlin Heidelberg, 1994).

[65] N. Fröman and P. O. Fröman, *Physical Problems Solved by the Phase-Integral Method* (Cambridge University Press, Cambridge, 2002).

[66] A. Zwaan, *Intensitäten im Ca-Funkenspektrum*, Ph.D. thesis, Utrecht university (1929).

[67] M. Zworski, *Semiclassical analysis* (American Mathematical Society, Providence, Rhode Island, 2012).

[68] B. C. Hall, *Quantum theory for mathematicians* (Springer, New York, 2013).

[69] T. M. Koskamp, M. I. Katsnelson, and K. J. A. Reijnders, arXiv preprint arXiv:2502.20512 (2025).

[70] A. Emelyanenko and L. Boinovich, Journal of Physics: Condensed Matter **20**, 494227 (2008).

[71] L. V. Berlyand and S. Yu. Dobrokhotov, Dokl. Akad. Nauk SSSR **296**, 80 (1987).

[72] G. A. Korn and T. M. Korn, *Mathematical handbook for scientists and engineers: definitions, theorems, and formulas for reference and review* (Dover publications, New York, 2000).

[73] V. I. Arnold, *Mathematical methods of classical mechanics*, 2nd ed. (Springer, New York, 1989).

[74] H. Goldstein, C. P. Poole, and J. L. Safko, *Classical Mechanics*, 3rd ed. (Addison Wesley, San Fransisco, 2002).

[75] T. Poston and I. N. Stewart, *Catastrophe theory and its applications* (Pitman, Boston, 1978).

[76] V. I. Arnold, S. M. Gusein-Zade, and A. N. Varchenko, *Singularities of Differentiable Maps, Volume 1, The Classification of Critical Points, Caustics and Wave Fronts* (Birkhäuser, Basel, 1985).

[77] L. D. Landau, E. M. Lifshitz, and L. P. Pitaevskii, *Electrodynamics of continuous media* (Pergamon Press, Oxford, 1984).

[78] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, Rev. Mod. Phys. **84**, 1419 (2012).

[79] K. J. A. Reijnders, D. S. Minenkov, M. I. Katsnelson, and S. Yu. Dobrokhotov, Ann. Phys. **397**, 65 (2018).

[80] S. Yu. Dobrokhotov and P. N. Zhevandrov, Russ. J. Math. Phys. **10**, 1 (2003).

[81] J. R. Taylor, *Scattering theory: the quantum theory of nonrelativistic collisions* (Dover publications, New York, 1972).

[82] M. Abromowitz and I. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, New York, 1964) pp. 358–364.

[83] B. Vainberg, *Asymptotic methods in equations of mathematical physics* (Gordon and Breach Science Publishers, New York, 1989).

[84] A. Sommerfeld, *Partial differential equations in physics* (Academic press, New York, 1949).

[85] S. H. Schot, Hist. Math. **19**, 385 (1992).

[86] M. V. Berry and K. E. Mount, Rep. Progr. Phys. **35**, 315 (1972).

[87] L. J. Curtis and D. G. Ellis, Am. J. Phys. **72**, 1521 (2004).

[88] V. I. Arnold, Funct. Anal. Appl. **1**, 1 (1967).

[89] R. E. Langer, Phys. Rev. **51**, 669 (1937).

[90] Wolfram Mathematica (1988–2023), `www.wolfram.com/mathematica`. The computations in this article were performed with version 13.0.0.0.

[91] M. Born and E. Wolf, *Principles of optics: Electromagnetic theory of propagation, interference and diffraction of light* (Cambridge University Press, Cambridge, 1999).

[92] H.-J. Stöckmann, *Quantum Chaos* (Cambridge University Press, Cambridge, 2007).

[93] C. López, Advanced Materials **15**, 1679 (2003).

[94] S. G. Menabde, I.-H. Lee, S. Lee, H. Ha, J. T. Heiden, D. Yoo, T.-T. Kim, T. Low, Y. H. Lee, S.-H. Oh, *et al.*, Nature communications **12**, 938 (2021).

[95] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics Non-relativistic Theory* (Pergamon Press, Oxford, 1981).

[96] K. J. A. Reijnders, T. Tudorovskiy, and M. I. Katsnelson, Ann. Phys. **333**, 155 (2013).

[97] N. Fröman, Journal of Physics A: Mathematical and General **12**, 2355 (1979).

[98] N. Fröman and P. O. Fröman, *JWKB Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965).

[99] A. M. Dykhne, Soviet Physics Jetp **13**, 1000 (1961).

[100] L. V. Keldysh, Pis'ma Zh. Eksp. Teor Fiz **29**, 716 (1979).

[101] Z. Fei, A. S. Rodin, G. O. Andreev, W. Bao, A. S. McLeod, M. Wagner, L. M. Zhang, Z. Zhao, M. Thiemens, G. Dominguez, M. M. Fogler, A. H. Castro Neto, C. N. Lau, F. Keilmann, and D. N. Basov, Nature **487**, 82 (2012).

[102] E. G. C. P. van Loon, H. Hafermann, A. I. Lichtenstein, A. N. Rubtsov, and M. I. Katsnelson, Phys. Rev. Lett. **113**, 246407 (2014).

[103] L. Xiong, C. Forsythe, M. Jung, A. McLeod, S. Sunku, Y. Shao, G. Ni, A. Sternbach, S. Liu, J. Edgar, *et al.*, Nature communications **10**, 4780 (2019).

[104] S. Das Sarma and E. H. Hwang, Phys. Rev. Lett. **81**, 4216 (1998).

[105] A. Bill, H. Morawitz, and V. Z. Kresin, Phys. Rev. B **68**, 144519 (2003).

[106] H. Morawitz, I. Bozovic, V. Kresin, G. Rietveld, and D. Van Der Marel, Zeitschrift für Physik B Condensed Matter **90**, 277 (1993).

[107] C. Song, X. Yuan, C. Huang, S. Huang, Q. Xing, C. Wang, C. Zhang, Y. Xie, Y. Lei, F. Wang, *et al.*, Nature Communications **12**, 386 (2021).

[108] C. Horie, Progress of theoretical Physics **21**, 113 (1959).

[109] I. Egri, Solid State Communications **44**, 563 (1982).

[110] F. Alpeggiani and L. C. Andreani, Phys. Rev. B **90**, 115311 (2014).

[111] A. P. Itin and M. I. Katsnelson, Phys. Rev. Lett. **115**, 075301 (2015).

[112] C. Dutreix and M. I. Katsnelson, Phys. Rev. B **95**, 024306 (2017).

Catching the Wave: a Guide through the Electron Sea

In physics, our understanding of the natural world is often built upon theoretical models, mathematical frameworks designed to capture the essential behavior of a system. This approach is particularly powerful in condensed matter physics, the field that explores the properties of materials that form the basis of modern technology. While experiments and large-scale computer simulations provide crucial data, new discoveries, and verification of theories, the journey of understanding often begins with an analytical theory that seeks to predict and explain physical phenomena from the fundamentals of nature. This thesis is a work of analytical theory, translating complex physical principles into mathematical formulas with the goal of producing interesting and verifiable predictions.

My research focuses on plasmons, which are collective, wave-like oscillations rippling through the sea of electrons in a conducting material. Much like waves of water traveling across the surface of an ocean, plasmons are propagating waves of electron density in metals and semiconductors. And just like ocean waves, which are essentially two-dimensional surface phenomena on a very deep body of water, the plasmons we study are confined to a two-dimensional plane. These plasmons are interesting, because they interact strongly with light, for example light from optic fiber internet cables. However, plasmons behave on much shorter length scales. This allows for miniaturization of optical chips, a significant advance for a field where component size is a primary drawback.

Now, a wave traveling across the open, uniform ocean is predictable, perhaps even a little boring and relatively simple to describe. The physics becomes far more interesting and complex when the wave encounters a disturbance or inhomogeneity in the environment, such as a coastline or an island. The same is true for plasmons, which are sensitive to local variations in the electron density or material they travel through. To understand such behavior in realistic materials, we need a theory that can handle spatial interruptions and incorporates inhomogeneous systems.

The specific theory employed in this thesis is the semiclassical approximation also known as the WKB approximation. This is a powerful mathematical framework for describing waves in non-uniform environments, with one important condition: the changes or variations must be smooth. In physical terms, this means that the properties of the material change over length scales that are much larger than the wavelength of the wave itself. Think of it this way: our theory is excellent at describing an ocean wave rolling up a long, gentle beach. It would be less suited for describing a wave crashing abruptly on a sea stack. Fortunately, at the microscopic scale of materials, changes are often quite smooth. Atoms at the boundary between two different materials tend to mingle, and the sea of electrons naturally smooths out any sharp edges. This makes the semiclassical theory a perfect fit for describing plasmons in these real-world, inhomogeneous systems. The theory itself involves a lot of complex and lengthy mathematics (which the interested reader can find in chapter 2 of this thesis!), but its application to three specific physical scenarios yields intuitive and beautiful results.

Scattering: what happens when a plasmon hits an "island"? Just as an ocean wave scatters and changes direction when it hits an island or smooth rock, a plasmon scatters when it encounters a local change in the material or electron density, altering its path and intensity. Our theory shows that plasmons are attracted to regions with lower electron density and repelled by regions with higher density. This can induce some interesting effects. A circular symmetric inhomogeneity of reduced electron density pulls and bends the incoming plasmon towards it like a sink. The plasmons turn inward towards the valley, before they are scattered away. For deep enough valleys they even turn 180 degrees, going backwards after scattering. An increase in electron density serves as a kind of potential barrier, pushing the plasmon outward, away from the peak. We demonstrated that the scattered waves can exhibit interference: scattered waves overlap and can combine constructively (amplifying each other) or destructively (canceling each other out), a hallmark of wave-like behavior.

Bound states: trapping a plasmon in a "canal". What if a wave isn't in an open unrestricted space like a sea or ocean, but is confined to a narrow channel like a river or canal flowing between two banks? The walls of this canal force the wave to bounce back and forth, and due to the boundaries only certain wave patterns, or "modes", are allowed to exist. We showed that the same thing can happen with plasmons. By creating two parallel walls of, for example, high electron density, which repel plasmons, we can trap a plasmon wave. However, creating these plasmon waveguides isn't as simple as just building higher walls. Due to the unique nature of these 2D waves, the barriers only become effective if the plasmon is already moving with some forward momentum along the length of the canal. Subsequently, this momentum creates what we call bound states, where the plasmon is guided along a specific path. The lowest energy allowed plasmon mode, is the one with only one node in the canal, and when the energy of the plasmon increases, so does the number of nodes in between the walls.

Periodic structures: when trapped plasmons "talk" to each other. Finally, we looked at something a bit more abstract, where the water analogy begins to fade and we need to incorporate quantum-like behavior like tunneling. Imagine an array of these previously discussed plasmonic waveguides placed side-by-side. If the barriers between the canals are not too high or wide, plasmons can “tunnel” from one canal to the next, a quantum mechanical phenomenon where a particle penetrates a barrier it classically cannot surmount. This interaction creates a collective behavior across the entire array, leading to the formation of a so-called band structure. This means that for certain energy ranges (the allowed bands), plasmons can propagate freely throughout the periodic structure. For other energies (the band gaps), propagation is forbidden. This is, once more, highly dependent on the velocity of the plasmon in the direction of the waveguides.

These three applications, scattering, bound states, and periodic structures, are demonstrations of a versatile theoretical framework for predicting plasmon behavior. The ability to guide plasmons in waveguides, for instance, is a key requirement for using them to transmit information on a microchip. For example, if we can excite plasmons with light from our long distance optical fibers, we can pass this information locally through plasmonic waveguides, interact with them and perform computational steps on a much smaller scale than conventional optics. At its core, this thesis develops a fundamental theory. The next logical step is the experimental verification of its predictions, which can in turn refine our understanding and guide the design of novel plasmonic devices and new technologies. The path from a fundamental theory to new technology is often long, but it begins with a robust and predictive mathematical model of the physical world and a desire to understand the beautiful and complex behavior of waves in the microscopic sea of electrons.

Een Handleiding door de Elektronenzee: Geen Golf te Hoog

In de natuurkunde is ons begrip van de wereld vaak gebaseerd op theoretische modellen: wiskundige kaders die zijn ontworpen om het gedrag van een systeem te beschrijven. Deze aanpak is bijzonder krachtig in de vastestoffysica, het vakgebied dat de eigenschappen van materialen onderzoekt die de basis vormen van moderne technologie. Hoewel experimenten en grootschalige computersimulaties cruciale data leveren, begint de ontdekkingsreis vaak met een theorie die natuurkundige fenomenen probeert te voorspellen en te verklaren vanuit de fundamentele wetten van de natuur. Dit proefschrift beschrijft zo een analytische theorie, dat complexe fysische principes vertaalt in wiskundige formules met als doel interessante en verifieerbare voorspellingen te doen.

Dit onderzoek richt zich op plasmonen: collectieve, golfachtige trillingen die door de zee van elektronen in een geleidend materiaal deinen. Net zoals dat water golft over het oppervlak van een oceaan, golft voor een plasmon de elektronendichtheid als het ware door een materiaal. De systemen die we in dit proefschrift bestuderen zijn tweedimensionaal. Dit betekent dat de plasmonen beperkt zijn tot een oppervlak. We kunnen dit wederom vergelijken met oceaangolven, welke ten opzichte van de diepte van de oceaan ook beschouwd kunnen worden als een oppervlaktefenomeen. Plasmonen zijn interessant, omdat ze een sterke interactie hebben met licht. Echter, een belangrijk voordeel van plasmonen is dat hun golflengte veel kleiner is dan die van licht met dezelfde frequentie. Hierdoor kunnen we mogelijk optische chips miniaturiseren, een belangrijke vooruitgang voor een vakgebied waar de grootte van componenten een primair nadeel is.

Een golf die door een uniform medium beweegt, is relatief eenvoudig te beschrijven. De fysica wordt veel interessanter en complexer wanneer de golf een verstoring of inhomogeniteit in de omgeving tegenkomt, zoals een kustlijn of een eiland voor een oceaangolf. Hetzelfde geldt voor plasmonen, die gevoelig zijn voor lokale variaties in de elektronendichtheid of het materiaal waar ze doorheen reizen. Om hun gedrag in realistische systemen te begrijpen, is een geavanceerde theorie nodig.

Het specifieke wiskundige hulpmiddel dat in dit proefschrift wordt toegepast, is de semiklassieke theorie. Dit is een krachtig wiskundig raamwerk voor het beschrijven van golven in niet-homogene omgevingen, op voorwaarde dat deze variaties geleidelijk zijn. Fysisch gezien betekent dit dat de eigenschappen van het materiaal veranderen over lengteschalen die veel groter zijn dan de golflengte van de golf zelf. Het is als het beschrijven van een oceaangolf die een lang, zacht strand oprolt, in tegenstelling tot een golf die abrupt op een rots slaat. Deze benadering is zeer geschikt voor veel elektronensystemen, waar de overgangen tussen verschillende componenten zelden abrupt zijn, omdat verschillende materialen vaak al een beetje mengen op het grensvlak. Het beschrijven van deze theorie vergt veel gecompliceerde formules (neem vooral een kijkje in hoofdstuk 2!), maar de toepassing van deze theorie op drie fysische scenario's levert mooie en intuïtieve resultaten.

Verstrooiing: het botsen van plasmonen op een 'eiland'. Een oceaangolf kan worden verstoord door een verandering in de omgeving, zoals het botsen tegen een eiland. Wanneer een plasmon een gelokaliseerde verandering in het materiaal tegenkomt wijzigt zijn pad en intensiteit, het verstrooit. Onze theorie toont aan dat plasmonen worden aangetrokken tot gebieden met een lagere elektronendichtheid en worden afgestoten door gebieden met een hogere elektronendichtheid. Een circulair symmetrische inhomogeniteit met een verlaagde elektronendichtheid trekt en buigt het inkomende plasmon naar zich toe als een afvoerputje. Wanneer de put diep genoeg is, kan de plasmon zelfs gedeeltelijk 180 graden omdraaien. Verder hebben we aangetoond dat de verstrooide golven gedeeltelijk kunnen overlappen en daardoor interferentie vertonen: ze kunnen op twee manieren combineren, namelijk constructief (elkaar versterken) of destructief (elkaar uitdoven), kenmerkend voor golven.

Gebonden toestanden: plasmonen gevangen in een 'kanaal'. Wanneer een golf door een kanaal stroomt, wordt zijn beweging beperkt door de waterkant, waardoor hij heen en weer kaatst tussen de randen. In principe staat deze beperking slechts een discrete set van gekwantiseerde stabiele golfpatronen, of 'modes', toe. We hebben aangetoond dat hetzelfde principe geldt voor plasmonen. Door een kanaal te creëren, afgescheiden door twee 'muren' van bijvoorbeeld hoge elektronendichtheden (die plasmonen afstoten), kunnen we de golf vangen. Het creëren van deze plasmon-golfgeleiders is echter niet zo eenvoudig als simpelweg hoge muren bouwen. Vanwege de unieke aard van deze 2D-golven worden de barrières pas effectief als het plasmon al met een voorwaartse impuls in de lengterichting van het kanaal beweegt. Deze structuur fungeert dan als een plasmonische golfgeleider, die de golf langs een gedefinieerd kanaal leidt in de vorm van deze gekwantiseerde gebonden toestanden.

Periodieke structuren: wanneer gevangen plasmonen 'communiceren'. Tot slot hebben we een complexer systeem onderzocht: een reeks van de zojuist besproken golfgeleiders naast elkaar. Als de barrières tussen de kanalen niet oneindig hoog zijn, moeten we kwantumachtig gedrag zoals 'tunneling' meenemen. Dit is een fenomeen waarbij een golf of deeltje door een barrière dringt waar het klassiek gezien niet doorheen kan. Deze interactie of communicatie tussen aangrenzende kanalen creëert een collectief gedrag over de hele reeks, wat leidt tot de vorming van een bandstructuur. Dit betekent dat voor bepaalde energiebereiken, de toegestane banden, plasmonen door de periodieke structuur mogen bewegen. Voor andere energieën, de verboden banden of ‘bandgaps’, is voortbeweging echter verboden. Met deze periodieke structuren kan je dus reguleren welke selectie aan energieën in het systeem kunnen propageren.

Deze drie toepassingen, verstrooiing, plasmonische golfgeleiders en periodieke structuren, zijn demonstraties van het veelzijdige theoretisch kader voor het voorspellen van het gedrag van plasmonen. De mogelijkheid om plasmonen in golfgeleiders te sturen is bijvoorbeeld een vereiste om ze te gebruiken voor het overbrengen of verplaatsen van informatie op een microchip. In de kern ontwikkelt dit proefschrift een fundamentele theorie. De volgende logische stap is de experimentele verificatie van de voorspellingen, wat op haar beurt ons begrip kan verfijnen en het ontwerp van plasmonische toepassingen in de technologie kan sturen. De weg van een fundamentele theorie naar zo’n nieuwe technologie is vaak lang, maar begint met een robuust en voorspellend model van het gedrag, noem het een handleiding van de golven in de elektronenzee.

Acknowledgments

First and foremost, I would like to thank my supervisor, Misha Katsnelson, for giving me the opportunity to pursue a PhD and for his invaluable support throughout this journey. Your vast knowledge of physics and your ability to generate new ideas seemingly effortlessly have been a constant source of inspiration. While the beginning of my PhD was marked by the challenges of the Covid pandemic and my own steep learning curve in plasmonics, I grew to deeply appreciate your didactic clarity and depth in our later interactions. Thank you for your guidance, patience, and for helping me grow into a more independent researcher.

Koen, thank you for your support and patience during my PhD. The transition from spintronics in my Master's to plasmonics for my PhD was no small feat, and your clear explanations and structured approach were instrumental in making it manageable. Your perfectionism, while occasionally frustrating (yes, I'm looking at those endless paper revisions), always led to better results, and I'm proud of what we achieved together.

I'd also like to thank the rest of the TCM group. In particular, Malte Rösner, for insightful discussions that helped shape our papers into more physically intuitive narratives. Thanks to Misha Titov and Andrey Bagrov for their valuable suggestions and conversations. Belinda, thank you for being the social heart of the group, always checking in on us beyond the research, and for your indispensable administrative support. I always knew I could count on you to bring much-needed structure and a personal touch to the group.

To the rest of the group: Yann, Achille, Manuel, Sergii, Askar, Danis, Sasha, Jan, Anna, Tom, Ilya, Ivan, Vladimir, and many others, thank you for creating such a lively and supportive atmosphere. You made the department a place where science and sanity could coexist. Honorable mention to Fabio (my partner in crime since day one), Robert, Joost, and the S-office in general: grazie, gracias, dankjewel, thank you for the mental support and the many useful discussions, especially towards the end of my PhD. You made the office a place where science and insanity could coexist.

Thanks to my friends who supported me over these four years. To my Nayade friends, to Haps, to my old housemates Vincent, Mark, and Diederik and Jos, and to my "vrienden van vroeger" Iris, Floor, Camiel and Job, and to many other friends: thank you for checking up on me, your presence made the stressful moments easier and the good times even better. I'm also grateful for your patience during the countless times I subjected you to explanations of my work. What might have been an on-the-spot physics lesson for you was an invaluable tool for my own understanding. Of course, I cannot forget my Burgundian friends Robyn, Lennart, Jule, Tom, Yela, Bas, and Daan (who even helped proofread parts of this thesis): I could always count on you guys.

To my family: thank you for your unwavering support. My siblings Janou, Maico, Giano, Dahli, and of course David, thank you for always being there. To my parents, Hans and Wilmy, who were always prepared to go the extra kilometer for me. I truly could not have done this without you. Eline, thank you for the dinners, the encouragement, and much needed distraction during my time in Nijmegen.

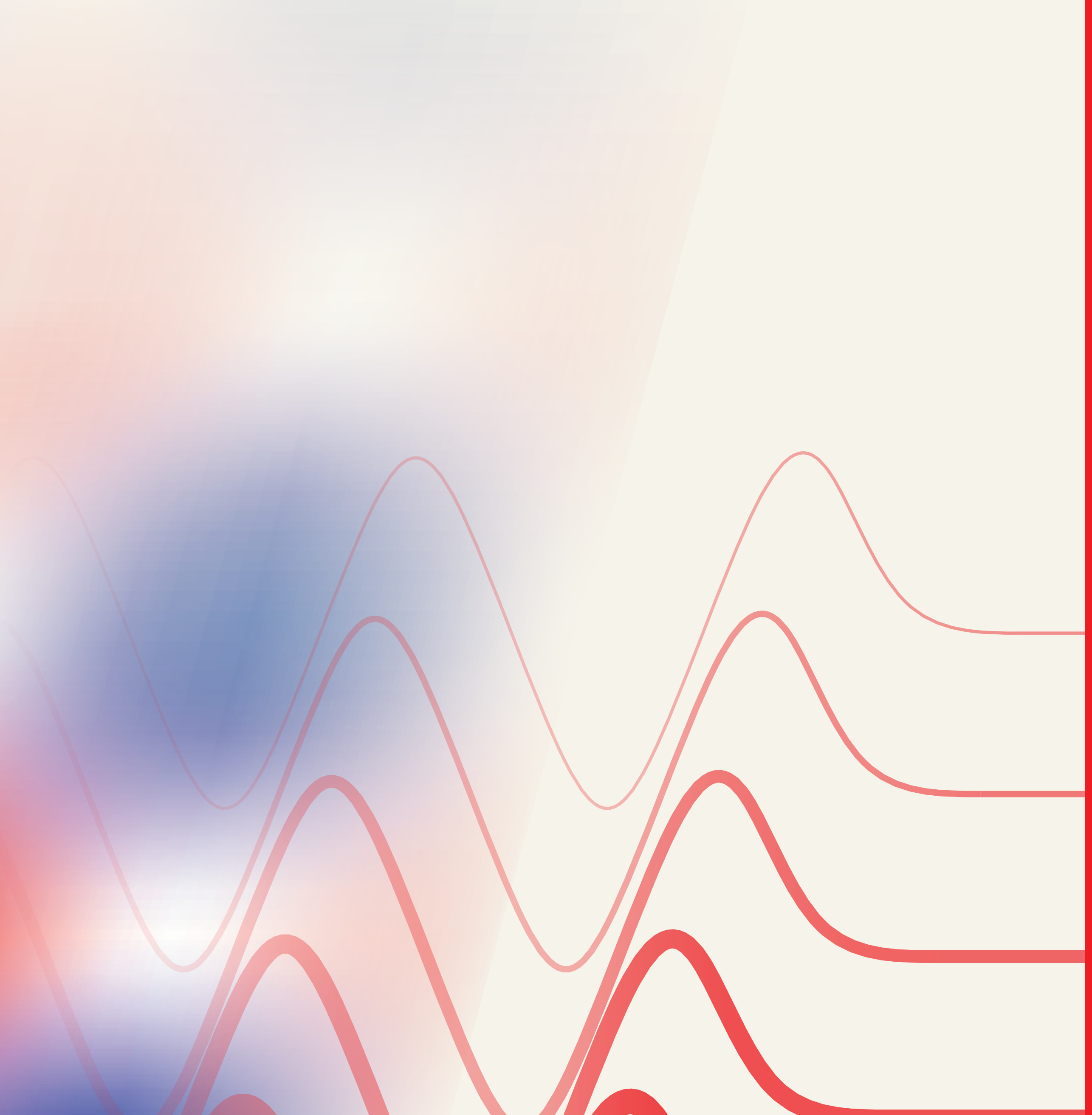
Finally, and most importantly, dankjewel lieve Lotte. You have been my biggest fan and most loyal supporter since day one, and it is to you that I owe the most. Your steadfast and selfless support during the final, intense months was what truly got me through. You put everything aside, welcoming me home late with a prepared dinner and a comforting conversation when I needed it most. I would not have managed to finish in time without your limitless support.

Curriculum Vitae

Tjacco Koskamp was born on August 8, 1995, in the small village of Beugen, The Netherlands, and grew up in the neighboring village of Haps. After graduating from his secondary school, SG Stevensbeek (later renamed Metameer), in 2014, he moved to Eindhoven to study Applied Physics at the Eindhoven University of Technology. For his bachelor's project, he joined the Physics of Nanostructures group (FNA), where under the supervision of Rembert Duine, he investigated the influence of DMI and spin-orbit torques on the current-driven spin-Cherenkov effect, receiving his Bachelor of Science in early 2018.

Tjacco continued in Eindhoven for his master's in Applied Physics, which gave him the opportunity to spend three months abroad. He did an internship at Tohoku University in Sendai, Japan, where he worked with Gerrit Bauer and Koji Sato on the dynamics of skyrmions. Back in Eindhoven, he returned to the Physics of Nanostructures group for his master's project, this time co-supervised by Rembert Duine and Maarten Beens. He explored spin-vorticity coupling in viscous electron systems, a project that deepened his interest in theoretical work and earned him a special "Theory for Technology" endorsement upon his graduation at the end of 2020.

For his PhD, which he started in early 2021, Tjacco switched to Radboud University in Nijmegen. Under the guidance of Mikhail Katsnelson and Koen Reijnders, his research focus shifted from spintronics to the fascinating world of plasmonics, studying the semiclassical approximation to describe plasmons in inhomogeneous 2D media. The results of this project are presented in this thesis. Beyond his research, he enjoyed his role as a teaching assistant for courses like Graphene, Statistical Physics, and Electrodynamics. In his final two years, he also served on the board of the PhD Panel for the Institute of Molecules and Materials, where he helped organize events for his fellow PhD candidates.



9 789465 151847 >

Radboud Universiteit